Proceedings of the

15th Conference of the Italian Association for Cognitive Sciences

AISC 2018 - IUSS Pavia

Edited by

Cristiano Chesi and **Giulia Piredda**

The new era of **Artificial Intelligence**:
a cognitive perspective

IUSS, P.zza Vittoria 15, Pavia

December 17-19, 2018

AISC ASSOCIAZIONE ITALIANA di SCIENZE COGNITIVE

This conference has been organized by **NETS**,
the IUSS Center for Neurocognition, Epistemology and Theoretical Syntax
http://www.nets.iusspavia.it

Pavia, December 17-19, 2018

ISBN: 9788890794384

Cite as appeared in
C. Chesi & G. Piredda (eds.) Proceedings of the 15th Conference of the Italian Association for Cognitive Sciences (AISC 2018). IUSS Pavia, 17-19 December 2018. ISBN: 9788890794384

Preface

This volume contains the papers presented at AISC 2018: 15th Annual Conference of the Italian Association for Cognitive Sciences (AISC) held on December 17-19, 2018 in Pavia.

There were 46 submissions. Each submission was reviewed by at least 2, and on the average 2.2, program committee members and reviewers. The committee decided to accept 39 papers. The program also includes 7 invited talks.

The conference topic of this year was ”The new era of Artificial Intelligence: a cognitive perspective”.

Big data, machine learning, AI algorithms are intensely disputed and acclaimed nowadays. Currently we are assisting to growing investments in computational models simulating neurocognitive functions (e.g. the Human Brain Project) and this trend will probably be enhanced in the near future (Communication Artificial Intelligence for Europe, 25 April 2018) especially in the fields of Machine Learning and Robotics.

The conference focused on the role of Cognitive Science in this debate, (re)discussing the new idea of AI (as compared to the one historically proposed by McCarthy, Minsky, Brooks, Lenat, Rumelhart and McClelland among others) bringing into discussion ideas from Neurocognition, Epistemology and formal Linguistics among other relevant Cognitive Science sub-fields.

This was the list of the invited speakers who presented their insightful perspective on the conference topic during the conference:

Prof.ssa Maria Rita Manzini (Università degli Studi di Firenze)
Prof. Marcello Massimini (Università degli Studi di Milano)
Prof. Silvestro Micera (Scuola superiore Sant'Anna Pisa)
Prof. Alfredo Paternoster (Università degli Studi di Bergamo)
Prof.ssa Gabriella Vigliocco (University College London)
Prof. Charles Yang (University of Pennsylvania)
Prof. Roberto Zamparelli (Università degli Studi di Trento)

We wish to thanks IUSS for economical and logistic support of this initiative. We are especially indebt with Virginia Borasi and Silvia Grecchi for their help in the organization of this conference.

The submission, review process and the creation of these proceedings has been made possible by EasyChair platform.

December 21, 2018 Cristiano Chesi & Giulia Piredda
Pavia

Program Committee

Pauli Brattico	IUSS University School for Adavanced School Pavia
Paolo Canal	IUSS University School for Adavanced School Pavia
Nicola Canessa	IUSS University School for Adavanced School Pavia
Stefano Cappa	IUSS University School for Adavanced School Pavia
Cristiano Chesi	IUSS University School for Adavanced School Pavia
Marco Cruciani	University of Trento
Filippo Domaneschi	Department of Philosophy - University of Genoa
Catricalà Eleonora	IUSS University School for Adavanced School Pavia
Elisabetta Gola	University of Cagliari
Matteo Greco	IUSS University School for Adavanced School Pavia
Antonio Lieto	University of Turin, Department of Computer Science
Paolo Lorusso	IUSS University School for Adavanced School Pavia
Fabio Paglieri	ISTC-CNR Rome
Giulia Piredda	IUSS University School for Adavanced School Pavia
Alfredo Tomasetta	IUSS University School for Adavanced School Pavia

Additional Reviewers

Borsa, Virginia
Conca, Francesca

Table of Contents

Lo sviluppo nel futuro delle scienze cognitive ... 1
Gabriella Airenti

Abilità narrative nei bambini con DSL in età prescolare: un'analisi dei comportamenti linguistici e non-verbali nei compiti di retelling ... 3
Francesca Beraldi, Gloria Gagliardi and Milvia Innocenti

Levels of conceptual format: the challenge of supramodality ... 5
Fabrizio Calzavarini

The computational modelling of inferential and referential lexical competence: a critical review ... 6
Fabrizio Calzavarini and Antonio Lieto

Solving Mathematical puzzles: the viewpoints of Artificial Intelligence and Psychological Cognitive Science ... 8
Federico Chesani, Elisa Gambetti, Fiorella Giusberti, Daniela Loreti and Paola Mello

True recursive domains in language and beyond ... 10
Cristiano Chesi and Valentina Bianchi

Expecting symmetries ... 12
Cristiano Chesi and Andrea Moro

Handwriting e Machine Learning come strumenti al supporto della diagnosi di Alzheimer . 14
Nicole Dalia Cilia, Claudio De Stefano, Francesco Fontanella and Alessandra Scotto Di Freca

Sviluppare l'intelligenza emotiva tramite la finzione. Il ruolo delle emozioni estetiche ... 17
Gianluca Consoli

The extended bodily self and the incorporation of personal property ... 19
Aurora De Bortoli Vizioli, Anna Maria Borghi and Luca Tummolini

Simulating Epistemic Bias in Academic Recruiting ... 21
Carlo Debernardi, Marco Viola and Eleonora Priori

Paranormal beliefs? A matter of core knowledge confusion (that can be simulated by a computational model) ... 23
Sara Dellantonio, Luigi Pastore and Claudio Mulatti

Evaluating fairness in 5-year-old children in a Child-Robot Interaction ... 25
Cinzia Di Dio, Federico Manzi, Shoji Itakura, Takayuki Kanda, Hiroshi Ishiguro, Davide Massaro and Antonella Marchetti

Ad hoc presupposition construction ... 27
Filippo Domaneschi and Diana Mazzarella

Francis Crick and the Hard Problem of Consciousness ... 28
Francesco Ellia

Beyond technological instrumentalism ... 30
Marco Fasoli

On the use of agent-based simulation for cognitive science (symposium) 31
Silvia Felletti, Paolo Pagliuca, Alessandro Sapienza and Daniele Vilone

The Motoric Component of Pictorial Experience 34
Gabriele Ferretti

Un modello induttivo-analogico della diagnosi nosologica 36
Francesco Gagliardi

Nuove simulazioni per il cervello: quale metafora per la mente? 37
Elisabetta Gola

Moral Action Recognition: A Challenge for Artificial Moral Cognition 39
Hasse Hämäläinen

DELTA: A Unifying Categorization Algorithm Integrating Prototypes, Exemplars and Theory-Theory Representations and Mechanisms 41
Antonio Lieto

Composing Prototypes by Coupling a Non monotonic Description Logic with Probabilities and Cognitive Heuristics 43
Antonio Lieto and Gian Luca Pozzato

Mappe concettuali vs ontologie. Un confronto sull'utilizzo di strumenti informatici per la didattica della storia e della filosofia 45
Antonio Lieto and Francesco Vittorio Rebuffo

Sounds in a vacuum: a defence of O'Callaghan's theory. 47
Giulia Lorenzi

Dalla pedagogia naturale alla pedagogia robotica 49
Emiliano Loria

Playing with physical and digital cards to enhance numerical cognition in children 51
Orazio Miglino, Massimiliano Schembri, Antonio Cerrato and Michela Ponticorvo

Constructing a Grammar Based on the Causations of Movements by Joints' Rotations from a Humanoid' Motion Planning 53
Hiroyuki Nishina

Are AI machine learning algorithms applied to text analysis enhancing prejudices? A critical discussion 55
Teresa Numerico

A cognitive model to understand emotions 57
Amitash Ojha and Elisabetta Gola

Fake news about fake news: a socio-cognitive perspective on four myths of online disinformation 59
Fabio Paglieri

Rhythmic and Morphosyntactic predictions: The anticipation skills of Italian children with developmental dyslexia 61
Valentina Persici, Natale Stucchi and Fabrizio Arosio

Wintertime Adaptation 63
Alessio Plebe

Explaining the behavior of learning classification systems: a black-box approach 65
Roberto Prevete, Andrea Apicella, Francesco Isgrò and Guglielmo Tamburrini

A Dynamic Cognitive Model of Bilingual Mental Lexicons 67
Cheng Qian

Can the Human Brain be replicated? 69
Steven S. Gouveia

The effects of stochastic environmental variations and plasticity on evolutionary adaptation and developmental memory 71
Stefano Tiso, Nicola Milano and Stefano Nolfi

Linguistic, social and interoceptive experiences during the processing of abstract concepts 74
Caterina Villani, Luisa Lugli, Fabrizio Campanile, Marco Tullio Luizza and Anna Maria Borghi

Cognitive functions and neural structures: population-bounded mappings 76
Marco Viola

Lo sviluppo nel futuro delle scienze cognitive

Gabriella Airenti[1]

[1]*Dipartimento di Psicologia, Center for Logic, Language, and Cognition, Università di Torino*

L'idea originaria da cui è nata la scienza cognitiva alla fine degli anni settanta del secolo scorso era di superare la psicologia tradizionale arricchendo lo studio della mente con il contributo di altre discipline come l'intelligenza artificiale, la linguistica, le neuroscienze, la filosofia e l'antropologia. Questo approccio, che allora definiva l'ambito della scienza cognitiva, è stato molto presto reinterpretato come la ricerca di un metodo unificante per lo studio della mente. Si è così sviluppato un legame privilegiato tra psicologia e intelligenza artificiale. La principale metodologia utilizzata prevedeva di costruire modelli computazionali che esplicitassero le rappresentazioni mentali che producono le diverse forme di azione umana (Pylyshyn, 1984).

Questo approccio simbolico si è sviluppato al prezzo di separare la mente dalla sua base biologica e dal contesto in cui l'attività umana si svolge. Non c'era spazio per sviluppo, apprendimento, variabilità e in generale per ogni tipo di cambiamento dovuto a cause biologiche o sociali. Anche l'avvento del connessionismo non ha cambiato questo stato di cose in modo sostanziale. Indubbiamente le reti neurali hanno dato un'importanza centrale all'apprendimento. Tuttavia le reti neurali si sono rivelate nel tempo come un metodo alternativo per costruire modelli computazionali senza manifestare la capacità di far emergere rappresentazioni strutturate (Gentner, 2010). Inoltre negli sviluppi recenti dell'intelligenza artificiale le reti neurali sono state largamente sostituite da varie tecniche di apprendimento statistico (Forbus, 2010).

Una svolta nelle scienze cognitive si è verificata con l'introduzione del concetto di mente incarnata (Varela *et al.*, 1991). Secondo questo approccio, non si possono modellare le funzionalità della mente trascurando il fatto che operano sul mondo esterno tramite il corpo. Questa svolta ha corrisposto a un'importanza maggiore assunta dalla robotica che si propone di costruire esseri artificiali che possano cooperare con soggetti umani in molteplici compiti che includono, per esempio, l'assistenza a persone anziane o portatrici di handicap. Attualmente è soprattutto la robotica sociale che cerca di stabilire una connessione con le scienze biologiche, la psicologia e le neuroscienze allo scopo di costruire robots che possiedano funzionalità che permettano loro di interagire con successo col mondo esterno fisico e sociale (Airenti, 2015; Wiese *et al*., 2017).

La questione che si può sollevare oggi riguarda la natura stessa della scienza cognitiva. La relazione privilegiata che la psicologia ha stabilito con le scienze dell'artificiale per un certo periodo è stata molto produttiva e ha generato idee interessanti e nuovi modi di vedere problemi classici, si pensi per esempio alla ricerca su memoria e rappresentazione della conoscenza. Tuttavia, appare anche chiaramente come l'esigenza di costruire modelli computazionali si sia rivelata un limite che ha ristretto l'ambito della ricerca. Già Hewitt (1991) notava la difficoltà di costruire sistemi artificiali che siano fondati, come i sistemi sociali, su concetti quali impegno, cooperazione, conflitto, negoziazione, ecc. Un gran numero di fenomeni essenziali per spiegare il funzionamento della mente umana sono stati quindi largamente ignorati perché difficilmente formalizzabili. Questo ha portato ad una visione rigida e statica della mente che non tiene conto del fatto che la mente stessa è il prodotto dell'interazione col mondo fisico e sociale in un continuo processo di sviluppo.

La psicologia dello sviluppo ha avuto per molto tempo un ruolo secondario all'interno della scienza cognitiva proprio a causa del fatto che, per definizione, essa pone al centro del suo interesse tutte le questioni che la scienza cognitiva ha tradizionalmente ignorato, vale a dire il concetto stesso di sviluppo e la parte che biologia e cultura hanno nelle sue manifestazioni. Il problema che gli psicologi dello sviluppo hanno incontrato riguarda il metodo. Gli psicologi dello sviluppo utilizzano metodologie diverse che comprendono certamente esperimenti ma anche lavoro sul campo, osservazioni controllate, parent-report. Lo sviluppo è un fenomeno complesso che può essere analizzato solo utilizzando diversi punti di vista. Un caso paradigmatico è quello dello studio della teoria della mente. Il problema della rappresentazione della mente dell'altro (o della propria) è emerso all'inizio nella ricerca sui primati (Premack e Woodruff, 1978) ed è diventato uno dei filoni principali della ricerca sullo sviluppo in generale. Attualmente, viene studiato nei primati sia umani che non umani, in gruppi di età diversi dai neonati agli anziani, in soggetti a sviluppo tipico e atipico, utilizzando metodologie diverse che comprendono tecniche sperimentali, ricerca sul campo e osservazioni cliniche.

La mente umana è complessa e i diversi metodi che sono stati proposti nelle diverse discipline che la studiano possono essere utili per aumentare la conoscenza che ne abbiamo. Dar conto di questa complessità era lo scopo principale che ha portato alla nascita della scienza cognitiva ed è il ritorno a questa prospettiva che può permettere di darle un futuro.

Riferimenti bibliografici

Airenti, G. (2015). The cognitive bases of anthropomorphism: from relatedness to empathy. *International Journal of Social Robotics*, 7(1), 117-127.

Forbus, K.D. (2010). AI and Cognitive Science: The past and next 30 years. *Topics in Cognitive Science*, 2, 345-356.

Gentner, D. (2010). Psychology in cognitive science: 1978-2038. *Topics in Cognitive Science,* 2, 328-344.

Hewitt, C. (1991). Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence*, 47, 79-106.

Premack, D., e Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515–526.

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: The MIT Press/Bradford.

Varela, F.D., Rosch, E., e Thompson, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.

Wiese, E. Metta, G., e Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology,* 8:1663. doi: 10.3389/fpsyg.2017.01663.

Abilità narrative nei bambini con DSL in età prescolare: un'analisi dei comportamenti linguistici e non-verbali nei compiti di *retelling*

Francesca Beraldi[1], Gloria Gagliardi[2], Milvia Innocenti[3]
[1]*Università degli Studi di Firenze,* [2]*Università degli Studi di Bologna,* [3]*AUSL Toscana Centro*

Introduzione

È di frequente riscontro nella pratica clinica che i bambini con disturbo del linguaggio [APA, 2013; Bishop *et al.*, 2017] vedano riequilibrate in seguito a trattamento logopedico gran parte delle abilità lessicali e linguistico-formali, ad eccezione però dell'ambito narrativo. Le abilità narrative non rappresentano soltanto un indicatore del livello di competenza linguistica, ma costituiscono una tappa evolutiva essenziale ai fini dell'organizzazione del pensiero logico e del ragionamento verbale: deficit in questo ambito, comportando difficoltà negli apprendimenti e in campo lavorativo, hanno spesso un forte impatto sulla sfera socio-relazionale [Tomblin *et al.*, 2003; Conti-Ramsden *et al.*, 2009; Durkin *et al.*, 2009]. Lo studio si propone perciò di valutare le abilità di *retelling* (ovvero di rievocazione di una breve storia) in bambini di età prescolare con diagnosi di DSL, analizzandone gli aspetti linguistici e pragmatici anche in relazione al *medium* (cartaceo o multimediale) con cui la storia viene loro presentata.

Metodo

Il campione è composto da 16 bambini monolingui (13 M; 3 F) di età compresa tra 4;2 e 5;4. 8 di essi, tutti maschi, sono stati reclutati tra i pazienti in trattamento presso la AUSL Toscana Centro in seguito a diagnosi di DSL di tipo espressivo; gli altri 8 (5 M; 3 F), che costituiscono il gruppo di controllo, hanno sviluppo cognitivo e linguistico tipico. Le produzioni narrative sono state elicitate utilizzando tre diverse prove:

- Bus Story test [Renfrew, 2015], prova standardizzata di *retelling* sollecitato da tavole cartacee figurate;
- *retelling* de "I tre porcellini", proposto attraverso un libro cartaceo illustrato;
- *retelling* del cortometraggio "Orso polare" (durata della storia: 90 sec.), proposto mediante tablet.

La generazione della storia è stata agevolata dalla permanenza del supporto durante lo svolgimento della prova.
Le sessioni sono state videoregistrate, trascritte ortograficamente con il software ELAN (https://tla.mpi.nl/tools/tla-tools/elan/) in formato CHAT-LABLITA [MacWhinney, 1991; Cresti & Moneglia, 2005] (fig.1) e analizzate sotto il profilo lessicale, morfo-sintattico, pragmatico e narrativo (tab.1).
L'effettiva significatività statistica delle differenze tra DSL e gruppo di controllo è stata valutata attraverso test statistici non parametrici in ragione della ridotta numerosità campionaria (χ^2 e Kolmogorov-Smirnov).

LIVELLO DI ANALISI	INDICI
Fluenza	fluenza narrativa false partenze pause (vuote/piene)
Tratti lessicali e morfo-sintattici	type/token ratio [Holmes & Singh, 1996] MLU [Brown, 1973] n. enunciati *verbless* n. enunciati interrotti n./percentuale di principali, coordinate e subordinate errori di morfologia libera/legata n. (e correttezza) dei clitici coerenza dei tempi verbali
Abilità Narrative	grammatica delle storie
Comunicazione non verbale	Gesto mimica: sorrisi, atteggiamenti di ricerca di conferma e espressioni concordi all'emozione del racconto contatto oculare

Tabella 1. Descrizione degli indici linguistici considerati nello studio

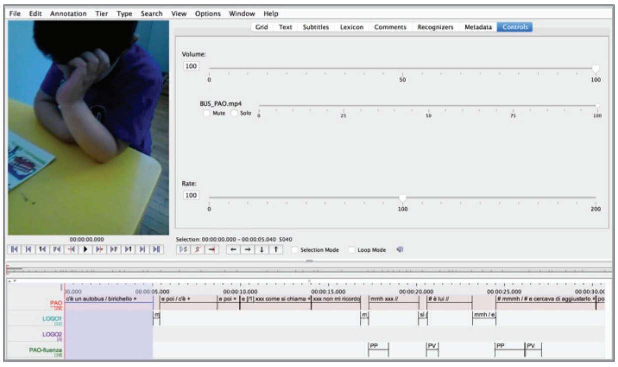


Figura 1. Trascrizione delle sessioni

Risultati

L'analisi conferma i dati riportati in letteratura: i bambini con DSL manifestano difficoltà persistenti nell'attività narrativa, poco fluente, e nell'utilizzo dei clitici [Bortolini et al., 2006], il più delle volte omessi. Per quanto riguarda i comportamenti non verbali, nei DSL si riscontra inoltre una forte prevalenza di gesti non connessi al discorso (auto- ed etero-adattatori), e un contatto oculare più breve con l'interlocutore.

Le modalità di presentazione della storia influiscono fortemente sull'attività narrativa: nell'intero campione, infatti, le performance sono risultate peggiori nella prova "Orso polare". Secondo le autrici tali discrepanze non sono imputabili alla diversa "difficoltà" dei racconti, tutti caratterizzati da una struttura eventiva elementare, né alla differente "notorietà" delle storie, per lo più ignote ai bambini. Tali dati, seppur parziali, sembrano confermare l'importanza di incoraggiare e sostenere la verbalizzazione di storie, preferendo in queste attività il libro cartaceo. Da un punto di vista più generale, le difficoltà evidenziate sono presumibilmente legate non soltanto a deficit di tipo linguistico, ma riconducibili a fragilità in altri ambiti cognitivi, in particolare a carico della memoria procedurale, delle funzioni esecutive (pianificazione e attenzione) [Pinton, 2018] e di Teoria della Mente, non esplicitamente indagate nello studio.

Bibliografia

APA - American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, Fifth edition (DSM-5)*. Washington, DC-London: American Psychiatric Publishing.

Bishop, D.V.M., Snowling, M.J., Thompson, P.A., Greenhalgh, T., & the CATALISE consortium (2017), Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10):1068–1080.

Bortolini, U., Arfé, B., Caselli, M.C., Degasperi, L., Deevy, P., Leonard, L.B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *International Journal of Language & Communication Disorders*, 41(6):695-712.

Brown, R. (1973), *A first language: The early stages*. London: George Allen & Unwin.

Conti-Ramsden, G., Durkin, K., Simkin, Z. & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders*, 44(1): 15-35.

Cresti, E. & Moneglia, M. (2005). *C-ORAL-ROM. Integrated Reference Corpora for spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Durkin, K., Simkin, Z., Knox, E. & Conti-Ramsden, G. (2009). Specific language impairment and school outcomes. II: Educational context, student satisfaction, and post-compulsory progress. *International Journal of Language & Communication Disorders*, 44(1): 36-55.

MacWhinney, B. (1995), *The CHILDES Project: Tools for analyzing talk*. Mahwah: Lowrence Erlbaum Associates.

Pinton, A. (2018). *I disturbi fonetico-fonologici nell'età dello sviluppo*. Roma: Carocci.

Renfrew, C. (2015). *Bus Story Test. A test of narrative speech*. Speechmark.

Tomblin J.B. Zhang X., Buckwalter P., O'Brien M. (2003), The stability of Primary Language Disorder four years after kindergarten diagnosis. *Journal of speech, language, and hearing research*, 46(6):1283-1296.

The computational modeling of inferential and referential competence

Fabrizio Calzavarini[1,2], Antonio Lieto[1,2]
[1]University of Turin, Italy, [2]Center for Logic, Language, and Cognition (LLC)

Introduction

In philosophy of language, a distinction has been proposed between two aspects of lexical competence, i.e. referential and inferential competence (Marconi 1997). The former accounts for the relationship of words to the world, the latter for the relationship of words among themselves. The distinction may simply be a classification of patterns of behaviour involved in ordinary use of the lexicon. Recent research in neuropsychology and neuroscience, however, suggests that the distinction might be neurally implemented, i.e., that different cognitive architectures with partly distinct neural realizations might be responsible for cognitive performances involving inferential and referential aspects of semantics. This hypothesis is strongly consistent with patient data supporting the notion of a functional double dissociation between inferential and referential abilities, and with a set of direct cortical mapping studies and neuroimaging experiments suggesting that inferential and referential abilities are underpinned by at least partly different regions of the human brain (e.g., Marconi et al. 2013; review in Calzavarini 2017).

The initial hypotheses formulated in the setting of the philosophy of language, along with the neuropsychological experimental evidence (about how referential and inferential competences may be neurally instantiated) can be the input to computational modelling activities involving the inferential and the referential aspects of lexical competence. The aim of the talk is to offer a critical discussion of the kind of formalisms that can be used to model the two aspects of lexical competence, and of the main difficulties related to the use of these computational techniques.

Inferential (=symbolic) vs referential (=connectionist) formalisms

The first conclusion of our discussion is that the distinction between inferential and referential semantics is instantiated in the literature of Artificial Intelligence by the distinction between symbolic and connectionist approaches. On the one hand, symbolic formalisms (e.g., meaning postulates, semantic networks, frames, distributional models) are successfully utilized to model a variety of inferential semantic tasks like semantic inference, definition naming, synonym judgments, word-word matching, and so on. The symbolic approach to inferential competence is not immune from problems. For instance, it is know that symbolic formalisms have difficulty in modelling prototypical knowledge and deafasible inference, two essential aspects of human lexical inferential competence (Lieto et al. 2012). The critical point here is that, for conceptual reasons, none of the symbolic formalism appears to be able to model the referential aspect of human lexical competence, i.e. our ability to apply words to objects in the world through perceptual recognition, particularly to visual recognition (Marconi 1997; Harnad 1980).

On the other hand, connectionist approaches (i.e. neural networks, including deep neural networks) are today very effective in modelling referential tasks like visual recognition and naming and are the most widely adopted framework in artificial vision (e.g., Szeliski, Springer, 2010). Differently from symbolic representations, neural networks receive input data directly coming from sensory systems, as images, signals, and so on, and thus the problem of grounding representations to entities in the external world is in some sense alleviated. The importance of neural networks for symbol grounding has been discussed by Harnad in a seminal paper (Harnad 1980). From this point of view, the main advantage of deep neural networks, and in particular of Convolutional Neural Networks, is that they are even closer to sensory data, and therefore they need less or no preprocessing of input data. Among the problems that affect this class of formalisms, however, is that one of not being able to model inferential mechanisms such as taxonomic relations and semantic composition (Lieta, Chella, & Frixione 2017).

The need of integration

The second conclusion of our discussion is that the modelling of lexical competence needs the advent of hybrid models integrating symbolic and connectionist frameworks. Such integration is hard to obtain at this state of knowledge (and this can provide, we claim, further indirect evidence for the cognitive reality of the inferential/referential distinction). While, in fact, existing hybrid systems and architectures such as CLARION (Sun 2006) or ACT-R (Anderson et al. 2004) are able to combine different kinds of representations, nonetheless this kind of integration is usually ad hoc based (Chella, Frixione, & Gaglio, 1998) or, as we will try to show in our presentation, is only partially satisfying. Our hypothesis is that Conceptual Spaces, a framework developed by Gärdenfors (2000) more than fifteen years ago, can offer a lingua franca that allows to unify and generalize many aspects of the representational approaches mentioned above and to integrate "inferential" (=symbolic) and "referential" (=connectionist) computational approaches on common ground. The integration referential/inferential integration via Conceptual Spaces seems nowadays more viable due to the enhancements provided by the existing systems. In particular, we claim that a straightforward way for such integration could combine the overall visual architecture provided in Chella, Frixione & Gaglio (2003) with the grounding mechanisms between conceptual spaces and symbolic representations available in the system DUAL-PECCS (Lieto et al. 2015).

References

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological review 111(4), 1036 (2004) 2.

Calzavarini, F.: Inferential and referential lexical semantic competence: A critical review of the supporting evidence. Journal of Neurolinguistics 44, 163–189 (2017) 3.

Chella, A., Frixione, M., Gaglio, S.: A cognitive architecture for artificial vision. Artificial Intelligence 89(1), 73–111 (1997) 5.

Chella, A., Frixione, M., Gaglio, S.: Anchoring symbols to conceptual spaces: the case of dynamic scenarios. Robotics and Autonomous Systems 43(2), 175–188 (2003)

Frixione, M., & Lieto, A. (2012). Representing concepts in formal ontologies. Compositionality vs. typicality effects. Logic and Logical Philosophy, 21(4), 391-414.

Gardenfors, P.: Conceptual spaces: The geometry of thought. MIT press (2000)

Harnad, S.: The symbol grounding problem. Physica D: Nonlinear Phenomena 42(1-3), 335–346 (1990)

Lieto, A., Radicioni, D. P., & Rho, V. (2017). Dual PECCS: a cognitive system for conceptual representation and categorization. Journal of Experimental & Theoretical Artificial Intelligence, 29(2), 433-452.

Lieto, A., Chella, A., & Frixione, M. (2017). Conceptual spaces for cognitive architectures: A lingua franca for different levels of representation. Biologically inspired cognitive architectures, 19, 1-9.

Marconi, D.: Lexical competence. MIT press (1997)

Marconi, D., Manenti, R., Catricala, E., Della Rosa, P.A., Siri, S., Cappa, S.F.: The neural substrates of inferential and referential semantic processing. cortex 49(8), 2055– 2066 (2013)

Sun, R.: The CLARION cognitive architecture: Extending cognitive modeling to social simulation. Cognition and multi-agent interaction pp. 79–99 (2006)

Solving Mathematical puzzles: the viewpoints of Artificial Intelligence and Psychological Cognitive Science[1,2]

Federico Chesani[a], Elisa Gambetti[b], Fiorella Giusberti[b], Daniela Loreti[c], Paola Mello[a]
[a]*Department of Computer Science and Engineering – DISI, University of Bologna*
[b]*Department of Psychology, University of Bologna*
[c]*CIRI Health Sciences & Technologies, University of Bologna*

The field of Artificial Intelligence (AI) has undergone undoubted progress in recent years, thanks to the advances offered by the current big-data era. Although the research has led, for example, to machines able to defeat human champions in several games, the challenge of a computer autonomously solving mathematical puzzles from their description in natural language text and diagrams is still open. For example, a computer recently beat a world champion in the Go game (Silver et al., 2016). However, if we consider the simple puzzle "Jacob, Lucy and Frank are three friends whose ages add up to 28 years. How many years later will their ages total 37 years?", currently there is no computer able to understand and find the solution. Conversely, primary school children, aged 9/10 years, can solve it without any complex computation ability, or accessing a huge quantity of data (Jonassen, 2000). In general, we would say that a shift from "big-data" towards "big-reasoning" is needed (Chesani et al., 2017).

Mathematical puzzles are iconic examples of the problem-solving capabilities we are looking for. They are usually described through natural language texts and diagrams. Overall, finding the answer requires neither advanced mathematical skills, nor extensive domain knowledge. Essentially it requires natural language understanding (reading skills and semantic processing), working memory, long-term and short-term memories, inductive and deductive reasoning (fluid intelligence), and quantitative knowledge (basic math skills) (Swanson and Beebe-Frankenberger, 2004). Hence, children's ability to solve such puzzles poses several questions to AI practitioners.

A first step would be to share (and possibly align) the AI viewpoint of mathematical puzzles with models and terms from psychological cognitive studies, such as the CHC theory of cognitive abilities (Flanagan, 2008), which is one of the more known psychometric-based taxonomy of human cognitive abilities.

A second step is to analyze the problem-solving process and understand how cognitive theories could foster new models in AI. In intelligent human behavior, "goals are achieved by assembling a series of sub-tasks, creating structured mental programs" that allow individuals to: (a) select information relevant to the current decision, (b) rapidly change the mental focus when the cognitive context modifies, (c) organize several steps or cognitive epochs following a sequence (Duncan et al., 2000). Recent studies have identified specific regions of the brain, such as the Multiple-Demand system of frontal and parietal cortex, which are strongly correlated with mental programming and standard tests of fluid intelligence (Duncan, 2010). It should also be considered that individuals differ in their problem-solving strategies depending on their cognitive abilities and the complexity of the problems. Identifying successful (and unsuccessful) strategies for solving puzzles, or searching for patterns, could suggest new insights for AI. For example, a new application for analyzing sequential eye movements showed that successful Raven solvers (Raven et al., 1998) use a constructive matching strategy (i.e., the formulation of the missing element on the basis of matrix information and looking for that element in the response area), rather than

[1] This work has been partially supported by the ASIA-GiM project, funded by the University of Bologna within the AlmaIDEA initiative.
[2] The authors wish to thank Ing. Riccardo Buscaroli, who contributed in the classification of mathematical puzzles.

a response elimination strategy (i.e., the inspection of alternatives in order to evaluate whether each fits into the empty matrix slot) (Raven et al., 1998).

Specialized approaches able to successfully solve various problems (related to mathematical puzzles) have been proposed and implemented in AI: the "three friends" problem, for example, can be viewed as a constraint-based problem, with three variables and few constraints. Current Constraint Solvers can deal with millions of variables and constraints successfully tackling problems of a higher difficulty level. However, the extreme specialization of AI has led to its fragmentation into several sub-fields, thus overlooking the viewpoint of a general intelligence approach, where many different cognitive skills, such as text understanding, diagram comprehension, knowledge elicitation and extraction, etc., are all merged together into a comprehensive, fully-fledged approach. Several Cognitive Architectures have been proposed in the past, often focusing on how to organize procedural/declarative knowledge and reasoning processes. Among the proposed solutions, we mention ACT-R (Anderson, 1983) and SOAR (Newell, 1990). Recently, a tentative of suggesting a unified, standard model of cognitive processes has also been proposed (Laird et al., 2017). Given the number of achieved results, these architectures can be considered a good starting point, although the problem we face is not about which architecture is better, but rather how to achieve, in a non ad-hoc way, integration in an AI system of the mentioned cognitive skills.

The ASIA-GiM project aims to investigate how recent AI achievements, and psychological cognitive models and studies could provide new insights towards the definition of fully-fledged intelligent agents able to solve mathematical puzzles. We present the initial observations the project builds on, its aim, and a primary classification of mathematical puzzles from the viewpoints of AI, and of Psychological Cognitive studies.

References

Anderson J. R. (1983): *The architecture of cognition*. Harvard University Press

Chesani F., Mello P., Milano M. (2017). Solving Mathematical Puzzles: A Challenging Competition for AI. *AI Magazine* 38(3), 83-96

Duncan J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences* 14(4), 172-179

Duncan J., et al. (2000). A neural basis for general intelligence. *Science* 289, 457–460

Flanagan D. P. (2008). The CattellHornCarroll Theory of Cognitive Abilities. *Encyclopedia of Special Education* 368-382

Hayes T. R., Petrov A. A., Sederberg P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. Journal of Vision 11(10):1-11

Jonassen D. H. (2000). Toward a design theory of problem solving. *Educational technology research and development* 48(4), 63-85.

Laird J. E., Lebiere C., Rosenbloom P. S. (2017). A Standard Model for the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine* 38(4).

Newell A. (1990). *Unified theories of cognition*. Harvard University Press

Raven J. C., Raven J., Court J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Section 4: Advanced progressive matrices. San Antonio, TX: Pearson

Silver D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587), 484-489

Swanson H. L., Beebe-Frankenberger M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology* 96(3), 471.

True recursive domains in language and beyond

Valentina Bianchi[1], Cristiano Chesi[2]
[1]University of Siena, [2]IUSS Pavia

On its path towards explanatory adequacy (and beyond, Chomsky 2001), syntactic theory focused heavily on the notion of recursion as the fundamental property of the simplest possible operation (Merge) for building syntactic structures. *Merge* is a binary operation creating sets (i.e. hierarchies) that applies freely to lexical items or to the results of previous *Merge* operations (in this sense it is considered "recursive", Zaccarella & Friederici 2015):

1. *Merge(the, apple) = {the, apple}*
2. *Merge(ate, {the, apple}) = {ate, {the, apple}}*
3. *Merge(Eva, {ate, ... }) = {Eva, {ate, {the, apple}}}*

Despite its centrality in a derivational approach to phrase structure, this minimal definition of Merge has been questioned both on formal and on cognitive grounds. From a formal perspective, it seems inefficient to let Merge apply freely, and then select the results on the basis of independent constraints (such as labeling, Rizzi 2016, or select, Collins & Stabler 2016; see Chesi 2015 for discussion). From the cognitive standpoint, there is a striking contrast between abstract competence (possibly leading to infinite recursion) and real performance (limiting recursion in most contexts, e.g. center embedding): this has been taken as evidence against the necessity for recursion tout court (since Christiansen 1992).

In this paper we argue that Merge must be prevented from creating illegal constituents. In particular, we propose that the application of Merge is sensitive to the selection properties of the items involved; then the selection properties single out a subset of lexical items that are able to introduce recursion.

In a nutshell:

a. We reduce selection to functional application: when a function-denoting item X is processed, it introduces the expectation for an argument Y of a specific semantic type.
b. Following Zamparelli 2000, Ramchand & Svenonius 2014 (a.o.), so called functional layers (e.g. CP, TenseP, DP) select each for one specific expansion in a non-recursive fashion (e.g., Aspect selects for one and only one property of events).
c. True recursive domains reduce to phases, namely domains introduced by lexical items that have the ability to select a higher functional layer and license, at the bottom of the functional hierarchy, another lexical item with the same property (nouns, verbs and predicative adjectives that can select QPs/DPs, CPs etc.)
d. The distinction between recursive and non-recursive items has a number of interesting cognitive implications, two of which we will discuss: an explicit definition of computational nesting; a possible parallel with the idea of expectation-driven semantic interpretation and of goal-driven actions and planning (when goals are intended as "anticipatory internal representations", Castelfranchi & Paglieri 2007).

References

Castelfranchi, C., & Paglieri, F. (2007). The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. Synthese, 155(2), 237-263.

Collins, C., & Stabler, E. (2016). A formalization of minimalist syntax. Syntax, 19(1), 43-78.

Christiansen, M. H. (1992). The (non) necessity of recursion in natural language processing. In Proceedings of the 14th annual conference of the cognitive science society (pp. 665-670).

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. science, 298(5598), 1569-1579.

Chesi, C. (2015) On directionality of phrase structure building. Journal of psycholinguistic research, 44(1):65-89.

Chesi, C., & Moro, A. (2018). Il divario (apparente) tra gerarchia e tempo. Sistemi intelligenti, 30(1), 11-32.

Heim, I. & Kratzer, A. (1998). Semantics in generative grammar. Blackwell.

Rizzi, L. (2016). Labeling, maximality and the head–phrase distinction. The Linguistic Review, 33(1), 103-127.

Zamparelli, R. (2000). Layers in the determiner phrase. Routledge.

Ramchand, G., & Svenonius, P. (2014). Deriving the functional hierarchy. Language Sciences, 46, 152-174.

Chomsky, N. (2001). Beyond explanatory adequacy. MITWPL. 20:1-28

Expecting symmetries:
Top-down derivation and the case of copular sentences.

Cristiano Chesi, Andrea Moro
University School for Advanced Studies IUSS Pavia

Non-feature-driven approach to displacement

From the classic bottom-up derivational perspective on syntactic structure building (Chomsky 1995), symmetry has been considered the locus of instability within a Dynamic Antisymmetry approach (Moro 2000, building on Kayne 1994) and the trigger for hierarchical (and, as a consequence, linear) reordering. Reordering leads to asymmetric configurations that are unambiguous in their dominance-precedence mapping (the two crucial dimensions of any structural representation) hence "stable".

A different perspective on "(in)stability"

As suggested in Chesi & Moro (2018), from a processing-friendly, more cognitively plausible and (at least) equally empirically adequate perspective, derivations unroll top-down (Chesi 2015), instead of bottom-up, hence the role of symmetry is reversed: instead of triggering displacement, a symmetrical configuration stops on-going displacements by satisfying a selection requirement using a pending (not yet selected) item in active memory. The idea of "active memory" is consonant with the standard filler-gap approach to movement in psycholinguistic and computational literature (Hawkins 1999) and it is implemented by means of a memory buffer in which constituents get stored whenever they are "unexpected": an unexpected/unselected constituent is, for instance, any DP that is not preceded by a selecting predicate; this is the case of the subject in SVO languages. For coherence with other formalisms (e.g. Stabler 1997) we will use the "=" diacritic for marking selection: $run_{=DP}$ will indicate that the verb *run* selects for a DP; if such DP is processed before the verb as in *John runs*, *John*, as soon as it is processed, is stored in memory and retrieved only after the first predicate *runs*, selecting one DP, is inserted in the derivation (in the current case, next word).

The (a)symmetry of predic(a)tion in copular sentences.

Building on the conjecture proposed in Moro (2004) here we explore the idea that symmetry instantiates the fundamental idea of predication: here we will argue in favor of a common underlying structure in which the predicate (independently on its verbal, adjectival or nominal status) will express the fundamental predicative function that will be explicitly marked in the lexicon.

The syntax of copular sentences

Considering the analysis of copular sentences proposed by Moro (1997), summarized in (2), both canonical (2.a) and inverse (2.b) copular constructions can be derived by removing either one of the offending DPs from the Bare Small Clause (BSC):

(1)	a. [$_{DP}$ le foto sul muro]	sono	[$_{BSC}$ _ [$_{DP}$ la causa della rivolta]]
	the pictures of the wall	are	the cause of the riot
	b. [$_{DP}$ la causa della rivolta]	sono	[$_{BSC}$ [$_{DP}$ le foto sul muro] _]
	the cause of the riot	are	the pictures of the wall

Despite the fact that both solutions are perfectly grammatical and instantiate an identical sequence of lexical categories, namely DP copula DP, the two structures show various asymmetries supporting the hypothesis that the two DPs play different roles in the structure. Here, *la causa della rivolta* (the cause of the riot) behaves as the predicate (selecting the other DP as an agent), while *le foto del muro* (the pictures of the wall) behaves as the grammatical subject (as shown in Italian by triggering verbal agreement). Independently from their surface positions, these roles are fixed.

Top-down analysis of copular sentences

From the top-down perspective, we rephrased theta-marking (i.e. the relation between the selecting predicate and the selected DP) as a selection specification: *causa* (cause) selects both for a "result", the PP *della rivolta* (of the riot), and for a "cause" *cause* (causa), the DP "le foto…"):

(2) $[_{nP}$ causa $_{=PP(result)}$] $[_{NP}$ (causa) $_{=DP(cause)}$]]]

To derive the canonical copular sentences, we assume that the copula selects for a nominal predicate[1]; as a consequence the main steps of the derivation are the following ones:

(3) i. $[_{DP}$ le foto sul muro] → **M(emory)**: $[_{DP}$ (le foto …)]
ii. $[_{DP}$ le foto sul muro] [sono $_{=DPpred}$]
iii. … [sono ~~$_{=DPpred}$~~$[_{DP}$ la $[_{nP}$ causa ~~$_{=PP(result)}$~~$[_{PP}$ della rivolta] $[_{NP}$ (causa) $_{=DP(cause)}$]]]]
iv. … [sono ~~$_{=DPpred}$~~$[_{DP}$ la $[_{nP}$ causa ~~$_{=PP(result)}$~~$[_{PP}$ della rivolta] $[_{NP}$ (causa) ~~$_{=DP(cause)}$~~$[_{DP}$ (le foto …)]]]]
← **M(emory)**: ~~$[_{DP}$ (le foto …)]~~

For the inverse copular construction, we assume that the last select feature (=DP_{cause}) is expanded only after the DP predicate (*la causa*) has been remerged as the selected nominal predicate of the copula (cf. Bianchi & Chesi's 2014 analysis of reconstruction under the subject of thetic predicates):

(4) i. $[_{DP}$ la $[_{nP}$ causa ~~$_{=PP(result)}$~~$[_{PP}$ della rivolta] $[_{NP}$ (causa) $_{=DP(cause)}$]]]] → **M(emory)**: $[_{DP}$ (la causa …)]
ii. … [sono $_{=DPpred}$]
iii. … [sono ~~$_{=DPpred}$~~$[_{DP}$ la $[_{nP}$ causa ~~$_{=PP(result)}$~~$[_{PP}$ della rivolta] $[_{NP}$ (causa) $_{=DP(cause)}$]]]]
← **M(emory)**: ~~$[_{DP}$ (la causa …)]~~
iv. … [sono ~~$_{=DPpred}$~~$[_{DP}$ la $[_{nP}$ causa ~~$_{=PP(result)}$~~$[_{PP}$ della rivolta] $[_{NP}$ (causa) ~~$_{=DP(cause)}$~~$[_{DP}$ le foto …]]]]

On the cognitive importance of this idea

This first attempt to derive the syntax of copular sentences in a top-down fashion within a Dynamic Antisymmetry approach leads to at least two general challenging issues: first, to extend this analysis based on instability of symmetrical relations to verbs other than the copula, hence in general on the syntactic mechanisms underlying the cognitive process of predication (see Chesi 2015, 2017 for discussion); second, this "symmetry point" greatly resemble to Resonant States (Carpenter & Grossberg 2016) suggesting the possibility of a specific application of the Adaptive Resonance Theory also to specific linguistic phenomena.

Selected references

Carpenter, G. A., & Grossberg, S. (2016). *Adaptive resonance theory* (pp. 1-17). Springer US.

Chesi C. (2017) Phase-based Minimalist Parsing and complexity in non-local dependencies. Proceedings of CLiC-it 2017. CEUR workshop proceedings, doi: urn:nbn:de:0074-2006-4.

Chesi, C. (2015) On directionality of phrase structure building. *Journal of psycholinguistic research*,44(1):65-89.

Chesi, C., & Moro, A. (2018). Il divario (apparente) tra gerarchia e tempo. *Sistemi intelligenti*, 30(1):11-32.

Chomsky, N. (1995) *The minimalist program*. MIT press.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language*, 244-285.

Kayne, R. S. (1994) *The antisymmetry of syntax*. MIT press.

Maienborn, C. (2003). Against a Davidsonian analysis of copula sentences. In *Proceedings of NELS*, 33:167-186.

Moro, A. (1997). *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge University Press.

Moro, A. (2000). *Dynamic antisymmetry*. MIT press.

Moro, A. (2004). Linear compression as a trigger for movement, in Riemsdijk, H. van - Breitbarth, A (eds.) *Triggers*, Mouton de Gruyter, Berlin.

Stabler, E. (1997). Derivational minimalism. In Retoré C. (ed.) *International Conference on Logical Aspects of Computational Linguistics* (pp. 68-95). Springer, Berlin, Heidelberg.

[1] The copula introduces a referential argument z of type K-state (Kim 1969, 76) which is characterized by the predicate P applying to the individual x: $\lambda P\ \lambda x\ \lambda z\ [z \approx [P(x)]]$ (z ranges over K-states) (Maienborn 2003)

Handwriting e Machine Learning come strumenti al supporto della diagnosi di Alzheimer

N.D Cilia, C. De Stefano, F. Fontanella e A. Scotto di Freca

Dipartimento di Ingegneria Elettrica e dell'Informazione (DIEI)
Universitá di Cassino e del Lazio meridionale
Via G. Di Biasio, 43, 03043, Cassino (FR), Italy

L'Alzheimer (Alzheimer Disease – AD) è una tra le malattie neurodegenerative più diffuse al mondo, il suo esordio non ha una fascia di età definita e il suo sviluppo è inesorabile. La compromissione delle facoltà cognitive è quasi totale; per questa ragione la sua diagnosi precoce ha largo seguito nella ricerca mondiale. Sotto le ormai conclamate teorizzazioni dell'embodied cognition, è maturata la convinzione che l'atto motorio sia anche e soprattutto un atto cognitivo. La scrittura è tra le abilità che risultano da una stretta relazione tra abilità cinestetiche e percettivo-motorie. In letteratura è evidente che queste abilità sono compromesse significativamente nell'AD, nella quale si registrano, in particolar modo, alterazioni nell'organizzazione spaziale e nel controllo fine del movimento (Neils-Strunjias *et al.* 2006). Tuttavia la maggior parte degli studi che analizzano gli effetti sulla cinematica della grafia sono stati condotti nei campi della medicina e della psicologia, dove tipicamente vengono utilizzati strumenti statistici per analizzare la relazione tra la malattia e le variabili prese in considerazione (Onofri *et al*. 2013). A questo grosso limite se ne aggiungono altri tre: innanzitutto non vi è uniformità nel protocollo sperimentale adottato, e questo non permette di avere una chiara visione di quali compiti siano ben discriminanti dell'insorgenza della malattia. Legato al disegno del dataset vi è poi la cardinalità dello stesso: la mancanza di uno studio avente e disposizione un ampio dataset di pazienti riduce, infatti, la possibilità di impiegare tecniche di machine learning a supporto della diagnosi. Infine non vi è uniformità tra le feature (caratteristiche del tratto grafico) prese in considerazione, le quali, se utilizzate come tratti distintivi, permetterebbero di correlare la variazione della scrittura con il deterioramento cognitivo. Scopo di questo lavoro è affrontare ogni questione aperta delle tre sopra esposte costruendo, innanzitutto, un protocollo sperimentale completo a cui sottoporre un ampio campione. In questo modo sarà possibile impiegare algoritmi di classificazione per distinguere il degrado cognitivo dei soggetti esaminati (per una disamina completa si veda De Stefano *et al*. 2017). I soggetti reclutati, di età compresa tra i 47 e gli 85 anni, non facenti uso di farmaci psicotropi, sono stati bilanciati per età ed istruzione. Ad ogni soggetto è stato inoltre assegnato un punteggio ottenuto dalla somministrazione del Mini-Mental State Examination (MMSE) come criterio di validazione della classificazione. Il protocollo sperimentale composto da 25 task comprende attività grafiche, di copia, di memoria e di dettatura. Esso è stato costruito a partire dalla letteratura a disposizione in modo che ogni gruppo di task permettesse di elicitare una diversa funzione cognitiva. All'interno dei gruppi, inoltre, si è fatto uso di diverse tipologie di parole (regolari, non regolari, non parole), di diversa lunghezza e con diversi tratti in ascender e descender (Schroter *et al.* 2003). Alcuni task hanno visto l'impiego di cue per monitorare la capacità nell'organizzazione spaziale del soggetto. I compiti sono stati presentati su fogli bianchi A4, posizionati su una tavoletta grafica (Wacom Bamboo Folio), la quale registra i movimenti della penna, in termini di coordinate spaziali (con z indicante la pressione esercitata durante la scrittura), utilizzata dal soggetto esaminato (Muller *et al*. 2017). La presenza di numerosi task, la cui esecuzione complessiva rimane inferiore ai 20 minuti e quindi di semplice impiego anche con i soggetti ad uno stadio avanzato, è volta all'individuazione di tutte quelle caratteristiche della scrittura che potrebbero essere utili per il rilevamento dell'AD. Dopo la fase di raccolta preliminare, l'obiettivo diventa quello di rimuovere i task che forniscono risultati meno rilevanti in modo da alleggerire il protocollo. Dalle tracce acquisite si sono ricavate feature basate sulla pressione, velocità, accelerazione, slant, jerk, ampiezza del tratto, numero di stroke, ecc… (Fernandes *et al*. 2017). Applicando diverse tipologie di algoritmi di classificazione (Werner *et al*. 2006; Yan *et al*. 2008) alle feature ricavate, aggiungendo età ed istruzione dei soggetti, si sono ottenuti risultati preliminari incoraggianti, i quali mostrano una percentuale di corretta classificazione, dei soggetti analizzati nella rispettiva classe di appartenenza (gruppo dei pazienti e gruppo dei controlli), pari all'88%. Questi risultati sembrano confermare l'ipotesi secondo la quale l'analisi della scrittura attraverso il machine learning può rappresentare un supporto alla diagnosi di malattie neurodegenerative e, in particolar modo, dell'AD. Obiettivo nel prossimo futuro sarà quello di cercare dei marker già presenti negli stadi prodromici della malattia, ovvero nei pazienti con Mild Cognitive Impairment, al fine di diagnosticare in maniera precoce l'insorgere del deterioramento cognitivo.

Bibliografia

De Stefano C., Fontanella F., Impedovo D., Pirlo G., Scotto di Freca A., 2017. A brief overview on handwriting analysis for neurodegenerative disease diagnosis. *Proceedings of the Workshop on Artificial Intelligence with Application in Health (WAIAH17)*, pp. 9-16.

Fernandes C., Lopes Lima J.M., 2017. Alzheimer's disease and handwriting-what do we know so far? *Proceedings of IGS 2017*, 130-134.

Muller S., Preische O., Heymann P., Elbing U., Laske C., 2017. Diagnostic value of a tablet-based drawing task for discrimination of ¨patients in the early course of alzheimer's disease from healthy individuals. *Journal of Alzheimer's disease*, 55, 1463-1469.

Neils-Strunjas J., Groves-Wright K., Mashima P., Harnish S., 2006. Dysgraphia in Alzheimer's disease: a review for clinical and research purposes. *J Speech Lang Hear Res,* 49, 1313-30.

Onofri E., Mercuri M., Salesi M., Ricciardi M., Archer T., 2013. Dysgraphia in relation to cognitive performance in patients with Alzheimer's disease. *Journal of Intellectual Disability-Diagnosis and Treatment*, 1, 113-124.

Schroter A., Mergl R., Burger K., Hampel H., Muller, H.J., Hegerl U., 2003. Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dementia and geriatric cognitive disorders,* 15, 132-42.

Werner P., Rosenblum S., Bar-On G., Heinik J., Korczyn A., 2006. Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment. *Journal of Gerontology: PSYCHOLOGICAL SCIENCES*, 61, 228-36.

Yan J.H., Rountree S., Massman P., Doody R.S., Li H., 2008. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. Journal of Psychiatric Research 42, 1203-1212.

Sviluppare l'intelligenza emotiva tramite la finzione

Gianluca Consoli
Dipartimento di psicologia dei processi di sviluppo e socializzazione
La Sapienza, Università di Roma

Questo articolo sviluppa la tesi in base alla quale una delle principali funzioni della finzione, per il tramite delle cosiddette "emozioni estetiche", è favorire l'esercizio dell'intelligenza emotiva. Tale esercizio consente di praticare tutti gli aspetti dell'intelligenza emotiva e di conseguenza ne facilita lo sviluppo. La tesi di fondo viene argomentata in riferimento all'evidenza sperimentale, oggi estremamente ampia, raccolta negli ultimi due decenni in relazione alle emozioni ordinarie, alle emozioni estetiche e all'intelligenza emotiva in diversi ambiti di ricerca delle scienze neurocognitive. È noto che intorno alla definizione dei principali costrutti teorici chiamati in causa, vale a dire "emozione", "finzione", "emozione estetica" e "intelligenza emotiva", è in corso un dibattito pluridecennale tanto esteso quanto ricco di aspetti controversi. L'intervento, però, non mira né a ricostruire questo dibattito né a contribuire a stabilire definizioni formali. Piuttosto, lo scopo precipuo è individuare degli elementi definitori di base condivisi che consentano di mettere in luce le profonde interrelazioni tra i fenomeni in gioco. In tale ottica, verranno fissate le caratteristiche chiave delle emozioni ordinarie che risultano rilevanti allo scopo di determinare quelle proprietà delle emozioni estetiche che, a loro volta, consentono di dimostrare la capacità di queste ultime di esercitare una diretta influenza sugli aspetti principali dell'intelligenza emotiva. In questo percorso argomentativo, si farà esclusivamente riferimento a ipotesi e costrutti teorici operazionalizzati nella ricerca sperimentale e corroborati da risultati sperimentali.
Nello specifico, a proposito del rapporto tra emozioni ordinarie ed emozioni estetiche, verrà argomentato che tutte le emozioni estetiche derivano il fondamento della loro identità dall'assunzione dell'atteggiamento estetico. In virtù di questo atteggiamento viene istituito uno spazio di lavoro globale di natura finzionale, nel quale cioè viene sospesa l'elaborazione ordinaria di tipo pragmatico, basata sulla conformità ai vincoli referenziali e finalizzata al conseguimento di scopi personali di natura pratica. Risulta in tal modo possibile fare esperienze vicarie e simulative delle credenze e degli scopi implicati nelle emozioni ordinarie senza conseguenze pratiche reali (Scherer e Coutinho 2013; Menninghaus et al. 2017). Su questa base, in linea con un consenso piuttosto significativo tra i vari esperti, vengono distinte – in via euristica e non normativa – due macrotipologie di emozioni estetiche, che si possono definire per comodità valutative e narrative. Nelle prime l'apprezzamento delle qualità intrinseche della rappresentazione finzionale è un fattore determinante dell'emozione (Zenter, Grandjean, Scherer 2008). Le seconde sono modellate sulle narrazioni finzionali e sulle condizioni del trasporto nel mondo della finzione e dell'identificazione con i personaggi (Miall e Kuiken, 2002).
L'intelligenza emotiva, concepita come la cooperazione tra l'intelligenza e le emozioni, viene generalmente definita come "la capacità di ragionare sulle emozioni e la capacità delle emozioni di supportare il pensiero. Essa include l'abilità di percepire accuratamente le emozioni, di accedere e generare le emozioni in modo da assistere il pensiero, di comprendere le emozioni e la conoscenza emotiva, di regolare in via riflessiva le emozioni in modo da promuovere la crescita emotiva e intellettuale" (Mayer e Salovey, 1997). Sulla base di questa definizione, l'intelligenza emotiva viene modellata in quattro dimensioni chiave, ciascuna delle quali è misurata con specifici test: (1) percepire le emozioni; (2) usare le emozioni per facilitare il pensiero; (3) comprendere le emozioni; (4) regolare le emozioni.
Integrando i risultati provenienti da diverse linee di ricerca – in particolare dagli studi sul rapporto tra finzione e cognizione sociale (Mar 2018); dagli studi sulla fluency e sui feeling of knowing (Silvia 2010); dagli studi sugli esperti (Fayn et al. 2018); dagli studi quali-quantitativi sulla defamiliarizzazione (Kuiken et al. 2010) – è possibile mostrare empiricamente che le emozioni estetiche impattano ciascuna delle dimensioni dell'intelligenza emotiva, ne favoriscono l'esercizio

e ne facilitano il miglioramento. Dunque, aldilà delle molte controversie relative a una pluralità di questioni specifiche, si può affermare senza tema di smentita che negli ultimi due decenni le scienze neurocognitive hanno pienamente dimostrato che la finzione, per il tramite delle emozioni estetiche, rivela un formidabile potere formativo sull'intelligenza, in particolare su quella emotiva: migliora la percezione e la comprensione delle emozioni, favorisce l'interazione tra le emozioni e i processi di pensiero, amplia la conoscenza e il repertorio delle emozioni, rafforza la regolazione delle emozioni.

References

Fayn, K., MacCann, C., Tiliopoulos, N., Silvia, P.J. (2015). Aesthetic emotions and aesthetic people: Openness predicts sensitivity to novelty in the experiences of interest and pleasure. *Frontiers in Psychology*, 6.

Kuiken, D., Miall, D., Sikora, S. (2004). Forms of self-implication in literary reading. *Poetics Today*, 25, 171-203.

Mar, R.A. (2018). Stories and the promotion of social cognition. *Current Direction in Psychological Science*, 27, 257-262.

Mayer, J., Salovey, P. (1997). What is emotional intelligence? In P. Salovey, D. Sluyter (Eds.), *Emotional Development and Emotional Intelligence: Educational Implications*, New York: Basic Books, 3-31.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Jacobsen, T., and Koelsch, S. (2017), The distancing-embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences*, 40.

Miall, D., Kuiken, D. (2002). A feeling for fiction: Becoming what we behold. *Poetics*, 30, 221-241.

Scherer, K., Coutinho, E. (2013). How music creates emotion: A multifactorial process approach. In T. Cochrane, B. Fantini, K. Scherer (Eds.), *The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control*, Oxford: Oxford University Press.

Silvia, P. (2010). Confusion and interest: The role of Knowledge emotions in aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts*, 4, 75-80.

Zenter, M., Grandjean, D., Scherer, K. (2008). Emotions evoked by sound of music: Characterization, classification, and measurement. *Emotion*, 8, 494-521.

The extended bodily self and the incorporation of personal property

Aurora De Bortoli Vizioli[1], Anna M. Borghi[1,2], Luca Tummolini[1]
[1] *Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy*
[2]*Department of Psychology, University of Rome La Sapienza, Rome, Italy*

Background and aims

In our everyday life we are typically able to discriminate between objects that belong to others and those that we feel are our own (Scorolli, Borghi & Tummolini, 2017). However the cognitive mechanisms that support this ability are still unclear. Single case studies with patients show that object ownership feelings can be temporarily lost after brain damage ("when I look at my belongings, I felt that they were not mine", Nascimento Alves et al, 2016). Intriguingly, the case of a somatoparaphrenic patient who denied ownership of her left hand as well as of personal property related with it (Aglioti et al, 2006) suggests that the experience of owning property can be directly based on the incorporation of objects in our body image (incorporation). To explore this possibility with a population of healthy participants, here we propose a new design aimed to create an "illusion" of ownership over objects owned by others and, conversely, an "illusion" of dis-ownership over self-owned objects (rings) by manipulating the sense of ownership of the body part associated with them (hand).

Methods

We employed a modified version of the rubber hand illusion (Botvinick & Cohen, 1998) in which each participant as well as the rubber hand wore a ring. Feelings of ownership and dis-ownership were measured through a word-picture matching task (adapted from Sui & Humphreys, 2012) in which participants had to judge whether an object picture (own hand/own ring vs. rubber hand/not-owned ring) and the labels ME/OTHER were correctly matched or not. In Experiment 1 (41 women, age M = 23.4 years, SD = 3.9), participants wore their own rings. The experiment was designed to test the effect of body-ownership manipulation on long-term object-ownership. In Experiment 2 (43 women, age M = 23.1 years, SD = 4.6), participants were given a choice between two rings. After the choice, the chosen ring was property of participant. In this case, the aim was exploring the effect of body-ownership manipulation on short-term object-ownership. The procedure of the two experiments was identical. Before the beginning of the experiment, the experimenter took some pictures of participants' left hand and ring. These pictures served as stimuli in the word-picture matching task. Participants then were asked to complete the first block of the task. After that, they repeated the task twice (block 2 and 3) but experienced two 2-min sessions of Rubber hand illusion (RHI) before each block. Blocks 1 2 3 were identical to each other. After the experiment, participants completed a questionnaire (adapted from Longo et al., 2008) rating their subjective experiences of the RH illusion.

Hypotheses

Ownership Hypothesis: if the illusion of ownership of the fake hand extends to the ring associated with it, we expect slower RTs in the synchronous than in the asynchronous condition when perceiving objects owned by someone else.

Dis-ownership Hypothesis: if the illusion of dis-ownership of the biological hand extends also to the ring associated with it, we expect slower RTs in the synchronous than in the asynchronous condition when perceiving self-owned objects.

Results

To assess whether owned objects become part of the conceptual self, we checked for presence of a self-prioritization effect (faster RTs and better accuracy in self-related matched trials) for both self-owned

hand and ring in the two experiments. We adopted a bootstrapping procedure combining accuracy and RT performance at baseline (block 1). Results of the bootstrap analysis show evidence of a self-prioritization effect in both experiments: faster RT and better accuracy when perceiving objects owned by the self compared to objects owned by others. This result suggests that the conceptual self is extended and includes both long-term (Experiment 1) and short-term (Experiment 2) owned objects.
To explore whether changes in the sense of body ownership affect the conceptual self, we assessed the effect of body-ownership manipulation (synchronous vs asynchronous stimulation) on RT performance in block 2 and 3. We adopted a mixed-effects model approach. In Experiment 1 response times on self- and other-ownership matched trials are slower in the Synchronous than in the Asynchronous condition, $b = 67.5$, $t= 2.480$, $p = 0.0163$. In Experiment 2 we do not find evidence that response times on self- and other-ownership matched trials are affected by the experimental condition, $b = 4.48$, $t= 0.156$, $p = 0.877$. Results of Experiment 1, but not of Experiment 2, confirmed both the ownership and dis-ownership hypotheses.

Conclusion

The mental representation of the self comprises conceptual information (semantic and episodic knowledge relevant to the self) as well as perceptual information regarding the body. These pieces of conceptual and perceptual information are malleable and linked (Maister et al, 2015). Here we have shown that the "conceptual" representation of the self includes also information regarding one's own personal property (Experiment 1) and that it is malleable: i.e. the conceptual representation of the self is quickly extended also to include recently acquired property (Experiment 2). Moreover, we have shown that changing the "bodily" representation of the self can affect the "conceptual" representation. Results indicate that undermining the bodily self (via body manipulation) has cascade effects on conceptual representations when the object has been associated with a body-part for a long time (Experiment 1) but not when it has been recently associated (Experiment 2). Taken together these findings suggest that the experience of object ownership can indeed be based on the incorporation of objects in our body image.

References

Aglioti, S., Smania, N., Manfredi, M., & Berlucchi, G. (1996). Disownership of left hand and objects related to it in a patient with right brain damage. *Neuroreport*, 8(1), 293-296.

Belk, R.W. (1988). Possessions and the extended self. *Journal of Consumer Research*, 15, 139–68.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, 391(6669), 756.

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, 107(3), 978-998.

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn Sci*, 19(1), 6-12.

Nascimento Alves, P., et al., (2016). Loss of object ownership feeling following a left hemisphere infarct. *Cortex*, 84, 132-134.

Scorolli, C., Borghi,A. M., & Tummolini, L. (2017). Cues of control modulate the ascription of object ownership. *Psychological Research*, 1-26.

Sui, J., He, X., & Humphreys, G.W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *J Exp Psychol Human* 38, 1105.

Simulating Epistemic Bias in Academic Recruiting

Carlo Debernardi[1], Eleonora Priori[2], Marco Viola[2]
[1]University of Florence, [2]University of Turin

During the last two decades, several countries adopted performance-based fundings based on formal evaluation of research and of researchers. However, the implementation of such evaluations raised several concerns. Gillies (2014) expressed the worry that such evaluations might lead some mainstream school of thought to suppress other heterodox schools (as happened in UK economics according to Lee and colleagues 2013). Such monopoly entails excessive conformity and thus impedes fertile competition between different schools of thought.

Epistemic bias and pluralism

This drive toward undesirable conformity might be due to an epistemic bias (rather than on nepotism), that spur evaluators to favor researchers and research products akin to their own position over distant ones. Viola (2017) marshaled some evidence supporting the existence of this bias in all the following kinds of evaluations: peer review, bibliometrics, and hiring procedures. He also argued that formal evaluations problematically reinforce existing biases (cf. Bonaccorsi 2015). Moreover, even small biases in each of these procedures might severely compromise pluralism due to self-reinforcement loops. In particular, biases in hiring might reinforce biases in every evaluative activity, since those who are hired today will be the evaluators of tomorrow.

An Agent-Based Model proposal

To articulate this hypothesis, we built an agent-based simulation inspired to the academic recruitment in some scientific communities in Italy. In the last years hiring practices of Italian universities underwent several modifications: for instance, the National Scientific Habilitation was introduced. This system requires the favorable opinion of 4/5 members of a panel in order to qualify as a candidate for *concorsi* (selection procedures for future professors); recently, the Habilitation has been slightly reformed, now requires only the favorable opinion of 3/5 members of the panel.

In our model, implemented with NetLogo 6.0.4, we explore the potential pluralism-threatening effect of two sources in epistemic bias: first, a bias in the evaluations of recruiters themselves (*ceteris paribus*, sameness of school of thought implies higher scores); and/or, a bias in publication and citations (the more people there are within a school of thought, the easier will be to get published and cited). The model allows for several variables: background conditions (number of agents initially belonging to each school of thought, turnover rate); strength of the two epistemic biases; governance of recruitment.

Preliminary results

Preliminary results show how even small biases makes it likelier (and faster) that a heterodox school of thought gets wiped off in almost every condition. Unsurprisingly, this effect is more pronounced when the two sources of biases co-exist.

Another relevant result concerns the difference between the hiring procedures. We built an experimental settings with two rival schools of thought starting with the same number of professors. With only the evaluation bias on, we analysed 300 simulations, looking for the ones ending with two coexisting schools of thought. We found a greater proportion of this kind of runs with the Habilitation requiring 3 favorable opinions from the panel. The pre-Habilitation setting scored roughly in the middle, while the Habilitation requiring 4 favourable opinions seems to be the procedure which most likely produces convergence to conformity.

References

Bonaccorsi, A. (2015). *La valutazione possibile. Teoria e pratica nel mondo della ricerca*, Bologna: Il Mulino.

Gillies, D. (2014). Selecting applications for funding: why random choice is better than peer review. *RT. A Journal on research policy and evaluation, 2(1).*

Kuhn, T.S. (1962/1970). *The Structure of Scientific Revolutions*. University of Chicago Press.

Lakatos, I. (1970). Falsificationism and the methodology of research programmes. In Lakatos, I., & Musgrave, A. (Eds.). *Criticism and the Growth of Knowledge*. Cambridge University Press.

Lee, F. S., Pham, X., & Gu, G. (2013). The UK research assessment exercise and the narrowing of UK economics. *Cambridge Journal of Economics, 37(4), 693-717.*

Travis, G. D. L., & Collins, H. M. (1991). New light on old boys: Cognitive and institutional particularism in the peer review system. *Science, Technology, & Human Values*, 16(3), 322-341.

Viola, M. (2017). Evaluation of Research (ers) and its Threat to Epistemic Pluralisms. *European journal of analytic philosophy, 13(2), 55-78.*

Paranormal beliefs?

A matter of core knowledge confusion (that can be simulated by a computational model)

Sara Dellantonio[1], Luigi Pastore[2], Claudio Mulatti[3]
[1] Università di Trento; [2] Università di Bari, [3] Università di Padova

The research on the cognitive bias involved in the endorsement of irrational beliefs is as vast and difficult to define as the set of possible irrational beliefs (see Vyse, 2013). One subset of irrational beliefs is that composed by beliefs about paranormal events, i.e. – according to Broad's (1953) definition – events that violate the fundamental and scientifically founded principles of nature. This is an interesting definition, which – with reference to the violation of principles of nature – establishes a link to a specific cognitive bias called 'core knowledge confusion' (e.g., Lindeman & Aarnio, 2007). The aim of this contribution is to explore the link between core knowledge confusion and paranormal beliefs and to show why and how the former leads to the latter. Moreover, we will also point out that this form of irrationality can be explained within the framework of a semantic model.

Core knowledge confusion

The term 'core knowledge' refers to a narrow but fundamental set of beliefs about evolutionarily important phenomena that determine our intuitive understanding of physics, biology, and psychology. These beliefs are usually learned by children before starting school and largely without instruction (Rakison & Poulin-Dubois 2001; Wellman & Gelman 1998) and can be considered the backbone of our conceptual system, engendering, shaping, and constraining further conceptual development (Carey 2009; Wellman, Hickling & Schult 1997). Core knowledge includes information about the differing nature and behavior of objects belonging to distinct ontological domains, such as: physical phenomena are material and objective, while mental phenomena (e.g. thoughts, desires, imagination, and representations) are subjective and immaterial; animate beings can commit purposeful acts, inanimate entities cannot; physical objects have an independent existence in space and can move other objects by physical force, but mental states cannot; living organisms can grow and heal or die, lifeless objects cannot; access to the physical world is necessary for perception, but desires and emotions can arise mentally.

If these ontological domains are confused with each other and we attribute the properties or behaviors of objects in one domain to entities belonging to another, we develop specific kinds of irrational beliefs, and in particular beliefs about paranormal events, i.e. events that violate the fundamental and scientifically established principles of nature (Broad 1953). If, for example, the biological processes of contagion or healing are transposed onto the psychological domain, we may believe that we can either contaminate or heal somebody through thought. If the psychological and the physical ontological domains are confused with each other, we may think that the mind can exert a force on external objects and make them move (psychokinesis) or that it can transfer thoughts from one mind to another as if they were objects (telepathy). Or we might believe that the position of the celestial bodies at birth determines the basic psychological characteristics of people (astrology) or that knowing future events is a purely mental process (precognition). Many other cases of irrational belief have been shown to be at least indirectly related to core knowledge confusion. Importantly, it has been shown that not only people who believe in horoscopes, but also individuals who believe in superstitions, conspiracy theories and precognition obtain high scores on tests that aim to evaluate core knowledge confusion (e.g. Lindeman & Aarnio 2007; Lindeman, Svedholm-Hakkinen & Lipsanen 2015). Thus, it is possible that core knowledge confusion can explain a wide number of irrational forms of reasoning.

A semantic model for paranormal beliefs

If we consider semantic knowledge as a vertically organized taxonomy with three dominant conceptual nodes – i.e. (intuitive or folk) physics, (intuitive or folk) biology, and (intuitive or folk) psychology – then ontological confusions can be attributed to sets of links that horizontally traverse the taxonomy to connect representations that *per se* are located 'far away' from each other in semantic space. The hypothesis we suggest here is that – at least in some cases – associating semantically 'far' representations might result in ontological confusion and thus in accepting some kind of irrational and specifically paranormal belief. Here we are not going to investigate how this semantic model could be structured (e.g., how the conceptual nodes are related to each other under normal conditions). We rather point out that the development of such a model would allow us – at least in principle – to design a computer simulation of the way in which we develop paranormal beliefs.

Our aim is to clarify a preliminary issue concerning the links that horizontally traverse the taxonomy to connect representations located 'far away' from each other in semantic space. Specifically, we will investigate how core knowledge confusion occurs and how it modifies the original semantic structure.

Conclusive remarks

We are hopeful that our analysis will help to clarify the issue of paranormal beliefs at two levels. On the one hand, by showing that core knowledge confusion offers a psychological explanation for how at least some paranormal beliefs originate. On the other hand, by presenting a semantic model that accounts for how core knowledge confusion concretely occurs.

References

Broad, C. D. (1953). *Religion, philosophy, and psychical research*. New York: Harcourt & Brace.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Lindeman, M. & Aarnio, K. (2007). Superstitious, magical, and paranormal beliefs: An integrative model. *Journal of Research in Personality*. 41(4): 741-744.

Lindeman, M., Svedholm-Hakkinen, A. M. & Lipsanen, J. (2015). Ontological confusions but not mentalizing abilities predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition.* 134: 63-76.

Rakison, D. H. & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*. 127(2): 209-228.

Vyse, S.A. (2013). *Believing in magic. The psychology of superstition*. Oxford: Oxford University Press.

Wellman, H. M., Hickling, A. K. & Schult, C. A. (1997). Young children's psychological, physical and biological explanations. *New Directions for Child and Adolescent Development*. 1997(75): 7-25.

Wellman, H. M. & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In W. Damon (ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 523-573). Hoboken, NJ, US: John Wiley & Sons Inc.

Evaluating fairness in 5-year-old children in a Child-Robot Interaction

Di Dio[1*], C., Manzi[1,5*], F., Itakura[2], S., Kanda[3], T., Ishiguro[3,4], H., Massaro[1], D., Marchetti[1], A.
[1]Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy
[2]Department of Psychology, Graduate School of Letters, Kyoto University, Kyoto, Japan
[3]Advanced Telecommunications Research Institute International Hiroshi Ishiguro Laboratories and Intelligent Robotics and Communication Laboratories, Kyoto, Japan
[4] Department of Adaptive Machine System, Osaka University, Toyonaka, Japan
[5] Institute of Psychology and Education, Neuchâtel University, Neuchâtel, Switzerland

Introduction

The relationship between humans and robots is increasingly becoming focus of interest for many research fields. In psychology, most studies investigating the Human-Robot Interaction (HRI) in adults have shown that humans behave similarly when interacting with another human or a robot (Kanda et al., 2004; Itakura et al., 2008a, b; Nishio et al., 2012), demonstrating how both physical and psychological features of the robots are important for the implementation of HRI (for a review, see Marchetti et al., 2018). By contributing to the identification of the precursors underlying human sociality and their ontogenesis, knowledge from developmental psychology has also proved to be central to the implementation of psychological models that can be used in HRI. In this respect, Itakura (2008) proposed a new research domain, Developmental Cybernetics, which hypothesizes that children understand non-human agents through mentalization (Manzi et al., 2017). With particular focus on the Theory of Mind (ToM), a recent investigation (Di Dio et al.) used the Ultimatum Game (UG) - that is related to the development of ToM in that it activates the ability of mental comprehension, detection and anticipation of the other's behavior (e.g., Castelli et al., 2010, 2014a,b, 2017) – to evaluate preschoolers' decision making when interacting with either another child or with a robot. During the UG, children played as both proponents and receivers. The aim of the study was to evaluate the mental characteristics that children ascribe to robots with respect to humans, comparing mental attributions with actual behavior during an interactive situation that requires reciprocity, such as that proposed in the UG. In so doing, children's justifications for their behavior during the UG were also analyzed when playing with the human and the robot as proposers and receivers.

Methods

Thirty-one (31) Italian kindergarten children participated in the experiment (13 females; mean age = 70.8 months, SD=2.99 months). The following tests were used: a) control tests: the PPVT, the subtest inhibition of NEPSY-II, the Backward Digit Span; b) ToM tests: Smarties and Sally-Anne tasks (classic storyboard version); first false belief task (novel video version); c) Attribution of Mental and Physical States scale (AMPS); d) the Ultimatum Game, played with a human and a robot as both receiver and proposer.

Results

In relation to AMPS, the GLM highlighted that children attributed higher mental and perceptive properties to the human than to the robot (Figure 1). On the other hand, comparison between the total amount offered and received when children played the UG with the human and the robot revealed no significant differences. Additionally, a chi-square analysis of the justifications at the UG coded as outcome, equity, and mentalization-based, showed that outcome-based justifications were the most used for all divisions when playing with both the human and the robotic partner. Equity and mentalization-based justifications were used for fair divisions independent of agency, substantially decreasing when children proposed or received a divisional offer that skewed towards iniquity.

Considerations

The AMPS showed that 5-year-old children recognized the robot as a different entity from the human in line with previous literature (e.g., Katayama et al., 2010). Nevertheless, children behaved similarly during the UG when playing with the human and the robot as either proposers or receivers, confirming a tendency to maximize gain (e.g., Fehr et al., 2008), and to partially disregard the other's "mind" when making "economic" choices. This idea was further reinforced by children's justifications for their behavior at the UG showing that, independent of the interactive agent and of the role played, at this age children mostly reason in economic/quantitative terms when playing, besides with human partners, also with robots. Most interestingly, nearly total absence of equity and mentalization-based justifications for unfair divisions suggests an attempt of children to resolve a social cognitive conflict emerging from the discrepancy between a "socially-expected" fair behavior and the actual "selfish" behavior. This interpretation - opposite to what observed above - underpins children's consideration of the other's "mind" and social judgment (Takagishi et al., 2010). Lack of major differences between the human and robotic partner also in this context corroborates the idea that the robotic partner is not treated, by young children, differently from the human partner.

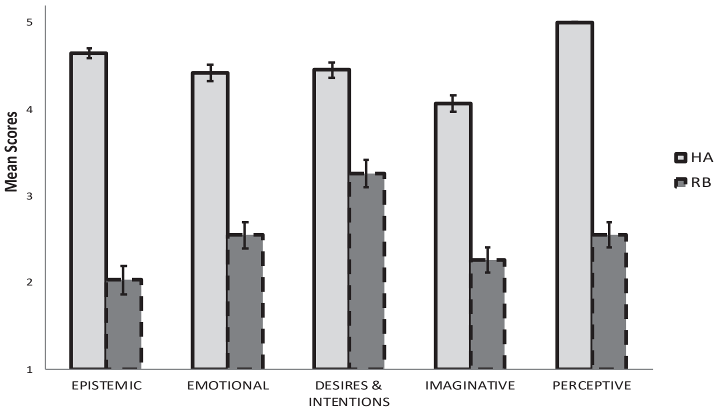


Figure 1: AMPS scores as a function of agency.

References

Castelli, I., Massaro, D., Sanfey, A.G., & Marchetti, A. (2014a). "What is fair for you?" Judgments and decisions about fairness and Theory of Mind. *Eur J of Developmental Psychology, 11*(1), 49-62.

Castelli, I., Massaro, D., Bicchieri, C., Chavez, A., & Marchetti, A. (2014b). Fairness norms and theory of mind in an ultimatum game: judgments, offers, and decisions in school-aged children. *PloS one, 9*(8), e105024.

Di Dio, C., Manzi, F., Itakura, I., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (under review). It does not matter who you are: fairness in preschoolers interacting with human and robotic partners. *International Journal of Social Robotics*.

Fehr, E., Bernhard, H. & Rockenbach, B. (2008). Egalitarianism in young children. *Nature,454*, 1079.

Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008a). How to Build an Intentional Android: Infants' Imitation of a Robot's Goal-Directed Actions. *Infancy, 13*(5), 519-532.

Itakura, S., Okanda, M., & Moriguchi, Y. (2008b). *Discovering Mind: Development of Mentalizing in Human Children*. In S. Itakura, K. Fujita (eds.), Origins of the Social Mind: Evolutionary and Developmental Views, 179-198. Springer Japan: Tokyo.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *International Journal of Human Computer Interaction, 19*(1), 61-84.

Katayama, N., Katayama, J.I., Kitazaki, M., & Itakura, S. (2010). Young Children's Folk Knowledge of Robots. *Asian Culture and History, 2*(2), 111.

Manzi, F., Massaro, D., Kanda, T., Tomita, K., Itakura, S., & Marchetti, A. (2017). Teoria della Mente, bambini e robot: l'attribuzione di stati mentali, *XXX Congresso AIP Sezione di Psicologia dello Sviluppo e dell'Educazione*. Messina, Italy: Alpes Italia srl.

Nishio, S., Ogawa, K., Kanakogi, Y., Itakura, S., & Ishiguro, H. (2012) Do robot appearance and speech affect people's attitude? Evaluation through the Ultimatum Game, *IEEE RO-MAN* (pp. 809-814). Paris, France: IEEE.

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., & Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of experimental child psychology, 105*(1-2), 130-137.

Ad hoc presupposition construction

Filippo Domaneschi[1] and Diana Mazzarella[2]
[1]*University of Genoa,* [2]*University of Neuchâtel*

Philosophers and linguists have long debated the phenomenon of 'presupposition', whereby speakers linguistically mark information as background or taken for granted. This debate has surprisingly left untouched the following two assumptions: (i) presuppositions are not part of what is *intentionally* communicated by the speaker, and (ii) the propositional content of a presupposition is semantically determined. These assumptions are well-established in the literature and they are endorsed even by scholars who offer a 'pragmatic' analysis of presuppositions. For instance, with regard to (i), Simons (2007) claims that presuppositions are "not part of the speaker's communicative intention" and they are transmitted simply as a "by-product". Or, with regard to (ii), she suggests that the presuppositional content is "calculated, presumably compositionally, on the basis of the content of the trigger plus the rest of the content expressed" (Simons, 2005).

In this paper, we challenge both these assumptions. We argue that presuppositions fall within the scope of 'ostensive-inferential communication' (Sperber & Wilson, 1986/1995) and illustrate the benefits of our proposal. Ostensive-inferential communication requires the communicator to display an overt behavior aimed at attracting attention to her communicative intention ('ostension'). Furthermore, it requires the audience to infer the communicator's intended meaning ('inference').

First, we show that by treating presuppositions as part of what is *ostensively* communicated by the speaker, we can provide a unified account of a variety of presuppositional uses discussed in the literature. This account covers the whole range of cases from noninformative, or 'common ground' presuppositions, to informative presuppositions, as well as cases of exploitative presuppositions.

Second, and more importantly, we suggest that by treating presuppositions as the output of an *inferential* process of pragmatic interpretation, we can shed a new light on their context-sensitivity. Traditional accounts of presuppositions view the context-sensitivity of (at least some) presuppositions as pertaining to the question of whether presuppositions are contextually defeasible. This question does not exhaust the relation between presuppositions and context. The propositional content of a presupposition is determined *ad hoc* via a process involving both semantic decoding and pragmatic inference, and involving a 'mutual parallel adjustment' (Wilson & Sperber, 2004) among presuppositions, the explicit content and the implicatures of the utterance.

Ostension and inference represent two sides of the same coin and can shed a complementary light on presupposition as a genuinely *communicative* phenomenon.

References

Simons, M. (2005). Presuppositions and relevance. In: Szabo, Z. (Ed.), *Semantics Versus Pragmatics*. (pp. 329-355). Oxford: Oxford University Press.

Simons, M. (2007). Presupposition and cooperation. Carnagy Mallon University, ms, https://www.cmu.edu/dietrich/philosophy/docs/simons/Presupposition%20and%20cooperation.non-anonymous.pdf

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Wilson, D., & Sperber, D. (2004). Relevance Theory. In L. R. Horn & G. L. Ward (Eds.), *Handbook of Pragmatics* (pp. 607–632). Oxford: Blackwell Publishing.

Beyond technological instrumentalism

Marco Fasoli
University of Milano Bicocca

"Technology is neutral, it has no ends of its own and exists only to accomplish human ends". This common-sense belief is often mentioned in the public debate, in order to attribute the entire responsibility of some human actions to the people. In the philosophy of technology, similar claims are usually defended by the so called "technological instrumentalism" (Pitt 2014). According to technological instrumentalism, technology is simply a mean and we are totally free to use any way we want. Conversely, according to "technological determinism" technology somehow determines how we use it. In this paper, I will analyze the starting claim and I will point out that it actually conflates two different theses. The first thesis is that technology cannot embody values, while the second one is that how an artifact is built does not influence how we employ it. Insofar as we usually don't' confer values to physical objects but to actions, choices, thoughts, etc., I will argue that the first thesis seems sound. Conferring values to objects would require the development of a counterintuitive notion of value, but we actually don't have some reasons to do this. On the other hand, I will point out that the second thesis is based on a naive conception of the human rationality. In order to understand how technology may foster some behaviors instead of others, we need to assume a more sophisticated model of the human mind as well as a more realistic conception of artifacts. In the last thirty years, cognitive sciences deeply revised our intuitions about the functioning of our mind, identifying a large spectrum of heuristics and biases (see Kahneman 2011). Furthermore, artifacts have affordances and affordances offer us specific repertoires of actions (Casati 2017, Fasoli 2018, Latour 1994). This change of paradigm looks similar to the transition from the classical economy to the behavioral economy (Thaler & Sunstein 1999). I will argue it would also have similar consequences, helping us to achieve a deeper and more realistic understanding of modern technologies.

References

Casati, R. (2017). Two, then Four Modes of functioning of the mind: towards an unification of "dual" theories of reasoning and theories of cognitive artifacts. In *Representations in Mind and World*, ed. J. Zacks and H. Taylor, 7-23. Routlege.

Fasoli, M. (2018). Super Artifacts: Personal Devices as Intrinsically Multifunctional, Meta-representational Artifacts with a Highly Variable Structure. *Minds and Machines*, 28(3), 589-604.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kline, N. (1959). The Challenge of the Psychopharmaceuticals. *Proceedings of the American Philosophical Society* 103,3: 455-462.

Kroes, P. and P.-P. Verbeek. (2014). *The Moral Status of Technical Artefacts*. Dordrecht, The Netherlands: Springer.

Latour, B. (1994). On Technical Mediation. *Common Knowledge* 3.2: 29-64.

Norman, D. (1993). *Things That Make Us Smart: Defending Human Attribute in the Age of Machine*. New York: Basic Books.

Pitt, J. (2014). 'Guns Don't Kill, People Kill'; Values in and/or Around Technologies. In *The Moral Status of Technical Artefacts*, ed. Kroes, Peter and Peter-Paul Verbeek, Dordrecht, 89-102, The Netherlands: Springer.

Thaler, R. H., & Sunstein, C. R. (1999). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT Yales University Press.

Francis Crick and the Hard Problem of Consciousness

Francesco Ellia[1]
[1] *Università di Bologna*

Francis Crick and the Hard Problem of Consciousness

It is remarkable to notice today that until the last decade of the last century both psychology and neuroscience tended to avoid any reference to consciousness. There were at least two reasons that motivated this practice, either because the problem of consciousness was considered a philosophical problem and so best left to philosophers, or because scientists conceded that it was a scientific problem, but science did not reach the maturity to face it yet. This changed over the last thirty years, during which there has been a growing interest in consciousness among the scientific community at many levels, including but not limited to: psychology, neuroscience, evolutionary biology and computer science. One of the events that legitimated consciousness as a matter worth of scientific investigation was that, among the others, Francis Crick devoted the last part of his career to study consciousness from a scientific standpoint. He set a research program based on the search for the neural correlates of consciousness (NCC) and presented his own theoretical model. His ideas were collected in his book *The Astonishing Hypothesis* (1994) and a series of subsequent articles, in several occasions co-authored by Christof Koch.

The scientific investigation of consciousness shares part of its story with one of the oldest philosophical debate: the so-called mind-body problem. In particular, whether or not consciousness can be fully explained scientifically has been long debated by philosophers. In his seminal book *The Conscious Mind*, David Chalmers named this problem the Hard Problem of consciousness. In fact, he distinguished between two different sets of problem: that the scientific investigation of consciousness poses. He called 'easy' (in a technical sense) problems those that are to explain the objective functions associated with consciousness in physical terms by specifying the neural mechanism that perform the functions. The easy problems may still represent a complex and difficult challenge for science, but at a conceptual level they fit very well in a theoretical framework we are familiar with. On the other hand, the Hard Problem does not, as it consists in to explain why and how these physical mechanisms are accompanied by subjective experience.

In this contribution I will first introduce Chalmers definition of hard problem and easy problems of consciousness. Then I will summarize Crick's theory of consciousness as presented in *The Astonishing Hypothesis.* I will argue then, that according to Chalmers definition of the hard problem, Crick's theory faces some of the easy problems of consciousness but does not attempt to engage with the hard problem. I will argue that one of the issues with Crick's theoretical approach is the lack of clarity in the definition of the phenomenon he attempts to explain.

References

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200-219.
Chalmers, D. (1996). *The Conscious Mind*. Oxford University Press.
Chalmers, D. (2000): What is a Neural Correlate of Consciousness? In *Neural Correlates of Consciousness*, edited by Thomas Metzinger, MIT Press.
Crick, F. (1994): *The Astonishing Hypothesis*. New York: Simon & Schuster.
Crick, F. and Koch, C. (1990): Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences,* 2, 263-275.
Crick, F. and Koch, C. (1995): Are We Aware of Neural Activity in Primary Visual Cortex? *Nature,* 375, 121-23.

Crick, F. and Koch, C. (1998): Consciousness and Neuroscience, *Cerebral Cortex* 8, 97-107.
Favela, L. (2006): *Biological Theories of Consciousness: the search for experience*. Master's thesis, San Diego State University.

On the use of agent-based simulation for cognitive science

Silvia Felletti[1,2], Paolo Pagliuca[1,3], Alessandro Sapienza[1,4], Daniele Vilone[1,5]
[1]*Istituto di Scienze e Tecnologie della Cognizione, CNR, Roma*

[2]*silvia.felletti@istc.cnr.it;* [3]*paolo.pagliuca@istc.cnr.it;*
[4]*alessandro.sapienza@istc.cnr.it;* [5]*daniele.vilone@istc.cnr.it*

1. Introduction

From the 1990s onward, the growth of computational capacity applied to research has brought social scientists to use advanced techniques to simulate and analyze agent interactions in social structures.
Agent-based models (ABMs) reflect relationships and interactions in the real world, and can be used to investigate social dynamics in populations ranging from relatively small (e.g., institutions and organizations dynamics) to global-size (e.g., migrations, demographic changes; see Squazzoni, 2012).
An ABM define agents' behavior, establishes the connections between them, sets the environmental variables, and runs simulations. The result is statistical data that can guide empirical research and validate theoretical models on human behavior and interaction.
In a similar vein, but with different scopes and potential applications, *evolutionary robotics* (ER) has been successfully applied in recent years on a variety of social, psychological and biological phenomena such as spatial navigation, risk attitudes, cooperation (Nolfi and Floreano, 2000). This technique allows to evolve populations of simulated robots (with a more or less complex neural network) in an environment which poses specific pressures (e.g. escaping from predators or searching for food), and simulate natural evolution (using mechanisms such as fitness, selection, random mutation) to observe what behavioral solutions are developed by the evolved population of individuals.
Besides providing empirical support to theoretical models, these computational techniques also offer some advantages over the experimental approaches typical of the behavioral sciences, particularly when the object of research is the evolutionary roots of a given behavior or phenomenon. They allow full observability, keeping track of all the data regarding agents and environment, and offer full control of the variables involved, without, however, being deterministic in their response (thus being adherent to real-world phenomena). They also present some limitations, due to their simplified and abstracted nature (see Paglieri et al., 2015).
In this symposium we will present three studies where ER or ABM were applied to psychological and social questions: risk attitudes, trust, and cooperation.

2. Risk attitudes in foraging animals: an evolutionary robotics approach Silvia Felletti and Paolo Pagliuca, ISTC-CNR

The study of risk attitudes in foraging animals (i.e., their preferences between fixed and variable foraging sources, when the expected value of the variable option is higher) has been dominated by Stephens' (1981) model, for which risk preferences will depend on *energy budget*, so that individuals with a low budget will choose the fixed option only if it is enough to take them above the threshold of starvation; otherwise they will switch to the risky option.
Evidence for this switch has been widely observed in smaller species (insects, small birds), which easily find themselves at risk of starvation. However, larger-bodied species' preferences seem to be more likely affected by diet quality. For example, chimpanzees are usually prone to choose riskier foraging strategies (hunting) when fruit is available and abundant, making them able to compensate for the energetic costs of hunting (Gilby & Wrangham, 2007).
Gilby and Wrangham argued that the scarcity/risk and the abundance/risk hypotheses are not necessarily mutually exclusive, and that animals (both small and large-bodied) could theoretically

be risk-prone when food is either very scarce or superabundant, with risk-proneness following a U-shaped function peaking when diet quality is either very high and very low.
Such hypothesis would be difficult to experimentally prove, in both orders of species. Particularly, inducing a near-starvation condition in bigger species would present practical and especially ethical obstacles.
In this talk we will illustrate how we applied evolutionary robotics to overcome these difficulties. We evolved and tested populations of individuals in different environments, and manipulated variables such as the capacity of individuals' energy storage, their metabolic rate (i.e., how fast they consume energy), the energetic value of food and its abundance in the environment. The aim of this study is to observe the effects of such variables on the emergence of different risk attitudes. In particular, we were interested in the following issues:

- Are both scarcity/risk and abundance/risk strategies adopted by small and large-bodied species, when in the same condition (starvation/abundancy)?
- Did the strategies evolve as a consequence of the environment characteristics or the size and metabolism of the species? What is the most determining factor?
- How would individuals evolved in abundant environments behave in scarcity (and vice versa)?

Preliminary results will be discussed.

3. The role of recommendation within the digitally infrastructured society
Alessandro Sapienza (speaker) and Rino Falcone, ISTC-CNR

The growth of the digital infrastructure has seen a significant part of our life moving on the web, giving us an enormous communicative possibility and enabling us to reach a very high number of individuals, which in turn requires reshaping the way we relate to each other.
In this talk I will focus on the use of individuals' recommendations, a necessary tool in every social environment, emphasizing the need for the identification of new specific methodologies in this infrastructure. In fact, because of our very nature, we are prone to generalize a methodology that works efficiently in a given context to others. But the classical approaches may not work in new contexts. We are interested in showing that:

1. It is not possible to apply this kind of recommendation in digital social context in the same way we do in real social context.
2. For trust propagation, it is fundamental to evaluate a node as recommender for a specific task, and not with respect to its ability on executing the same task, which is a common approach in literature.

The study we present involved the verification of our assumptions by the means of an agent-based simulation and a trust model based on Castelfranchi & Falcone (2010).
Consistently with our predictions, we proved that, receiving an equal quality and quantity of information, what makes the difference is the way we process this information.
We discovered that querying a huge percentage of the agents in the network, although including even the less reliable ones, allows to obtain better performance, rather than restricting the query to the most reliable agents, a the small portion of the network. In fact, the structure of a distributed network impose physical limits on an individual’s knowledge.
In a further experiment, we analyzed the fallacy of confusing the trustworthiness of an agent concerning its ability as a recommender and the one describing how good it is in performing the requested task. Our results suggest that using the second in place of the first leads to a highly unreliable result.

4. Hierarchical invasion of cooperation in complex networks
Speaker: Daniele Vilone

The emergence and survival of cooperation is one of the hardest problems still open in science, and in particular in Social Sciences because cooperation among humans involves non-trivial cognitive patterns (Smith & Szathmáry, 1995).

Several factors such as the existence of punishment, repeated interactions, topological effects and the formation of prestige may all contribute to explain the counter-intuitive prevalence of cooperation in natural and social systems (Rand & Nowak, 2013). The characteristics of the interaction networks have been also signaled as an element favoring the persistence of cooperators. In this talk I will consider the invasion dynamics of cooperative behaviors in complex topologies. The invasion of a heterogeneous network fully occupied by defectors is performed starting from nodes with a given number of connections (degree) K. The system is then evolved within a Prisoner s Dilemma game and the outcome is analyzed as a function of K and the degree of the ' nodes adopting cooperation. Carried out using both numerical and analytical approaches, the results I will present show that the invasion proceeds following preferentially a hierarchical order in the nodes from those with higher degree to those with lower degree; however, the invasion of cooperation will succeed only when the initial cooperators are numerous enough to form a cluster from which cooperation can spread: this implies that the initial condition has to be a suitable equilibrium between high degree and high numerosity (Vilone, Capraro, & Ramasco, 2018). These findings have potential applications to the problem of promoting pro-social behaviors in complex networks.

5. *References*

Castelfranchi, C., & Falcone, R. (2010). Trust theory: A socio-cognitive and computational model (Vol. 18). John Wiley & Sons.

Gilby, I. C., & Wrangham, R. W. (2007). Risk-prone hunting by chimpanzees (Pan troglodytes schweinfurthii) increases during periods of high diet quality. Behavioral Ecology and Sociobiology, 61(11), 1771-1779.

Nolfi, S., Floreano, D., & Floreano, D. D. (2000). Evolutionary robotics: The biology, intelligence, and technology of self-organizing machines. MIT press.

Paglieri, F., Parisi, D., Patacchiola, M., & Petrosino, G. (2015). Investigating intertemporal choice through experimental evolutionary robotics. Behavioural processes, 115, 1-18.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. Trends in cognitive sciences, 17(8), 413-425.

Smith, J. M., & Szathmary, E. (1997). The major transitions in evolution. Oxford University Press.

Squazzoni, F. (2012). Agent-based computational sociology. John Wiley & Sons.

Stephens, D. W. (1981). The logic of risk-sensitive foraging preferences.

Vilone, D., Capraro, V., & Ramasco, J. J. (2018). Hierarchical invasion of cooperation in complex networks. Journal of Physics Communications, 2(2), 025019.

The Motoric Component of Pictorial Experience

Gabriele Ferretti,
University of Florence

This paper suggests that motor processing plays an important role in our experience of pictures. The standard literature on picture perception has never explicitly addressed this idea, as depicted objects are taken to be passive objects of perception: they do not foster in us the perception of action possibilities (Matthen 2005; Nanay 2011; Ferretti 2016a, 2016b, 2017a, 2017b, 2018, forthcoming) and we cannot perceive significant spatial shifts when moving with respect to them (Nanay 2011; Hopkins 2012). For this reason, it might sound extremely controversial to say that motor processing plays an important role in generating pictorial experience.

Contrary to this general stance, this paper defends, for the first time, this apparently very controversial claim, never addressed in the literature: that the specific and crucial relations between vision and motor processing are crucially involved in our perception of objects in a picture.

The reader should note that a closer look to this claim reveals that it should not be considered as a very odd claim. Indeed, the philosophical analysis of several studies from cognitive science suggests that vision and motor processing are deeply linked (Nanay 2013; Noë 2004; Briscoe and Grush 2015; Ferretti 2016c; Zipoli Caiani and Ferretti 2017; Ferretti and Zipoli Caiani 2018). But we also know that the representations by means of which we can access a pictorial content are *visual* representations (Nanay 2011, 2012, 2017; Ferretti 2016, 2017a, 2017b, 2018, forthcoming). This suggests that motor processing might be involved in picture perception in a non-trivial manner. However, nobody has ever explicitly discussed the role of motor processing in our experience of pictures.

Starting from this lack in the literature, I first discuss two ways in which vision and motor processing are crucially linked. First of all, (i) vision allows us to visually represent the action possibilities offered by the objects we face in our environment. Second, (ii) as we move with respect to an object, we experience a change in sensory stimulation that is correlated to the way such a movement is performed. These two ways in which vision and motor processing are crucially linked are taken to be at the basis of our everyday perception of real objects. I suggest that they also play a crucial role in order for us to obtain a pictorial content.

I first specify that the literature on picture perception suggests that, when we are in front of a picture, we visually represent two things: the vehicle that bears the pictorial content, i.e. the picture's surface and the pictorial content itself, i.e. the depicted object. The literature also suggests that the way we visually represent the surface can influence the way we obtain the pictorial content.

At this point, I show that the crucial relations between vision and motor processing, described by (i) and (ii), are at the basis of the representations we use in picture perception. In particular, I describe how they are respectively in play during the representation of the picture's surface, as well as during the representation of the depicted object.

In order to strengthen my point, I also show that, when we cannot properly exercise the two ways in which vision and motor processing are crucially linked, described by (i) and (ii), respectively, during the representation of the picture's surface, as well as during the representation of the depicted object, this leads to a breakdown in pictorial experience, of the kind obtained with pictorial illusions *à la trompe l'oeil*, which are able to foster in us the (illusory) visual experience of real presence, which resembles, almost perfectly, the one we are (veridically) acquainted with during real object perception.

References

Briscoe R. and 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/.

Ferretti, G. (2016a) Visual feeling of presence. *Pacific Philosophical Quarterly*, DOI: 10.1111/papq.12170.

Ferretti, G. (2016b): 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. (2016c). Through the Forest of Motor Representations. *Consciousness and Cognition*, 43, 177-196. DOI: http://dx.doi.org/10.1016/j.concog.2016.05.013.

Ferretti, G. (2017a). Are Pictures Peculiar Objects of Perception? *Journal of the American Philosophical Association*, 3(3), 372–393. DOI: 10.1017/apa.2017.28.

Ferretti, G. (2017b). Pictures, Emotions, and the Dorsal/Ventral account of Picture Perception. *Review of Philosophy and Psychology*, 8(3), 595-616. DOI: 10.1007/s13164-017-0330-y.

Ferretti, G. (2018). The Neural Dynamics of Seeing-In, *Erkenntnis*, DOI: https://doi.org/10.1007/s10670-018-0060-2.

Ferretti, G. (Forthcoming). Perceiving Surfaces (and What They Depict). In *The Senses and the History of Philosophy,* Edited by Brian Glenney and José Filipe Silva, Routledge.

Ferretti, G. and Zipoli Caiani, S. (2018). Solving the Interface Problem without Translation: the Same Format Thesis. *Pacific Philosophical Quarterly*. DOI: 10.1111/papq.12243.

Hopkins, R. (2012) Seeing-in and Seeming to See. *Analysis*, 72, 650–59.

Matthen, M. (2005) *Seeing, Doing and Knowing: A Philosophical Theory of Sense Perception*. Oxford: Oxford University Press.

Nanay, B. (2011) Perceiving Pictures. *Phenomenology and the Cognitive Sciences*, 10, 461–80.

Nanay, B. (2012). The philosophical implications of the Perky experiments. *Analysis*, 72, 439-443.

Nanay, B. (2013). *Between perception and action*. Oxford: Oxford University Press.

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

Noë, A. (2004). *Action in perception*. Cambridge: The MIT Press.

Zipoli Caiani, S. and Ferretti, G. (2017). Semantic and Pragmatic Integration in Vision for Action. *Consciousness and Cognition,* 48, 40–54. DOI: http://dx.doi.org/10.1016/j.concog.2016.10.009.

Un modello induttivo-analogico della diagnosi nosologica

Francesco Gagliardi
Independent Scholar, Comitato Direttivo dell'AISC (Associazione Italiana di Scienze Cognitive)
ORCID: 0000-0002-4270-1636, fnc.research@gmail.com

Lo scopo di questo lavoro è fornire una spiegazione filosofico-cognitiva della diagnosi nosologica legando i classici temi filosofici delle forme del ragionamento scientifico (come induzione, deduzione, abduzione, analogia) con gli aspetti cognitivi del ragionamento di senso comune (come la categorizzazione, i prototipi, le argomentazioni). In filosofia della medicina sono state proposte varie teorie della diagnosi per analizzare e spiegare il ragionamento diagnostico; una di queste è la diagnosi nosologica: un particolare tipo di diagnosi a-teorica e basata sulla similarità. In questo lavoro mostriamo come la diagnosi nosologica si possa considerare come un procedimento scientifico composto sia da processi induttivi di categorizzazione e concettualizzazione che di argomentazioni per analogia; i processi di categorizzazione coinvolti sono spiegabili usando due delle teorie dei concetti proposte nella scienza cognitiva: la teoria dei prototipi e la teoria degli esemplari; mentre le argomentazioni usate sono analogie sia simmetriche che anti-simmetriche. Infine, presentiamo un modello della diagnosi nosologica che lega in un quadro coerente queste precedenti analisi; in questo modello la diagnosi nosologica basata sul riconoscimento del quadro morboso tipico, la sindrome, è spiegata in termini di teoria dei prototipi e di argomentazioni analogiche anti-simmetriche; mentre la diagnosi nosologica basata su di un precedente caso clinico è spiegata con la teoria degli esemplari e con le usuali argomentazioni analogiche simmetriche.

References

Gagliardi, F. (2018) Argomentazioni analogiche e processi di categorizzazione nella diagnosi nosologica. *Rivista Internazionale di Filosofia e Psicologia*, 9(2):151-161. DOI:10.4453/rifp.2018.0016

Nuove simulazioni per il cervello: quale metafora per la mente?

Elisabetta Gola[1]
[1]*Università di Cagliari*

1. La nascita della scienza cognitiva e di nuovi modi di pensare al mentale

La scienza cognitiva nasce negli anni Sessanta in un contesto in cui informatica, linguistica, antropologia, neuroscienza, filosofia, psicologia si erano allineate quasi magicamente sull'intendere gli oggetti delle loro analisi legati da un filo conduttore comune: gli aspetti strutturali e procedurali rappresentati in forme simboliche avevano consentito di aprire nuove prospettive teoriche sul mentale, prospettive di cui il comportamentismo aveva decretato la non indagabilità nel decennio precedente (Gardner, 1987).

La filosofia da sempre si interroga su quali siano i principi di funzionamento della mente, dei concetti, delle rappresentazioni, dei processi percettivi e di astrazione, ma con la nascita dei tentativi di applicazione del metodo scientifico alle discipline umanistiche e la nascita, in particolare, della psicologia scientifica, i vincoli metodologici tipici delle scienze naturali portano a mettere in discussione la possibilità di conoscere oggetti che non si possono osservare, misurare, sperimentare. Gli approcci strutturalisti continuavano a proporre come oggetto d'analisi strutture che potevano essere solo immaginate (per esempio un contenuto mentale, esaminabile con l'introspezione). Parole quali istinto, sentimento e i termini per descriverne le varie specie, continuavano ad essere usate. Fu Watson, nel 1913, col suo testo "Psychology as the behaviorist views it" a sostenere che l'oggetto della psicologia è il comportamento, l'unico fenomeno indagabile perché decomponibile in parti indagabili scientificamente. Il comportamento può infatti essere esaminato come risposta dell'organismo agli stimoli, ed entrambi (stimolo S e risposta, R) possono essere oggetto di sperimentazione, validazione di ipotesi, soggetti ad esperimenti ripetibili. Il manifesto di Watson fu consolidato nel tempo da studiosi quali Skinner e Thorndike e il comportamentismo divenne la corrente mainstream in psicologia.

La mente, il cui funzionamento non è possibile conoscere, viene considerata una scatola nera per la sua capacità di elaborare informazioni, ma senza che sia possibile conoscere tale processo.

Se pure era vero che non se ne poteva dire "scientificamente" nulla rispetto agli esseri umani, lo schema Stimolo-Scatola nera-Risposta descriveva il processo comportamentale dei calcolatori che negli anni '60 non erano più solo un modello matematico, ma cominciavano ad essere dei sistemi artificiali dotati di comportamento: lo stimolo era un ordine espresso con i sistemi di input, la risposta un'azione che il computer eseguiva (stampare, chiudere un file, aprire un file, etc.). E la scatola nera? Tutto il lavoro di elaborazione dei componenti 'interni' del calcolatore: unità logica, memoria, processore, etc. Frutto di tecnologie ingegneristiche, i processi di elaborazione simbolica dei calcolatori non avevano lo stesso difetto degli analoghi processi di cui si supponeva l'esistenza nella mente umana. Verosimilmente, se i calcolatori avevano un sistema di elaborazione simbolico che consentiva comportamenti complessi, come l'Intelligenza Artificiale cominciava a mostrare, anche gli esseri umani, nell'ipotesi peggiore, dovevano avere una struttura analoga.

2. Metafore concettuali

L'isomorfismo tra processi computazionali e processi mentali era la chiave interpretativa più naturale in quel momento storico e da un punto di vista epistemologico trovava supporto nell'idea che i modelli teorici avessero natura metaforica, specialmente nelle fasi iniziali della scoperta di una teoria scientifica (Black, 1962).

La nascita della scienza cognitiva coincise così con l'emergere di una metafora che ha dominato gli studi sui processi cognitivi, arrivando sino ai giorni nostri: la metafora della mente come calcolatore (Gozzi, 1989, 1991).

Si tratta di una metafora concettuale (Lakoff e Johnson, 1980), ossia una proiezione delle conoscenze relative a un dominio di ispirazione (source domain) a un dominio di arrivo (target domain). Nella maggior parte dei casi il dominio di ispirazione è più concreto e/o più conosciuto e serve proprio per catturare le proprietà (più astratte e/o meno note) del dominio di arrivo. Le conoscenze sul sistema solare sono state così utilizzate per formulare le prime teorie sull'atomo, per esempio da Bohr nel 1913. Strada facendo, le metafore alla base delle ipotesi scientifiche, si son rivelate 'false', ma son comunque rimaste alla base dei processi che hanno portato alla formulazione delle teorie validate (la meccanica quantistica nell'esempio della fisica atomica).

Menti e calcolatori

Solitamente le metafore concettuali (e i modelli esplicativi), come appunto la metafora "L'atomo è un sistema solare", sono direzionali (Lakoff e Johnson, 1980; Gibbs, 2008; Shen e Porat, 2017), ma la metafora della mente come calcolatore presenta invece delle peculiarità. Se inizialmente la proiezione dal dominio sorgente (calcolatore) al dominio target (mente) si giocò sulla possibilità di conoscere il lavoro simbolico della mente, assimilato all'elaborazione delle informazioni nei calcolatori, ben presto si verificarono due fenomeni divergenti rispetto all'evoluzione standard di un modello:

1. La metafora venne usata anche nella direzione opposta: l'intelligenza artificiale cercava ispirazione nelle ipotesi teoriche (filosofiche, psicologiche, antropologiche, etc.) per riprodurre la mente umana nei calcolatori, rovesciando così la prospettiva.
2. Mentre inizialmente la mente veniva vista, in analogia col calcolatore, come il software che 'girava' nell'hardware cerebrale, con l'evolversi delle neuroscienze portò a contestare questa dicotomia, così come il dualismo mente-corpo.

La combinazione di queste due visioni, porta da un lato all'evoluzione della prospettiva connessionista (Rumelhart e McClelland, 1986), e dall'altra a una rilettura della metafora del calcolatore (Smolensky, 1991): la mente è un prodotto dell'attività cerebrale, quindi se l'intelligenza artificiale vuole raggiungere la competenze del comportamento cognitivo degli esseri umani deve passare per la simulazione delle reti neurali. È il computer che riproduce il comportamento del cervello, facendo appello alle conoscenze sul funzionamento della trasmissione di segnale tra neuroni, e che ipotizza, spesso indipendentemente dalle conoscenze di natura psicologica o filosofica, comportamenti 'intelligenti' da eseguire adottando una tecnica di rinforzo molto più vicina al comportamentismo che al cognitivismo.

Nuove metafore per la mente

E oggi, a quale punto di evoluzione è arrivata la metafora del calcolatore rispetto allo stato dell'arte delle odierne tecnologie intelligenti? (Awati e Buckingham Shum, 2018). L'aumento delle capacità di calcolo e di memorizzazione delle informazioni, nonché la strutturazione in rete delle informazioni, hanno generato di per sé, quasi spontaneamente, comportamenti 'collettivi' intelligenti. I big data, opportunamente analizzati, ci danno per esempio informazioni sulle intenzioni di gruppi di persone (persino gli orientamenti di voto). Le indagini corpus based in linguistica permettono di individuare fenomeni che prima non era possibile trattare o venivano trattati in modo inadeguato proprio per l'impossibilità di un'analisi empirica. Eccetera.

Pertanto, alla lice di quanto affermato, si ritiene necessario un cambiamento di paradigma interpretativo nelle scienze cognitive, che consenta di capire l'evoluzione di concetti quali quello di "intelligenza", "apprendimento", "comunicazione", etc. alla luce di quanto i big data e le possibilità di analisi automatica, il pattern matching e il machine learning permettono di fare. Questo ci porterà a superare la metafora del calcolatore, o -nell'ipotesi più debole e minimalista- ridefinirne la direzionalità e i termini. Insomma, in modo più o meno radicale, sembra arrivato il momento di cambiare metafora.

References

Awati, K. and Buckingham Shum,, S., *Big Data Metaphors We Live By*, https://towardsdatascience.com/big-data-metaphors-we-live-by-98d3fa44ebf8

Gardner, H. (1987). *The Mind's New Science: A History Of The Cognitive Revolution*, Basic Books.

Gibbs, R. (2008). "Metaphor as structure mapping", in *The Cambridge Handbook of Metaphor and Thought*, Cambridge University Press.Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, Cambridge, (MA): MIT Press.

Shen, Y., & Porat, R. (2017). Metaphorical Directionality: The Role of Language. In B. Hampe (Ed.), *Metaphor: Embodied Cognition and Discourse*, Cambridge: Cambridge University Press, 62-81.

Smolensky, P. (1991). The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn'. In T. Horgan and J. Tienson (eds.), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academic Publishing Company, 281-306.

Moral Action Recognition: A Challenge for Artificial Moral Cognition

Hasse Hämäläinen[1]
[1]*Jagiellonian University*

Recently, the has been a rapid succession of advances in the ability of artificial neural networks (NNs) to *identify* and *label* various aspects of reality by abstracting from perceptual data. Currently, they can identify not only static objects such as cars and animals, but also produce increasingly accurate *descriptions* of what is taking place in a given image. As a continuation of this development, today the theorists of AI are even asking whether an NN could be made to autonomously evaluate what would be an appropriate reaction to the descriptions produced by another NN, e.g. whether or not a described acting should be regarded as 'permissible' (see e.g. Guarini 2010, Conitzer *et al*. 2017). Attention dedicated to this question has increased as a result of e.g. legislation such as NetzDG in Germany that obliges social media providers to remove all harmful content within a short time from uploading.

Taking a step from mere identification to evaluation is not only a major technological challenge, but also ethical one - because the future in which machines assess the morality of our behaviour on our behalf does not seem a particularly attractive prospect for those that still place value to human moral autonomy. Although some propose that equipping the autonomous systems that identify which content is (not) permissible with "kill switches" suffices to address this concern (Howard and Muntean 2016), I argue that a more viable and ethically sound approach would be to recognise that an autonomous NN that evaluates content on the basis of the descriptions of situations is, as a matter of fact, not possible to implement (even) with current NN techniques due to a crucial difference between moral cognition and the kinds of cognition - such as image recognition - that are informed by probability estimations.

Although NNs are not the only way to attempt to model moral cognition, most current research of artificial moral cognition is focused on modelling it by means of them. The alternative way would be to build a set of if/then -rules. However, since there are so many things that can be morally relevant in a description of an action, the set would need to be arbitrarily limited, e.g. to apply only to descriptions that are 'easy' for machines to unpack (e.g. remove any content with a slur). The supposed advantage of NNs over if/then -rules is that they seem to avoid similar arbitrariness (e.g. Churchland 2015). The success of NNs with recognising what acting a random, unseen image represents (Vinyals *et al*. 2014) - which is a close to impossible task for a set of if/then rules - could be presented as evidence for this.

Usually, NN-based methods for modelling moral cognition have many interconnected layers of 'neurons', that is, functions that reject/deamplify or pass forward/amplify their input value on the basis of a numerical weight assigned to it. In such a model, *the input* to the first layer of a NN could consist, e.g. in the descriptions of actions (Guarini 2010), or in the 'patterns' of features (that another NN has) attributed to actions (Howard and Muntean 2017). *The output* of the NN would consist in labelling the data that has passed all the layers with e.g. string 'remove'. The labels assigned to the actions by the NN would be checked, for example, by people with knowledge of moral theory and weights adjusted as many times as needed to ensure that the output of the network would meet their standards. Once this condition is met, the network can be let to operate without human supervision, i.e. autonomously - and later on, its output could serve as the gold standard in the training of the new generation of moral NNs.

This path to artificial autonomous moral cognition seems straightforward, and it may seem that above all practical problems (e.g. the lack of suitably large publicly available training datasets) still prevent moral NNs from producing the anticipated results. However, I suggest that there might be also a more fundamental problem. Unlike the proponents of moral NNs think, I claim that *also* moral NNs suffer from certain arbitrariness. So as to see this, we may compare them with image recognition NNs.

When faced with an unseen image, a trained image recognition NN scans its for the 'activation maps' of pixel values that it already 'knows' to represent a certain class of objects. If the activation map is only distantly (e.g. only 10% of the pixel values match) similar to the 'known' pattern, the NN assigns an equally small weight (0.1) to that pattern in order to 'mark' that the *probability* of the two representing the same class is low. Typically, NN recognises an image to contain a representation of a certain object *iff* the probability rises above a threshold set by the designers of the network (e.g. 0.9).

However, I claim that the above situation is *not* analogous with moral cognition. For an action can have default moral relevance *even if* its description had only very little (if anything) in common with the descriptions of the previous morally relevant actions. I show - with the help of examples - that doing something *unexpected* is a feature of many morally outrageous actions. Since a NN that has been trained with the data about the previous actions is unable to recognise unexpected actions, it is bound to ignore the moral relevance of such actions. I show that an unfortunate consequence of the failure of the NN to recognise a morally relevant action would be that it may continue to remove content that might be far more innocuous than the ignored, unexpected content, thus leaving *more space* for that particular content to spread. In this way, the introduction of autonomous moral NNs to keep harmful content in reins, I argue, could paradoxically exacerbate the problem of such content spreading across internet.

Since autonomous moral cognition requires 'sensitivity to the unpredictable', it is not possible to realise such cognition with NN techniques that rely on probabilities. However, on a positive note, I also think that autonomy might not be even needed for successful artificial moral cognition: perhaps it would suffice to implement to a NN an ability that Aristotle calls *synesis* - i.e. ability to discern how a real 'virtuous person' would evaluate the situation at hand. However, until there is no viable method to implement *synesis* to NNs, my conclusion implies that there is a reason to return to if/then -rule based systems, or stick to designing moral NNs in such a way that they leave the final decisions to humans.

References

Aristotle (1920). *Ethica Nicomachea*. Edited by I. Bywater. Oxford: Oxford University Press.

Churchland, P. M. (2015). Rules: The Basis of Morality… ?. In T. Metzinger and J. M. Windt (eds.). *Open MIND: 6(T)*. Frankfurt am Main: MIND Group.

Conitzer, V., W. Sinnot-Armstrong, J. S. Borg, Y. Deng and M. Kramer (2017). Moral Decision Making Frameworks for Artificial Intelligence. AAAI 2017 Conference paper. Available at https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14651/13991.

Guarini, M. (2010). Particularism, Analogy and Moral Cognition. *Minds and Machines*, 20(3), 385-422.

Howard, D. and I. Muntean (2016). A Minimalist Model of the Artificial Autonomous Moral Agent (AAMA). AAAI 2016 Conference paper. Available at https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12760/11954.

Howard, D. and I. Muntean (2017). Artificial Moral Cognition: From Functionalism to Autonomous Moral Agents. In Powers, T. (ed.). *Philosophy and Computing: Essays in Epistemology, Philosophy of Mind, Logic, and Ethics.* Berlin: Springer.

Vinyals, O., A. Toshev, S. Bengio and D. Erhan (2014). Show and Tell: A Neural Image Caption Generator. Available at https://arxiv.org/abs/1411.4555.

DELTA: A Unifying Categorization Algorithm Integrating Prototypes, Exemplars and Theory-Theory Representations and Mechanisms

Antonio Lieto
Università di Torino, Dipartimento di Informatica e ICAR-CNR (Palermo)

This paper details how to reconcile, under a heterogeneous representational perspective, different theories of typicality about conceptual representation and reasoning that have been traditionally seen as incompatible. In particular, it provides a novel theoretical hypothesis - as well as a novel categorization algorithm called **DELTA** - (i.e. unifie**D** Cat**E**gorization a**L**gorithm for he**T**erogeneous represent**A**tions) able integrate, in a cognitive artificial agent, the representational and reasoning assumptions of the theory-theory of concepts with the those ascribed to the prototype and exemplars-based theories.
The DELTA algorithm (detailed below) has the goal of selecting, given a certain stimulus d perceived from the environment, the most appropriate **typicality-based** representation available in the declarative memory of a cognitive agent (i.e. a prototype, an exemplar or a theory-like structure).

Data: Stimulus d; list of candidate representations: $closed^{\mathcal{S}1}$.
Result: A typicality based representation of a category.

$closed^{\mathcal{S}1} = \{\emptyset\}$
$\mathcal{S}1_{\mathrm{EX}} \leftarrow categorizeExemplars(d)$;
if $firstOf(\mathcal{S}1_{\mathrm{EX}},\ closed^{\mathcal{S}1}).distance(d) < similarityThreshold$ **then**
  return $firstOf(\mathcal{S}1_{\mathrm{EX}},\ closed^{\mathcal{S}1})$;
else
  $\mathcal{S}1_{\mathrm{PR}} \leftarrow categorizePrototypes(d)$ **return** $firstOf(\mathcal{S}1_{\mathrm{PR}},\ closed^{\mathcal{S}1})$;
end
if $firstOf(\mathcal{S}1_{\mathrm{PR}},\ closed^{\mathcal{S}1}).distance(d) > ConceptualCoherenceThreshold$ **then**
  return $firstOf(\mathcal{S}1_{\mathrm{PR}},\ closed^{\mathcal{S}1})$;
else
  $\mathcal{S}1_{\mathrm{T}} \leftarrow categorizeTheory(d)$;
  return $firstOf(TheoryBasedCategorization,\ closed^{\mathcal{S}1})$;
end

Algorithm 1: A Unified categorization algorithm for prototypes, exemplars and theory-like representations.

Following a preference that has been experimentally observed in human cognition (Medin and Schaffer, 1978), DELTA assumes that exemplars and prototypes-based categorization are executed first (and in the following order: first exemplars-based categorization is attempted and then a prototype-based one) but also include a theory-like mechanism able to eventually discard the categorization result based on prototypes in favor of more coherent theory-like representations (as originally shown in the experiments by (Keil, 1989). The heuristics used for the choice between prototype and theory-like representation is grounded on the notion of Conceptual Coherence (Thagard, and Verbeurgt, 1998). In particular: the prototypical answer is maintained in case the considered stimulus results to be "coherent enough" with respect to the corresponding micro-theory related to the selected prototype, otherwise it is overridden by the theory-like representation which is closer to the stimulus.

Representational assumptions for DELTA

The DELTA algorithms relies on the representational hypothesis according to which conceptual structures, in natural and artificial systems, are *heterogeneous proxytypes* (see Lieto, 2014 for details). In this view, a concept is composed by heterogeneous bodies of knowledge containing different types of information associated to the the same conceptual entity. Each body of knowledge is a proxytype (i.e. an element of a representational network stored in the agent's long term memory corresponding to a particular category and that can tokenized in working memory to "go proxy" for that category). Furthermore, each body of conceptual knowledge is assumed to be featured by specific processes in which such representations are involved (e.g., in cognitive tasks like recognition, learning, categorization, etc.). Such heterogeneous perspective has been explicitly taken into account in the DUAL-PECCS system (http://www.dualpeccs.di.unito.it). In particular, the current version of DUAL-PECCS exhibits both prototype and exemplars-based categorization (see figure 1 below).

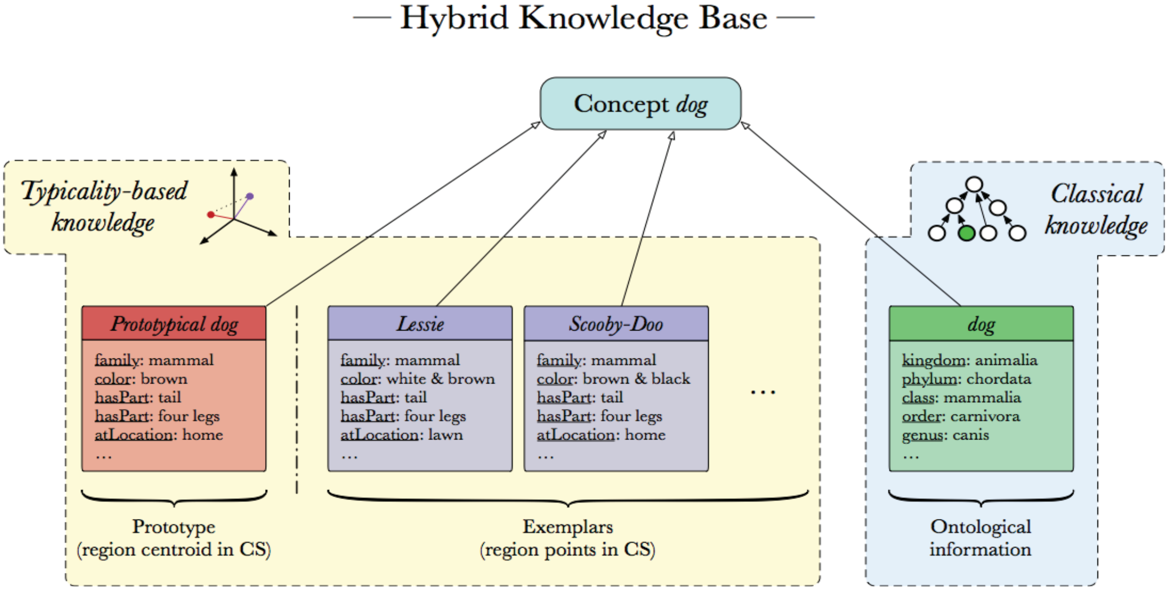


Fig. 1. An example of heterogeneous representation of the concept DOG in DUAL-PECCS

A missing part of its the current heterogeneous conceptual architecture concerns the representation of the default knowledge in terms of theory-like representational structures (while it already integrates classical, prototypical and exemplars based knowledge representation and processing mechanisms). To overcome these limits, I propose: i) to use graphical models (Danks, 2004) in order to represent the type of common-sense knowledge assumed in the theory-theory hypothesis; ii) to adopt the proposed unifying algorithm to explicitly integrate different types of typicality-based categorization mechanisms. A preliminary test on the task of common-sense linguistic categorization (involving only prototypes and exemplars based representations and reasoning procedures) has obtained promising result when compared with human performances (with an overlapping of the 89% of the responses, see (Lieto, Radicioni, and Rho, 2017)).

References

Danks, D. (2004). Psychological theories of categorizations as probabilistic models, CMU Technical report.

Keil, F. (1989). Concepts, kinds, and cognitive development. Cambridge, MA. . MIT press.

Lieto, A. (2014). A computational framework for concept representation in cognitive systems and architectures: Concepts as heterogeneous proxytypes. *Procedia Computer Science*, *41*, 6-14.

Lieto, A., Radicioni, D.P., Rho, V. (2017). Dual PECCS: A Cognitive System for Conceptual Representation and Categorization. *Journal of Experimental & Theoretical Artificial Intelligence* 29(2), 433–452.

Lieto, A. (2019). Heterogeneous Proxytypes Extended: Integrating Theory-like Representations and Mechanisms with Prototypes and Exemplars, in *Proceedings of BICA 2018, 9th Internat.Conference on Biologically Inspired Cognitive Architecture*s, Springer, Advances in Intelligent Systems and Computing.

Medin, D.L., Schaffer, M.M. (1978). Context theory of classification learning. *Psychological review* 85(3), 207.

Thagard, P., Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science* 22(1), 1–24.

Composing Prototypes by Coupling a Non Monotonic Description Logic with Probabilities and Cognitive Heuristics

Antonio Lieto*^ e Gian Luca Pozzato*
** Università di Torino, Dipartimento di Informatica*
^ ICAR-CNR, Palermo

Combining typical knowledge to generate novel concepts is an important creative trait of human cognition. Dealing with such ability requires, from an AI perspective, the harmonization of two conflicting requirements that are hardly accommodated in symbolic systems: the need of a syntactic compositionality (typical of logical systems) and that one concerning the exhibition of typicality effects (Frixione and Lieto, 2012). According to a well-known argument (Osherson and Smith, 1981), in fact, prototypical concepts are not compositional. The argument runs as follows: consider a concept like pet fish. It results from the composition of the concept pet and of the concept fish. However, the prototype of pet fish cannot result from the composition of the prototypes of a pet and a fish: e.g. a typical pet is furry and warm, a typical fish is grayish, but a typical pet fish is neither furry and warm nor grayish (typically, it is red).
In this work we provide a logical framework able to account for this type of human-like concept combination. We propose a nonmonotonic Description Logic (from now on DL) of typicality called $\mathbf{T^{CL}}$ (**Typicality-based Compositional Logic**).
This logic combines three main ingredients (for the details see Lieto and Pozzato, 2018; Lieto and Pozzato submitted). The first one relies on the DL of typicality ALC + TR introduced in (Giordano et al., 2015). In this logic, "typical" properties can be directly specified by means of a "typicality" operator T enriching the underlying DL, and a knowledge base (KB) can contain inclusions able to represent that "typical Cs are also Ds". In the logic ALC + TR one can consistently express exceptions and reason about defeasible inheritance as well.
As a second ingredient, we consider a distributed semantics similar to the one of probabilistic DLs known as DISPONTE (Riguzzi et al, 2015), allowing to label ontological axioms with degrees representing probabilities, but restricted to typicality inclusions. The basic idea is to label inclusions of the type "typical Cs are also Ds" with a real number between 0.5 and 1, representing its probability, assuming that each axiom is independent from each others (the actual probabilistic values are assumed to come from an application domain). The resulting knowledge base defines a probability distribution over scenarios.
As an additional element of the proposed formalization we employ a method inspired by cognitive semantics (see Hampton, 1987 for a review) for the identification of a dominance effect between the concepts to be combined. In particular, for every combination, we distinguish a HEAD, representing the stronger element of the combination, and a MODIFIER. The basic idea is: given a KB and two concepts CH (HEAD) and CM (MODIFIER) occurring in it, we consider only some scenarios in order to define a revised knowledge base, enriched by typical properties of the combined concept.

Selection Criteria

Given a KB K and given two concepts CH and CM occurring in K, our logic allows defining the compound concept C as the combination of the HEAD (CH) and the MODIFIER (CM), where $C \sqsubseteq CH \sqcap CM$ and the typical properties of the form $T(C) \sqsubseteq D$ to ascribe to the concept C are obtained in the set of scenarios obtained by applying the DISPONTE semantics, that: i) are consistent with respect to K; ii) are not trivial, i.e. those with the highest probability, in the sense that the scenarios considering all properties that can be consistently ascribed to C, or all the properties of the HEAD that can be consistently ascribed to C are discarded; iii) are those giving preference to the typical properties of the HEAD CH (with respect to those of the MODIFIER CM) with the highest probability.

Composing the PET FISH

Let K be a Knowledge base containing the rigid inclusion (*) $Fish \sqsubseteq \forall livesIn.Water$ and the following typical inclusions equipped with probabilities:

1. $0.9 :: \mathbf{T}(Pet) \sqsubseteq \forall livesIn.(\neg Water)$
2. $0.8 :: \mathbf{T}(Pet) \sqsubseteq Affectionate$
3. $0.7 :: \mathbf{T}(Fish) \sqsubseteq \neg Affectionate$
4. $0.8 :: \mathbf{T}(Pet) \sqsubseteq Warm$
5. $0.6 :: \mathbf{T}(Fish) \sqsubseteq Greyish$
6. $0.9 :: \mathbf{T}(Fish) \sqsubseteq Scaly$
7. $0.8 :: \mathbf{T}(Fish) \sqsubseteq \neg Warm$

In this case we have $2^7 = 128$ different scenarios obtained as in the DISPONTE semantics. In our logic, we can discard those that are not consistent, are trivial and privilege the MODIFIER with respect to the HEAD. It turns out that our logic is able to select the scenario with the following typical properties:

3. $0.7 :: \mathbf{T}(Fish) \sqsubseteq \neg Affectionate$
6. $0.9 :: \mathbf{T}(Fish) \sqsubseteq Scaly$
7. $0.8 :: \mathbf{T}(Fish) \sqsubseteq \neg Warm$

On the other hand, the composed concept PET FISH also inherits the rigid inclusion ⊑ ∀livesIn.Water from (*).

Notice that in, our logic, adding a new inclusion **T**(PET ⊓ FISH) ⊑ Red, would not be problematic (i.e. $\mathbf{T}^{\mathbf{CL}}$ tackles the phenomenon of prototypical attributes emergence). The proposed logic has been recently applied to a number of cognitive phenomena including: conjunction fallacy, metaphors generation, and iterative conceptual combination (Lieto and Pozzato, submitted).

References

Frixione, M., Lieto, A. (2012). Representing concepts in formal ontologies: Compositionality vs. typicality effects. *Logic and Logical Philosophy* 21(4), 391–414.

Osherson, D.N., Smith, E.E. (1981) . On the adequacy of prototype theory as a theory of concepts. *Cognition* 9(1) 35–58.

Giordano, L., Gliozzi, V., Olivetti, N., Pozzato, G.L. (2015) Semantic characterization of Rational Closure: from Propositional Logic to Description Logics. *Artificial Intelligence* 226, 1–33

Hampton, J.A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition* 15(1), 55–71.

Riguzzi, F., Bellodi, E., Lamma, E., Zese, R.: Reasoning with probabilistic ontologies. In Yang, Q., Wooldridge, M., eds.: *Proceedings of IJCAI 2015*, AAAI Press (2015) 4310–4316.

Lieto, A., Pozzato, G.L.(2018). A description logic of typicality for conceptual combination. In: *Proceedings of the 24th International Symposium on Methodologies for Intelligent Systems, ISMIS 2018*, Lecture Notes in Artificial Intelligence, Springer.

Lieto, A., Pozzato, G.L.(201x). A Description Logic Framework for Commonsense Conceptual Combination Integrating Typicality, Probabilities and Cognitive Heuristics, submitted.

.

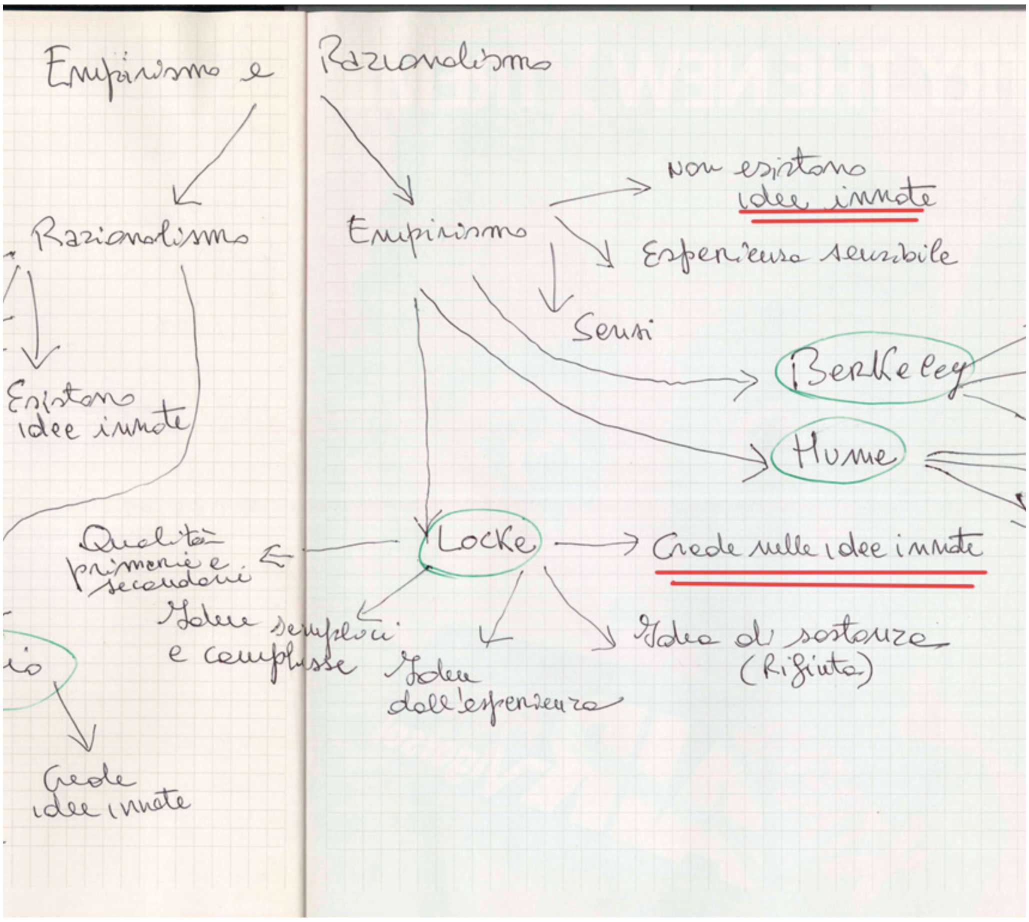
Mappe concettuali vs ontologie. Un confronto sull'utilizzo di strumenti informatici per la didattica della storia e della filosofia

Antonio Lieto*^ e Francesco Rebuffo*
** Università di Torino, Dipartimento di Informatica*
^ Liceo Scientifico "Guido Parodi", Acqui Terme

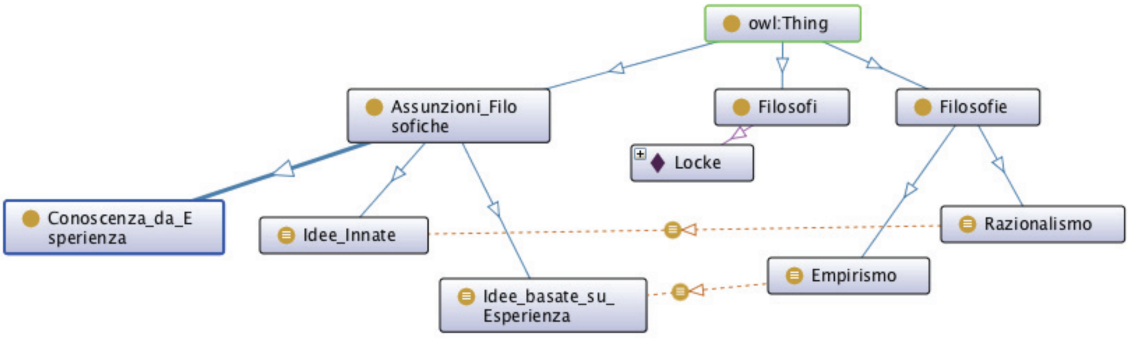
Questo lavoro propone un confronto tra diversi strumenti utilizzabili per modellare la conoscenza di dominio in ambito didattico: le mappa concettuali, Novak e Cañas (2006), (uno strumento tradizionalmente utilizzato nelle scuole) e le ontologie computazionali (dei sistemi formali di modellazione concettuale, attualmente molto usati nei sistemi di intelligenza artificiale per le loro capacità di "ragionamento automatico", si veda Guarino, (1995)).

Nello specifico, questo articolo presenta il risultato di un un doppio esperimento sul campo condotto presso il Liceo Scientifico "Guido Parodi" di Acqui Terme in cui gruppi di studenti paragonano lo strumento della mappa concettuale e quello dell'ontologia nella risoluzione di due problemi di "misconcezione" (o errata concettualizzazione): uno indotto attraverso la consegna di appunti e materiali didattici contenenti informazioni volontariamente contraddittorie tra loro (caso che potrebbe corrispondere alla situazione in cui uno studente prende - per qualche motivo - degli appunti in modo scorretto) e l'altro legato ad una complessità concettuale intrinseca all'argomento.

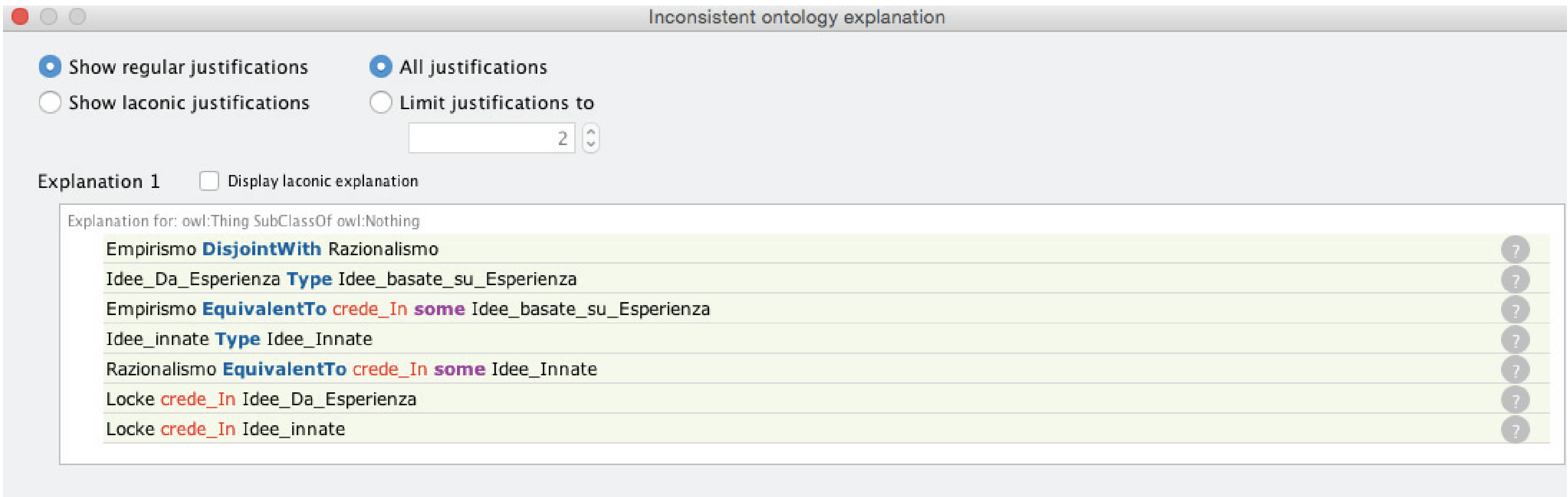
In basso, un esempio reale di confronto tra i due strumenti. Nel primo caso un esempio cartaceo di una mappa concettuale realizzata con carta e penna dai ragazzi di una classe quarta del liceo scientifico sul tema Empirismo e Razionalismo dove vengono indicate correttamente le tesi principali dell'Empirismo e dove si induce attribuendo a Locke la credenza nelle idee innate nonostante poco sopra sia stato affermato che per gli empiristi non esistono idee innate. La contraddizione viene evidenziata in fase di correzione ma non era stata rilevata dagli studenti neanche quando lo stesso schema era stato fatto modellare tramite C-MAPS (il principale software di moderazione di mappe concettuali).



Nel secondo caso, per lo stesso tema di modellazione assegnato, gli studente hanno creato - con il supporto di un ricercatore - una ontologia (visibile in basso) in grado di rilevare errate concettualizzazioni.



In questo caso: se si inserisce l’individuo Locke all’interno della classe ontologica “Empirismo” e poi si prova (indotti da una concettualizzazione errata) ad assegnare a Locke la proprietà “crede in dee innate” il software mostra l’inconsistenza della base ontologica perché Locke essendo “Empirista” può solo credere in “Idee basate su esperienza”.
Grazie al reasoner ontologico, infatti, quando l’utente prova ad assegnare all’istanza “Locke” l’informazione contraddittoria la stessa contraddizione emerge immediatamente (a differenza della mappa) e viene spiegato agli studenti il motivo di tale inconsistenza (come si vede nella figura in basso)



Il principale risultato emerso da tale lavoro mette, dunque, in luce il ruolo che le ontologie e le tecnologie semantiche possono avere in ambito didattico al fine scovare eventuali errori di concettualizzazioni (misconceptions). Il mero utilizzo di mappe concettuali (sia fatte a mano che fatto con strumenti come C-Maps), invece, non permette agli studenti di accorgersi di aver appreso concettualizzazioni sbagliate su un determinato dominio di conoscenza.

Riferimenti

1) Joseph D. Novak & Alberto J. Cañas (2006). "The Theory Underlying Concept Maps and How To Construct and Use Them", Institute for Human and Machine Cognition. Accessed 24 Nov 2008.
2) Guarino, Nicola. "Formal ontology, conceptual analysis and knowledge representation." International journal of human-computer studies 43, no. 5-6 (1995): 625-640

Sounds in a vacuum: a defence of O'Callaghan's theory.

Giulia Lorenzi
University of Bergamo

In the contemporary debate about the metaphysics of sounds, the dispute about the existence of sounds in a vacuum has played a central role. In this field of studies, the two most articulated positions are expressed in Casati and Dokic's theory (1994) and O'Callaghan's (2007), and they differ completely over this ontological trait. Indeed, Casati and Dokic support a distal theory of sounds as events and they suggest that a sound is "only" a vibrating event of a source. While O'Callaghan proposes a distal theory of sounds as *relational* events and he considers the presence of a surrounding medium to be a necessary condition for the existence of a sound.

Casati and Dokic (2014) submit the thought experiment of sounds in a vacuum (which appeared for the first time in Berkeley's *Three Dialogues between Hylas and Philonous*) to O'Callaghan to bring out the difference between their positions and to suggest that their rival defends a counter-intuitive position. In particular, they present the perceptual situation of a sound of a tuning fork in an evacuated jar and imagine opening and closing the lid of the jar. They think that the vast majority of perceivers will come to say that, in that particular case, there is only one sound that cannot be perceived continuously. To support their position, Casati and Dokic also compare the experience of the jar with three visual cases of the tunnel effect.

On the other hand, O'Callaghan tries to find good reasons to reject Casati and Dokic's objection. Firstly, O'Callaghan discusses the visual analogies presented by his rivals. Secondly, he show that the thought experiment cannot lead to certain metaphysical results. Nevertheless, he does not try to argue that, in the jar situation, it is possible to perceive a sequence of sounds, instead of a single sound that pops into and out of perception.

Here, I focus mainly on the perceptual aspects, rather than on the metaphysical elements, involved in the thought experiment. Firstly, my aim is to demonstrate that Casati and Dokic's perceptual intuition is less reasonable than they think. I deploy Bozzi's phenomenological account (2018) of the aggregation and separation of sounds to show that it is plausible to perceive a sequence of different sounds instead of a single sound. Secondly, I want to suggest that this difficult case forces us to consider another relevant and oft-ignored element in this debate: the metaphysical and perceptual nature of silence (Sorensen 2009). Indeed, in their accounts of sounds both O'Callaghan and Casati and Dokic do not provide an analysis of silence. My final goal is to show that Casati and Dokic's position leads to difficulties in the explanation of the metaphysics and the perception of silence and, for this reason, it becomes more counter-intuitive than O'Callaghan's.

On the basis of this argumentation, I want to provide a new argument in favour of O'Callaghan's general theory of sound as a relational event.

References:

Bozzi, P., (2018). Two factors of unifications for musical notes: closeness in time and closeness in tone. Translated and commented by L. Zecchinelli in *Paolo Bozzi's Experimental Phenomenology* (ed. I. Bianchi and R. Davies). London: Routledge, 273-304.

Casati, R., Dokic, J., (1994). *La Philosophie du Son*. Nîmes: Editions Jacqueline Chambon.

Casati, R., Dokic, J., (2014). Sounds. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.).

O'Callaghan, C., (2007). *Sounds. A Philosophical Theory*. Oxford: Oxford University Press.

Sorensen, R., (2009). Hearing Silence: The perception and introspection of absences. In Nudds M., O'Callaghan C., *Sounds & perception. New Philosophical Essay*. Oxford: Oxford University Press, 7.

Dalla pedagogia naturale alla pedagogia robotica

Emiliano Loria

Consorzio FINO – Università di Genova

Nasciamo immersi in un mondo di artefatti, oggetti artificiali il più delle volte sofisticati, di non immediata comprensione, impiegati con diverse funzioni in una complessa rete di azioni, spesso riconducibili ad interscambi comunicativi, che solo in parte hanno una connotazione prettamente verbale. E cominciamo ad apprendere da subito, probabilmente grazie a dispositivi innati che orientano il comportamento lungo particolari direzioni adattive all'interno della rete sociale, e culturale, in cui veniamo immediatamente collocati dalla nascita.

Gli psicologi ungheresi Csibra e Gergely (2006; 2011; Gergely&Csibra 2005; Gergely&Csibra 2006) ipotizzano l'esistenza di un sistema cognitivo di *pedagogia naturale* caratterizzato da un trasferimento di conoscenze da adulti a bambini che avviene attraverso relazioni comunicative triadiche, atte a rendere l'apprendimento infantile rapido ed efficace. Tale relazione, detta pedagogica, si innesca grazie all'impegno dell'adulto di produrre opportuni segnali ostensivi diretti al bambino e focalizzati su un referente esterno; e, d'altra parte, si basa sull'innata capacità dei bambini di cogliere tali segnali comunicativi e leggerne la natura referenziale.

L'impegno dell'adulto espresso benevolmente attraverso sorrisi, vocalizzazioni, sguardi diretti, gesti deittici, è *normalmente* letto e vissuto dai bambini in maniera positiva, attirando e spostando la loro attenzione sul referente, che è il reale oggetto della comunicazione. Il sistema di inferenze, computazioni ed elaborazioni in merito al contenuto informativo (Gergely 2007; Cesana-Arlotti et al. 2018) può avviarsi conseguentemente a questa fase relazionale preparatoria. In altri termini, la forza performativa dell'atto comunicativo - che potremmo parafrasare nei termini di 'forza illocutoria dell'enunciazione' (Domaneschi 2014) - è indotta, dunque, da un insieme di pratiche comunicative verbali e non-verbali (es.: contatto visivo, *joint attention*, linguaggio *motherese*, gesti deittici, parole di incoraggiamento, di saluto espresse nella lingua d'uso, di nominazione del bambino interlocutore, ecc.). Tali pratiche comunicative hanno un ventaglio di scopi che vanno dalla manifestazione esplicita di intenzioni comunicative dell'adulto (Csibra 2010), dallo stringere una relazione di fiducia/deferenza col bambino (Recanati 1997; Sperber et al. 2010), all'innesco di una aspettativa referenziale orientata al contenuto dell'enunciato (che può consistere, ad esempio, in una dimostrazione di funzionamento di un oggetto), secondo quanto predetto dalla teoria della pertinenza (Sperber&Wilson 1995). In alcuni casi, tuttavia, l'intera gamma di questi atti pragmatici (Airenti 2017) può non essere compresa dal bambino. Si immaginino, ad esempio, i casi di cecità e sordità dalla nascita, o quei casi di compromissioni cognitive che si registrano nell'autismo. Quest'ultima patologia, in particolare, rende i bambini incapaci di interpretare correttamente quei segnali comunicativi più sopra riportati (cfr., Cummings 2014; 2017). Il loro apprendimento, dunque, è seriamente ostacolato.

Al fine di aggirare tali iniziali difficoltà di apprendimento si sono avviate con successo, da una decina di anni almeno, politiche di sviluppo di una robotica tesa a reinventare modalità di trasmissione di saperi, che semplificano di molto i tratti peculiari della *comunicazione umana uno-a-uno*. L'impiego di robot educativi è sempre più diffuso, anche in Italia. Si veda il caso, ad esempio, del progetto inglese "Aurora" (http://aurora.herts.ac.uk), oppure del nostrano "Robotiko" (https://www.robotiko.it/robot-per-bambini-autistici). Uno degli obiettivi della robotica educativa è quello di promuovere l'imitazione, l'associazione di causa ed effetto tra azione e comportamento, comprensione/acquisizione di prime parole, stimolare l'apprendimento sociale (Dautenhahn&Werry 2004; Di Lieto et. al. 2017; Sullivan 2008; Thorell et al. 2009).

Che cosa determina un tale successo? Cosa viene sottratto o aggiunto rispetto alla comunicazione tra essere umani, e che cosa persiste? È lecito parlare di assenza di empatia tra robot e bambino come elemento facilitante del rapporto comunicativo-educativo? Ipotizzo che rivolgersi al livello di empatia, e quindi direttamente o indirettamente al coinvolgimento di fattori legati al *mindreading* – sistema cognitivo che si suppone essere compromesso nell'autismo-, sia fuorviante. Piuttosto, sarebbe meglio circoscrivere l'ambito di comunicazione robot-bambino al livello percettivo della relazione comunicativa in quanto tale, mettendo in rilievo i tratti pragmatici salienti della comunicazione pedagogica descritta dalla teoria della pedagogia naturale, e ad alcuni *biases* cui essa conduce.

References

Airenti, G. (2017). Pragmatic Development. In: Cummings, L. (ed.) (2017). *Research in Clinical Pragmatics*. Cham: Springer, 3-28.

Arduino, M. (2014). *Il bambino che parlava con la luce. Quattro storie di autismo*. Torino: Einaudi.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., Bonatti, L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, 359(6381), 1263–1266.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25(2), 141–168.

Csibra, G., Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Munakata, Y. Johnson, M.H. (eds.). *Processes of Change in Brain and Cognitive Development. Attention and Performance*, XXI. Oxford: Oxford University Press, 249-274.

Csibra, G., Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transaction of the Royal Society B*, 346, 1149-1157.

Cummings, L. (2014). *Pragmatic Disorder*. Dordrecht: Springer.

ID. (ed.) (2017). *Research in Clinical Pragmatics*. Cham: Springer.

Dautenhahn, K., Werry, I. (2004). Towards interactive robots in autism therapy. Background, motivation and challenges. *Pragmatics & Cognition*, 12(1), 1–35.

Di Lieto, M.C., Inguaggiato, E., Castro, E., Cecchi, F., Cioni, G., Dell'Omo, M., Laschi, C., Pecini, C., Santerini, G., Sgandurra, G., Dario, P. (2017). Educational Robotics Intervention on Executive Functions in preschool children: a pilot study. *Computers in Human Behavior*, 71, 16-23.

Domaneschi, F. (2014). *Introduzione alla pragmatica*. Roma: Carocci.

Gergely, G. (2007). Learning "about" versus learning "from" other minds: Natural pedagogy and its implications. In: Carruthers, P., Laurence, S., Stich, S. (eds). *The innate mind: vol. 3: foundations and the future*. Oxford: Oxford University Press, 170-198.

Gergely, G., Csibra, G. (2005). The social construction of the cultural mind. Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, 6(3), 463-481.

Gergely, G., Csibra G., (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: Enfield, N.J., Levenson, S.L. (eds.). *Roots of Human Sociality: Culture, Cognition, and Human Interaction*. Oxford: Berg Publishers, 229-255.

Parise, E., Csibra, G. (2013). Neural responses to multimodal ostensive signals in 5-monthold infants. *PloS one*, 8(e72360).

Recanati. F. (1997). Can we believe what we do not understand?. *Mind & Language*, 12, 84–100.

Sperber, D., Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., Wilson, D. (2010) Epistemic vigilance. *Mind & Language*, 25(4), 359–393.

Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills and systems understanding. *Journal of Research in Science Teaching*, 45(3), 373-394.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12(1), 106–113.

Playing with physical and digital cards to enhance numerical cognition in children

Orazio Miglino[1,3], Massimiliano Schembri[2], Antonio Cerrato[1], Michela Ponticorvo[1]
[1]*Natural and Artificial Cognition Lab, University of Naples Federico II – Naples, Italy*
[2]*Federica Web Learning, University of Naples Federico II – Naples, Italy*
[3]*Institute of Cognitive Sciences and Technologies, National Research Council – Rome, Italy*

Introduction

Numerical cognition can be studied through different methods including developmental and simulative ones (Ponticorvo et al., 2015). Before language acquisition, toddlers are sensitive to numerical properties; they detect differences between small numbers and they have expectations about quantity, if their expectations are violated they show disappointment or astonishment. For example, in a comparison of different sets of objects they seem attracted by the most numerous one and if they see two items placed in a hidden box they look startled if they notice, after a few time, only one item in the box (Dehaene, 2001). This evidence correlates with the theory of innate numerical abilities (Butterworth, 1999; Girelli et al., 2000), according to which we got, since the first weeks of life, innate numerical skills to classify small sets of elements (4-5 items); later, the culture teaches us how to use the numbers competency in a more advanced manner.
Quantification processes are mediated by numerical representations in analogical and/or non-verbal ways (Dehaene, 1992) and they are not based on proper arithmetical skills but rather on other specific abilities, such as the *subitizing* (a rapid and accurate judgement about the numerosity of a small set of items; Kaufman et al., 1949) and the numerosity estimation. After the utilization of these *informal* mathematical abilities, numerical cognition gets enriched by symbolic processing operations.

A ludic way to enhance numerical cognition

Here we propose a card game designed for children aged 5 years and above; the game presents a physical and a digital aspect, providing the interaction of two users: an artificial agent and a natural one (the child playing the game). The real user has a deck of physical cards composed by the four arithmetic operations (addition +, subtraction -, multiplication *x* and division *:*) plus the card representing "greater than" (>) and "less than" (<), with which he/she has to answer the mathematical questions proposed by the digitalized opponent (e.g. "*if 2 ? 7 = 9, what operation was used?*"). The particularity of the artificial agent is that it can choose between 4 different codes to represents the numbers (numerical, textual, analogical-abstract and analogical-iconographic) making the questions more challenging for the children. Moreover, the artificial agent has been implemented with an adaptive tutor able to ask questions based on the mathematical competency of the real user. For example, the game starts with the artificial agent proposing questions to the children with a single type of code (e.g. the numerical one) and single type of mathematical operation (e.g. addition). Considered the possibility to keep trace of the children's answers inside the database, only if they correctly respond for a certain amount of items, the tutoring system can increase the difficulty (modifying its internal degree of complexity) of the game switching to another type of numerical representation (e.g. changing from numbers to letters or introducing a double type of numerical code in the same question) and/or passing to another type of mathematical operation. In this way the game proceeds in respect of the unique level of the children, that acquire gradually the familiarity with arithmetics. In addition, the artificial agent has been equipped also with an emotional module that reflects the users' performance.
We developed this game in accordance to the methodology of game-based learning (GBL; Pivec, 2007) that represents for learners an opportunity to build their knowledge in an amusing setting, with the possibility to also experiment failures and receive feedbacks in an informal context, perceived as *safer* than the traditional and educative one. Nowadays, teachers and educationalists no

longer look at games and videogames as a distraction (nevertheless it persists skepticism about isolating effect that can have on children and the difference of games vs the standardized pedagogical approach), but like a powerful instrument for conveying concepts, also difficult, in an entertaining way. GBL practices can be successfully applied to the teaching of maths, as showed by Drigas and Pappas (2015) in their small literature review about GBL for mathematics. The two scholars collected information from different researches (most of whom targeted on primary and middle school children) showing that GBL has positive effect on students' mathematical skills, as well on students' cognitive and mental skills. At the same time, educational math games enhance students' motivation in learning the mathematical concepts. We agree with the opinion of the researchers, inasmuch games could represent a useful auxiliary learning tools to build an innovative teaching model for maths.

Future Directions

Since that till now we have not tested this game nor collected data in any educational settings, we want to start a proper empirical study with the first version of the game. We plan to test it with children with and without diagnosis of learning impairments (e.g. special educational needs) to investigate their *number sense*.

Moreover, the use of our game can be extended also with older adults; there are many studies (e.g. Verghese et al., 2003) indicating the benefits of leisure activities (such as playing cards) in elderly, because games can help older people in maintaining their mental acuity. In addition, the frequency of playing games can be considered as a protective factor against neurodegenerative disease.

Playing cards is a way of interacting with numbers in a ludic way; for this reason, card games represent a powerful educational tool to stimulate the numerical cognition in children and a suitable setting to study numerical abilities.

References

Butterworth, B. (1999). The mathematical brain. London: Macmillan.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1-42.

Dehaene, S. (2011). The number sense: How the mind creates mathematics. OUP USA.

Drigas, A., & Pappas, M. (2015). On line and other game-based learning for mathematics. *International Journal of Online Engineering (iJOE)*, 11(4), 62-67.

Girelli, L., Lucangeli, D., Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of experimental child psychology*, 76(2), 104-122.

Kaufman, E.L., Lord, M.W., Reese, T.W., Volkmann, J. (1949). The discrimination of visual number, in American Journal of Psychology. *The American Journal of Psychology*, 62(4), 498–525.

Pivec, M. (2007). Play and learn: potentials of game-based learning. *British Journal of Educational Technology*, 38(3), 387-393.

Ponticorvo, M., Gigliotta, O., Miglino, O. (2017). Simulative Models to Understand Numerical Cognition. *Natural and Artificial Computation for Biomedicine and Neuroscience*, 75-84. Cham: Springer International Publishing.

Verghese, J., Lipton, R. B., Katz, M. J., Hall, C. B., Derby, C. A., Kuslansky, G., Buschke, H. (2003). Leisure activities and the risk of dementia in the elderly. *New England Journal of Medicine*, 348(25), 2508-2516.

Constructing a Grammar Based on the Causations of Movements by Joints' Rotations from a Humanoid' Motion Planning

Hiroyuki Nishina
Saitama University

Aziz-Zadeh et al. (2006) reported that action sentences were understood by the activation of different sectors of the premotor cortex, depending on the effectors used in the sentences. This tempts semanticists to posit the bodies of those understood them as a semantic model on which to valuate them. Calvin, in Calvin and Bickerton (2000), pointed out that embedding is observed in an algorithm for throwing in playing bocce.

Compact humanoid model robots are now convenient for mimicking human actions on a small scale. Impressed by embedded motions in actions, we simulated human actions on a humanoid, and tryied to construct a kinematic grammar based on the causation, by each joint, of its distal edges. This grammar generates a phrase marker as a tree, representing a posture in an action.

Each action executed by a humanoid is planned as a sequence of time points, at each of which the degrees by which the 17 servomotors as its joints rotate are aligned (Kondo (2007). As data, we used differential rotation values, for pushing-up, by subtracting the degrees value in the previous time point from that in the present time point in each raw of joint's rotation degrees. See the following:

NO	Angle Diff	CH1(Sh)	Ch2(S)	Ch3(E)	Ch6(N)	Ch7(Sh')	Ch8(S')	Ch9(E')	CH13(I)	Ch14(H)	Ch15(K)	Ch16(A)	Ch17(An)	CH19((I')	CH20(H')	Ch21(K')	Ch22(A')	Ch23(An')
1	[t0, t1]	3	16	−90	0	−175	−197	−90	−91	−150	−167	−69	−92	−92	−30	−13	−111	−88
2	[t1, t2]	74	0	0	0	−78	0	0	0	−65	0	−31	0	0	66	0	32	0
3	[t2, t3]	72	0	0	0	−72	0	0	0	8	110	−41	0	0	−8	−112	42	0
4	[t3, t4]	−59	−11	0	0	51	11	0	0	70	−73	122	0	0	−69	−78	−128	0
5	[t4, t5]	0	0	0	0	0	0	0	0	0	0	−69	−2	0	0	0	72	0
6	[t5, t6]	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
7	[t6, t7]	0	60	90	0	0	−55	89	0	0	0	0	0	0	0	0	0	0
8	[t7, t8]	0	−60	90	0	0	55	−89	0	0	0	0	0	0	0	0	0	0
9	[t8, t9]	0	60	−90	0	0	−55	89	0	0	0	0	0	0	0	0	0	0
10	[t9, t10]	0	−60	90	0	0	55	−89	0	0	0	0	0	0	0	0	0	0
11	[t10, t11]	0	60	−90	0	0	−55	89	0	0	0	0	0	0	0	0	0	0
12	[t11, t12]	0	−60	90	0	0	55	−89	0	0	0	0	0	0	0	0	0	0
13	[t12, t13]	0	60	−90	0	0	−55	89	0	0	0	0	0	0	0	0	0	0
14	[t13, t14]	0	−60	90	0	0	55	−89	0	0	0	0	0	0	0	0	0	0
15	[t14, t15]	0	60	−90	0	0	−55	89	0	0	0	0	0	0	0	0	0	0
16	[t15, t16]	0	−60	90	0	0	55	−89	0	0	0	0	0	0	0	0	0	0
17	[t16, t17]	0	0	0	0	0	0	0	0	0	0	69	2	0	0	0	−72	0
18	[t17, t18]	59	0	0	0	−51	0	0	0	−78	75	−121	0	0	76	−78	125	0
19	[t18, t19]	−124	9	0	0	124	−11	0	0	0	−112	40	0	0	1	112	−39	0
20	[t19, t20]	−22	0	0	0	26	0	0	0	65	0	31	0	0	−66	0	−32	0

Table 1. Differential Planning for Push-up

It is important to note that the movements of the edges caused by each joint's rotation propagate from the center toward the periphery of the humanoid's skeleton. On each of the five paths consisting of the head and the limbs of the humanoid, the center keeps its immediately distal joint, which keeps its immediately distal joint, and so on, … , keeps the endpoint, thus transitively. On these paths, some rotating or halting joints "causing" their immediately distal joints or endpoints "to go leftward/rightward" or "stay" in the sense of the causative analysis of Lakoff, G. (1970) and Parsons, T. (1985). Led by the causation of motion as the hint for representing the embedded motions in terms of joints'rotations as syntactic embedding, we constructed a kinematic tree, which has a five conjunct-coordinated structure. Each conjunct recursively embeds a transitive clause whose main verb is "keep"/"move leftward or rightward" having a complement as their objects.

Assuming that each joint is the subject of "keep" (cause to stay) or "move leftward or rightward" (cause to l/r-go), which takes an object complement, and that its distal joints and endpoint, as a whole, constitute its object enables us, for any joint, to form a pair of that joint and its remnant distal vertices (joints and endpoint). With this grammatical mechanism, we were enabled to represent embedded rotations as an embedded clauses. This makes it become easier to compare the grammar of language and that of action.

We mapped the humanoid' skeleton to a syntactic tree, whose initial symbol is the center of the skeleton, from which its head-part and limbs are hung down as five conjoined clauses, as was shown by

J. Ross (1967). As rewrite rules, two rule schemata, resembling X-bar schemata (Chomsky, N. (1970)), generate conjuncts: J→j J^{-1}, J^{-1}→ keep / l/r-move J'^{-1}. Here, j, where $J^{(')}$, and $J^{(')\,-1}$ act like NP, S, VP respectively. The differential rotation degrees of each joint in the table of the motion planning for pushing-up are underrepresented by the verbs, "keep" and "l/r-move", showing only the counterclockwise/clockwise direction of its rotation. Representing the remnant vertices distal to a joint, j, as a retracted ordered subset of vertices (center, joints, and endpoints), $V^{-1}=(V,<)-\{v_1\}$, enabled us to treat the causation of edge movement by a joint's rotation as a cyclic phenomenon and embed a rotation into another. Joints are named, and semantically translated as individuals, and each verb in the tree is translated as λ*P*/*y** [λ*x*[*keep'*/ *l/r-move'*(*P*/*y**)(*x*)]](*x*: subject variable, *y*: lowest verb object variable, *P*: embedded object clause variable). In each of the five conjuncts, we cyclically apply the lambda formula to every clause in a bottom up way, replacing its variable with the constants in it (Montague, R. (1974), as is shown in *Figure 1*. Conjoining the conjunct formulas leads to the total representation of one posture in pushing-up. To represent its semantics of pushing-up, or that of any of other actions, we connect sequentially such kinematic trees, following the planned order, where kinematics is shown by the distributional change of three modes of rotation over the joints.

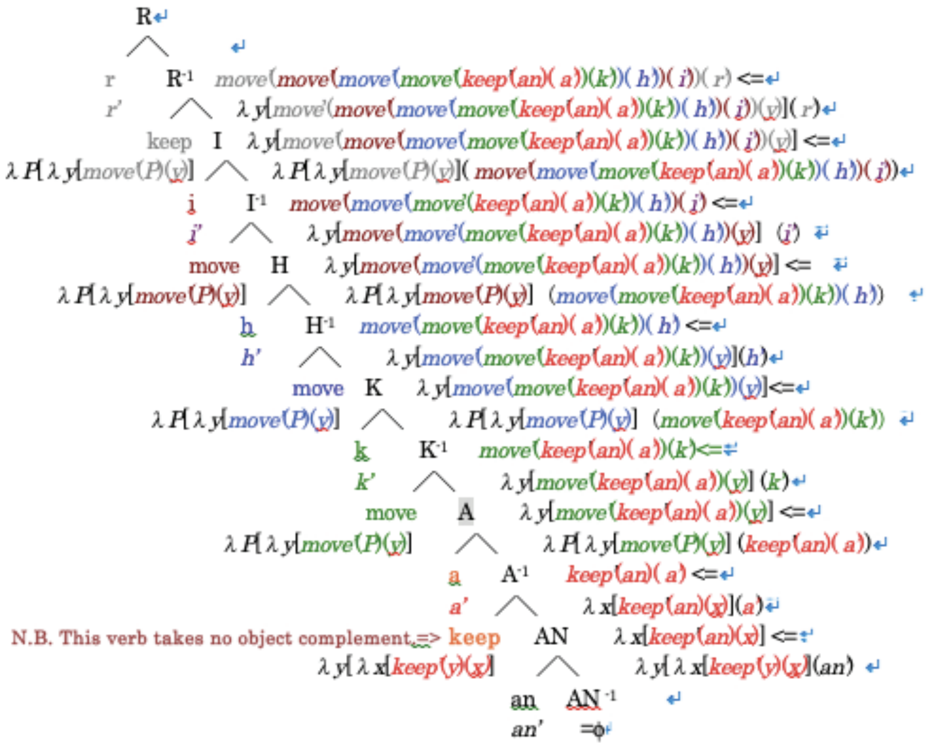


Figure 1. Partial Syntax and Semantics from the Differential Panning for an Action (for the Left Leg)

Okanoya (2004) showed that Bengalese finches' love songs has a "grammar", which is regular, meaning that its complexity is not high. Berwick, R. C, Okanoya K., Beckers G. J. L., Bolhuis J. J. (2011), however, pointed out that considering terminal symbols as semantic units is important in evaluating grammars. Song variants do not result in distinct meanings with completely new semantics. Fitch and Hauser (2004) reported that the signals to which tamarines were able to respond as the cue of being fed turned to be the strings, synthesized from human consonants and vowels, that were generated by a regular grammar. The "meaning" of all of the successful sentences functioned merely as the cue for being fed. As for actions, humans are able to assign them different meanings. Particularly, if they are human, it is the case. Human actions are, of course, distinguishable by the difference in the motions of the part(s) of their performers in one posture or in the whole action. Having this rich semantics, a kinematic grammar such as we constructed or the one in a larger scale seems to be possible as an origin of language, though they are not so strong in their complexities.

Are AI machine learning algorithms applied to text analysis enhancing prejudices? A critical discussion

Teresa Numerico[1]

[1] *Università di Roma Tre*

AI Machine learning algorithms risk applying biases and stereotypes to their knowledge inferences, based on statistical language associations. The ratio they use is the definition of statistical vector relationships between words, grounded on their reciprocal distance, measured within the training set of texts. The Stanford Natural Language Processing Group created one of the most famous algorithms to process words and to measure their distance, in order to grasp their semantic relationships. Its name is GloVe (Global vectors for word representation). GloVe is trained on various text corpora such as Wikipedia and Gigaword. The problem is that using the distance between natural language words to train an unsupervised learning algorithm there is the concrete probability of embedding in its knowledge base all the biases and stereotypes that are hidden within the data sets (Caliskan *et al.* 2017).

There is, in fact, evidence that shows that even without explicit will, humans tend to implement in their language habits all sorts of bad practices related to stereotypes regarding persons of different ethnic origins (typically afro-americans), or females, or any other social minority. Everybody, it is suggested by a variety of empirical studies, is implicitly carrying some biases. In a psychological study, the authors demonstrated an increase in the probability of interest in identical CVs of prospective employees just by changing the ethnic origin of the first names of the candidates (Lavergne, Mullainathan 2004).

According to Gillespie (2014), the way the training set of data is conceived has a large unexamined impact on the result of the training of the related algorithms. The suggestion that AI algorithms are less biased than humans with respect to prejudices and stereotypes needs to be demonstrated; the evidence at the moment shows the exact opposite (Caliskan et al. 2017, Bolukbasi T. et al. 2016).

It is crucial, in my opinion, to cast some doubt on the epistemological credibility, on the transparency of methods and on the fairness of the unsupervised training outcomes of the algorithms. We need to set a control strategy that can demonstrate the effectiveness and safety of results, and, above all, their assumed neutrality when treating sensitive data; for example, when we choose new employees or when we decide on the qualifications of a person for getting a mortgage, or, even worse, when we ask a black box algorithm to suggest the duration of the sentence for a convicted person (Pasquale 2015).

There are some researchers that propose a supervised correction of the perverse effects of the training by a cognitive intervention whose aim is the explicit elimination of prejudices from the data set (Bolukbasi T. et al. 2016), while others are still confident in the capabilities of the algorithms to offer better judging performances compared to biased human decisions (Sumpter 2018, chap. 14).

It is fundamental to discuss the organization of these algorithms and, above all, the effects of the implicit biases embedded in training textual databases used as a reference standard. I think that we should be very cautious in trusting a blunt device for taking decisions in situations where it is unlikely that we completely understand its logical processes (deep learning algorithms methods imply that even their programmers admit that they ignore the details of the learning processes they implement), as Norbert Wiener himself, already in 1950, wisely suggested.

The aim of this paper is not to propose an explicit correction to the text analysis algorithms but to discuss a cultural orientation towards the unproven belief that algorithms are more neutral and more effective than humans in all areas of knowledge, and in particular in text analysis.

We have to remember that algorithms are programs that historically rely on the project of formal language that was one of the core elements of formal logic of last century. The refinement of formal language proposed by logicians (such as Gottlob Frege, David Hilbert, Bertrand Russell) had the clear objective of getting rid of natural language within mathematics. According to those logicians, natural language was not a trustworthy tool for preserving information in the context of demonstration. It is then unlikely that software, which is directly inspired by formal logic practices about the definition of a formal language, could offer a correct interpretation of text.

In textual analysis the interpretative process is very important. Natural language understanding is based on the adoption of a critical approach, which is maybe the major objective of the long years of formative education for human beings. Natural language is contextual, multi-layered and ambivalent, and only after a long formative experience are human beings able to deal with it proficiently.

As Alan Turing had suggested as early as 1948, in order to display intelligence an agent needs to acquire two different capabilities: discipline and initiative. Text analysis is definitely based mainly on the initiative function, because it implies an efficient exercise of critical reasoning and subjective evaluation.

Even if human beings are not free from bias and prejudices when they represent the world, human cerebral plasticity, the habit to recognise and change the mistakes they commit and human cognitive openness, mean that it is likely that trained human beings are the best candidates to interpret a text correctly. As O'Neil (2016) advocated, humans can efficiently discard behavioural patterns that they previously followed. Human beings are mobile in their choices. They can change their minds after a deep investigation into the causes of their former beliefs and justifications. Ambivalent, context-sensitive natural language needs to be analysed with a high level of flexibility that algorithms cannot offer due to their formal language rigidity.

Moreover it is impossible to underestimate the role of programmers in the creation of the algorithms. They are human beings, and, as Michael Polanyi suggested in 1966, they often make decisions on the basis of their tacit knowledge. They obey their implicit biases as they implement the rules with which algorithms ought to comply. In the case of GloVe project the implicit rule that gives meaning to the agency of the algorithm states that: the more that words are close to each other the more they convey a related concept. This rule is very simplistic when applied to the complexity, richness and tortuosity of natural language. The difference between algorithms and human beings is that, once established, it is impossible to change the behaviour of an algorithm except by reprogramming it, while human beings with their intellectual elasticity can react differently and make new decisions that are not dependant on past choices, without the need for reprogramming and without having to define the desired results in advance.

Reference list

Bolukbasi T., Chang K.W., Zou J., Saligrama V., Kalai A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Advances in neural information processing systems*, pp. 4349-57.

Caliskan A. Bryson J. Narayanan A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 183-186.

Gillespie T (2014). The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) Media Technologies: Essays on Communication, Materiality, and Society (pp.167-194). Cambridge, Mass: MIT Press.

Lavergne, M. Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American economic review* 94, n. 4: 991-1013.

O'Neil C. (2016). *Weapons of math destruction*. London: Penguin.

Pasquale F. (2015). *Black box society*. Cambridge Mass: Harvard Univ. Press.

Polanyi M. (1966). *The Tacit dimension*. Garden City, N.Y.: Doubleday.

Sumpter D. (2018). *Outnumbered*. London: Bloomsbury.

Turing A. M. (1948). Intelligent Machinery. Report, National Physics Laboratory, in B. Meltzer D. Michie (Eds.), *Machine intelligence*, 5 Edinburgh: Edinburgh Univ. Press, 1969: 3-23; in Copeland J. B (ed.) (2004). *The essential Turing*. Oxford: Clarendon Press. 410-432

Wiener N. (1950/1954). *The human use of human beings*. Boston: Houghton Mifflin.

A Cognitive Model to Understand Emotions

Amitash Ojha, Elisabetta Gola
University of Cagliari, Cagliari, Italy

Emotions are one of the most central and pervasive aspects of human life. Normal people 'feel' and 'express' a wide range of emotions. While emotions deepen and enrich our experience, they also have a profound effect on other cognitive functions such as decision-making, reasoning, language comprehension, etc. (Okon-Singer, Pessoa & Shackman, 2015; Pessoa, 2013) It has been argued that cognition and emotions are complementary to each other and one cannot be properly understood or modeled without understanding the other. While several cognitive processes are being modeled, emotion being the fundamental aspect of human nature cannot be overlooked - especially, at a time when the technology is moving towards advanced intelligent systems and smarter robots. Therefore, to develop 'human like intelligence' and for a 'qualitative human-machine interaction', it is important that machines are also trained to understand human emotions.

Human emotion understanding and its modeling is tough because of its complexity. But in last few years, several attempts have been made to train artificial agents to recognize human emotions using various cues such as facial expressions, gestures, language and different bio-signals (Jung, Lee, Yim & Kim, 2015; Zhang & Lee, 2012; Piana, Stangliano, Odone & Camurri, 2016; Trigeorgis et al., 2016)). Some of the attempts have been made to learn emotions from different modalities. Although these emotion recognition systems work well in certain conditions, they still cannot be said to have achieved their goals because none of them replicate or reflect empirically examined cognitive structure of human emotion processing.

Traditionally emotions have been seen as biological processes that fundamentally differ from cognitive functions. The prevalent idea is that cognitive functions have structure and emotions do not; thus emotions cannot be mathematically modeled. This dissociation between emotion and cognition is evident in psychological literature. To add to this, researchers focusing on the relationship between emotion and cognition have argued that 'affect' and 'cognition' are separate and partially independent systems. However, in last few decades this view has been challenged and researchers have tried to outline the cognitive structure of emotion suggesting a vital role of emotion in overall cognition (Leventhal, 1984). Some have even gone to the extent of arguing that emotional processes cannot be separated from basic cognitive functions; they are highly intertwined Lazarus, 1982).

With this background, the primary goal of this paper is to present an emotion understanding model, which is based on the widely accepted Leventhal's cognitive model of emotion processing (Leventhal, 1982). Leventhal's multilevel process theory of emotion derives from the perceptual motor model of emotion (Leventhal, 1984). Our proposed model replicates processes that are akin to processes assumed in this model to be active while processing emotions in human beings. We also discuss the structure of cognitive model of emotion processing and how our model replicates it.

References

Leventhal, H. (1982). The integration of emotion and cognition: a view from the perceptual–motor theory of emotion. In *Affect and cognition: the 17th annual Carnegie Symposium on cognition. Hillsdale (NJ)* (Vol. 7, pp. 121-156).

Leventhal, H. (1984). A perceptual-motor theory of emotion. *Advances in experimental social psychology, 17*, 117-182.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American psychologist, 37*, 1019.

Okon-Singer, H., Pessoa, L., & Shackman, A. J. (2015). *The neurobiology of emotion-cognition interactions*: Frontiers Media SA.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration*: MIT press

Jung, H., Lee, S., Yim, J., Park, S., & Kim, J. (2015). Joint fine-tuning in deep neural networks for facial expression recognition. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2983-2991).

Zhang, Q., & Lee, M. (2012). Emotion development system by interacting with human EEG and natural scene understanding. *Cognitive Systems Research, 14*, 37-49.

Piana, S., Staglianò, A., Odone, F., & Camurri, A. (2016). Adaptive body gesture representation for automatic emotion recognition. *ACM Transactions on Interactive Intelligent Systems (TiiS), 6*, 6.

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on* (pp. 5200-5204): IEEE.

Fake news about fake news: a socio-cognitive perspective on four myths of online disinformation

Fabio Paglieri[1]
[1]*Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Roma*

One of the most problematic and controversial impacts of cognitive technologies on societies is online disinformation (Del Vicario et al., 2016): in its 2013 Global Risks report, the World Economic Forum devoted a chapter to outline the danger of "Digital Wildfires in a Hyperconnected World", and since then the buzz about fake news has steadily increased (Floridi, 2014). As a side effect, ironically yet not surprisingly, there are now fake news about fake news. This paper argues against four widespread equivocations:

1. *Fake news are a contemporary plague*. This is historical fiction: false information has always circulated in human societies, and its social impact was no less in the past than it is today. The Constitutum Constantini, a forged document used for centuries to justify the temporal power of the Catholic Church, is just the most obvious example (see also Buonanno, 2009).
2. *Fake news are becoming more frequent in our information ecology*. False information is certainly more abundant nowadays than in the past, but is its increase greater than the increase in information in general, be it accurate or inaccurate? There is no support for a positive answer to the latter claim; on the contrary, it has been argued (Hemp, 2009; Gleick, 2011) that the real challenge for our cognitive system is the amount of information we are exposed to, not its quality.
3. *Fake news are a cancer to be eradicated*. The most frequent reaction against fake news is extermination: either by fighting them online, e.g. via debunking, or by punishing them through legal means, e.g. as recently proposed in Italy by two Senators of the past legislature (Zanda & Filippin, 2017). Unfortunately, effectiveness of debunking is dubious at best (Lewandowsky et al., 2012; Chan et al., 2017), whereas targeting online disinformation via legal means is at risk of trivializing the epistemic complexity of information assessment, jeopardizing freedom of expression, or both. In contrast, online disinformation is an unavoidable feature of free speech, one that needs to be valued as a pedagogical tool to teach people how to deal with it – much like reinforcing the immune system of children by controlled exposure to germs, rather than sterilizing their environment.
4. *Fake news are bad because they spread falsehood*. Falsehood has been around for a very long time and does not make fake news especially alarming nowadays. The real danger in contemporary disinformation is oversimplification: the dominant mantra about knowledge is that it is easy – easy to both access and assess. Hence the widespread disregard for experts, and the conviction that well-intentioned laypeople can do wonders in all areas: from science to politics, from education to world economy. This belief is mistaken and dangerous, yet it cannot be eradicated by an equally oversimplified alternative (basically, "shut up and trust your betters"). Some recent works on the psychology of reasoning (Paglieri, 2016; Mercier & Sperber, 2017) suggest a different treatment for online disinformation, one aimed at making people more aware of the complexities of knowledge.

Based on these considerations, I will try to articulate some practical suggestions on how our online epistemic practices could be usefully amended and improved. My take-home message can thus be summarized in a series of relatively simple maxims, or slogans:

- *Online reputation is fleeting – and that's good!* The fact that downplaying or hiding one's mistakes and blunders is relatively easy online is often quoted as a critical shortcoming of online interactions, since it facilitates all kinds of shady behaviours by minimizing the likelihood of detection and reproach. There is, however, an essential upside to the fleeting nature of online reputation: precisely because it frees us from several reputational concerns, it should also minimize the biasing effect of those concerns on our epistemic processes. Unfortunately, the relevance and stability of one's own online reputation is often overestimated by users: when interacting on social media, we perceive ourselves as being constantly in the public eye, even though most of the people we interact with have (i) minimal interest in us as stable partners, (ii) limited occasions to engage with us, and (iii) very short attention and memory span for online events. Thus many reputational concerns that handicap our social epistemology, e.g. by making us unduly resistant to publicly manifest a change of opinions or acknowledge a mistake, are ironically much less justified online than offline. Citizens need to be made aware of this paradoxical aspect of social media interaction: online reputation matters much less than we think.

- *Beware of friendly fire!* Most of the online disinformation we are exposed to reaches us through our "friends", not our "enemies", and that makes it harder to detect it as misinformation – both because we tend to trust a friendly source, and because online "friends" share with us values and worldviews, thus crafting the message in ways likely to elicit acceptance. But online friends are dangerous not only because they may unwillingly feed us bad information: on top of that, our discussion with them tends to replicate the narrow-minded, closeted, mutually reinforcing dynamics that we are quick to stigmatize in people that do not share our ideas. As a case in point, often the only difference in an online discussion among, respectively, no-vax and pro-vax people is in the standpoints they share and the sources they trust, not in how they talk to each other, nor in how they refuse to seriously engage with dissenting voices.
- *The method is the message!* Related to the previous point, it is essential to educate users in appreciating the importance of how they argue, and not just of what they stand for. How many times have you read online comments sounding as follows: "These people are braindead morons! They never listen to reason, instead they spend all their times throwing insults at us!"? The tragedy is that, in the heat of the debate, we are blind to the irony of such blatant self-contradictions. Relatedly, it is common to criticize others for their superficial interest in relevant sources, while at the same time failing to identify one's own sources, or maybe even citing sources that we have not verified first-hand – typically because sources, in online debates, are used more as ammunitions to win a fight ("this link will shut you up!") than as vehicles of shared knowledge. If we seriously want to stand for a better way of creating, assessing, and circulating information online, we must start from our own epistemic backyard, making sure our practices are beyond reproach, and thus teaching by example.
- *It's the journey, not the destination!* A big problem with the contemporary obsession with fake news is the underlying idea that the pot of gold at the end of the rainbow is truth, a mythical prize that online disinformation prevents us from achieving. On the contrary, the real danger of our times is the lack of effort that most people exhibit when assessing the information they are exposed to; but making that effort, as any professional researcher knows first-hand, does not guarantee true knowledge, and even less diminishing uncertainty – indeed, most of the times the output of our quest for knowledge are just more (and better) questions. This is not problematic, however, because the real reward is the refinement of our cognitive capabilities, which can only be achieved by putting them to the test of rigorous, time-consuming, thought-provoking processes of information gathering and critical evaluation. Those processes ought to be the focus of our concern, not the "quota of truth" coming out of them.

References

Buonanno, E. (2009). *Sarà vero. Falsi, sospetti e bufale che hanno fatto la storia*. Torino: Einaudi.

Chan, M., Jones, C., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28(11)*, 1531–1546.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America, 113(3)*, 534.

Floridi, L. (2014). *The fourth revolution. How the infosphere is reshaping human reality*. Oxford: Oxford University Press.

Gleick, J. (2011). *The information: a history, a theory, a flood*. New York: Pantheon Books.

Hemp, P. (2009). Death by information overload. *Harvard Business Review, 87(9)*, 82–89.

Lewandowsky, S., Ecker, U., Seifert, C., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest, 13(3)*, 106–131.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge: Harvard University Press.

Paglieri, F. (2016). *La cura della ragione*. Bologna: Il Mulino.

World Economic Forum (2013). *Global Risks 2013, 8th edition*. Geneva: World Economic Forum.

Zanda, L., & Filippin, R. (2017). *Norme generali in materia di social network e per il contrasto della diffusione su internet di contenuti illeciti e delle fake news*. DDL S. 3001, disegno di legge presentato al Senato il 14 dicembre 2017.

Rhythmic and Morphosyntactic predictions: The anticipation skills of Italian children with developmental dyslexia

Valentina Persici, Natale Stucchi and Fabrizio Arosio

Università degli Studi di Milano-Bicocca

In both music and language incoming input can be predicted on the basis of the features of the input that has already been parsed. These types of structure-based predictions take place due to mechanisms of 'online' integration triggered by the structural combinatory features of the information already processed. Recently, it has been discussed whether these anticipation mechanisms, which are thought to be relevant for the acquisition of language-related skills such as reading and writing, are shared in language and musical rhythm (Guasti, Pagliarini, & Stucchi, 2017). This possibility may contribute to explain why deficits in reading and writing are associated with deficits in rhythmic processing, as it has been shown in children with Developmental Dyslexia (DD; Thomson et al., 2006; Thomson & Goswami, 2008).

We investigated how morphosyntactic and rhythmic information triggers anticipation in language and music by administering a picture selection task and a rhythmic task to a group of 15 Italian monolingual children with a diagnosis of DD and to 15 age-matched controls (TD) (mean age: 10;1 years, SD=1.09 years). In the picture selection task children were presented with two pictures and heard sentences in which a gender-marked clitic pronoun or determiner provided information on a following agreeing noun (*tocca* <u>*la*</u> *grande mela* - touch the^{fem} big apple^{fem}); we manipulated the gender information associated with the nouns in three conditions: phonologically unmarked (G), phonologically marked (GP), phonologically marked and semantically determined (GPS); we measured accuracy and response times. In the rhythmic task children heard sequences of sounds, in which a warning tone provided information on the occurrence of a future sound (called the imperative). Participants were asked to tap in time with the imperative.

Preliminary analysis revealed that the two groups did not differ significantly in the determiner processing task. However, in the clitic processing task, while DD children were slower in processing morphosyntactic information critical for anticipation in the G ($p < .01$) and GP ($p < .05$) conditions as compared to TD children, the two groups were equally efficient in using the semantic information ($p = .4$). Interestingly, DD children were also less accurate than TD controls in the rhythmic task. Moreover, only TD children exhibited a consistent negative asynchrony as adults do (Aschersleben, 2002).

Results suggest that DD and TD children make use of different anticipation strategies: while morphosyntactic combinatory information is used by TD children, children with DD seem to rely on semantic information. Furthermore, the difficulties encountered by DD children in using structure-based anticipation strategies do not seem to be limited to language, but to also extend to non-linguistic domains when the same type of anticipation skills are required.

Figures

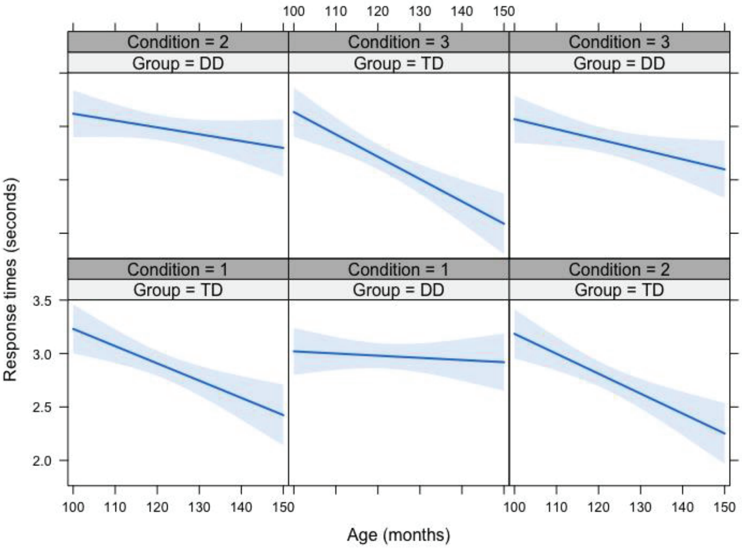


Figure 1. Determiner processing task: performance of the TD and DD groups in the three experimental conditions.

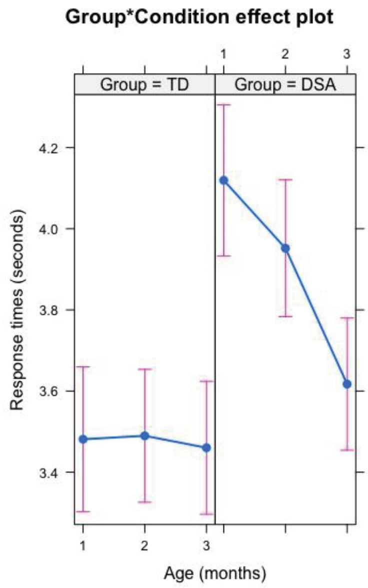


Figure 2. Clitic processing task: performance of the two groups in the three conditions.

References

Aschersleben, G. (2002). Temporal Control of Movements in Sensorimotor Synchronization. *Brain and Cognition* 48: 66-79. https://doi.org/10.1006/brcg.2001.1304.

Guasti, M. T., Pagliarini, E., & Stucchi, N. (2017). Language, reading and motor control: get rhythm! *Biolinguistics*, *11*, 1-12.

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading* 29, 334-348. https://doi.org/10.1111/j.1467-9817.2006.00312.x.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris* 102, 120-129. https://doi.org/10.1016/j.jphysparis.2008.03.007.

Wintertime Adaptation

Alessio Plebe
Department of cognitive Science – University of Messina
aplebe@unime.it

Recurrent periods when AI fell out of favor came to be known as "AI winter". One of the most severe "winters" came by mid 1980s, when the overly enthusiasm from some AI scholars and the enormous journalistic attention were rapidly waning. AI was widely said to have failed. Then AI resurged in the 1990s when artificial neural networks burst onto the stage, but after the new hype a second "winter" loomed over at the beginning of this century.

Seemingly, AI has never enjoyed a fortunate period like today. Since 2014 AI deserved the covers of journals such as *Science*, *Nature*, *The Economist*. For Klaus Schwab (2016), Executive Chairman of the *World Economic Forum*, AI is driving an unprecedented revolution in industry, as well as in the social assets of most countries in the world. According to CBInsights, worldwide investment in private companies on AI has grown from 589 million dollars in 2012 to over 5 billion in 2016.

Regardless of this apparent healthiness, murmurs of disappointment are being heard once again. The journalists are chasing again cases of AI failures, ludicrous like Facebook algorithm censoring the Venus of Willendorf as pornography, or dramatic like the deaths caused by Tesla and Uber autonomous vehicles. Concerns derive from economical analysis too. Brynjolfsson et al. (2017) title their analysis *AI and the Modern Productivity Paradox*, where the paradox is the clash between the impressive performances of AI in many domains, and the decline of measured productivity growth in the last two decades.

Signs of a new upcoming winter? This might be the case for computer vision scientist Amnon Shashua, founder of Mobileye company which equipped Tesla with the AutoPilot system. He foresees the risk of a "AI winter" in the domain of autonomous vehicles (Shalev-Shwartz, Shammah, & Shashua, 2017).

In this paper I hold that AI has developed a degree of tolerance to seasonal variations such that, even if a new winter would happen, AI will readily survive it. My confidence results from several reflections that will be detailed in the paper, in the remaining of this abstract I will address only one, which I deem as the most important.

The main reflection results from a scrutiny of the progress that enabled the current success in AI and a comparison with previous periods. The current progress stems from algorithms collectively called *deep learning* (Hinton & Salakhutdinov, 2006), a direct derivation of the venerable *Parallel Distributed Processing*, to which Hinton himself contributed. The arrival of PDP triggered the harsh debate between "symbolic" and "connectionist" approaches (Fodor & Pylyshyn, 1988), which divided the AI community, and was detrimental of its overall evolution. The tone and temper of the discussion around deep learning have drastically changed. Instead of a polarity in terms of "symbolic" and "connectionist" it is more appropriate the philosophical polarity between empiricism and rationalism, with the former grounding deep learning and the latter grounding various forms of rule-based AI. For sure, few scholars on the rationalist site align themselves with those presaging another winter, a distinguished representative is Gary Marcus (2018). Marcus lists ten shortcomings of deep learning, all in some way related to the lack of rationality, such as the inability to learn from explicit definitions or to distinguish between causation and correlation. Kotseruba and Tsotsos (2018), in a recent review of about hundred AI cognitive architectures, while including several models based on neural networks, deliberately excluded deep learning "Since the current deep learning techniques are mainly applied to solving practical problems and do not represent a unified framework".

Notwithstanding, I found distinctive of the current contraposition that many scholars at both sides of AI are now acknowledging the value of each other and the need for an integration. For example, from the rationalist side, Lake et al. (2017) clarify that "Although we are critical of neural networks in this article, our goal is to build on their successes rather than dwell on their shortcomings. We see a role for neural networks in developing more human-like learning machines". On the opposite side, several of the leading inventors and developers of deep learning are wary of the limitations of a pure empiricist framework. For example, François Chollet (2018) writes that "In general, anything that requires reasoning – like programming or applying the scientific method – [...] is out of reach for deep-learning models".

This sort of awareness empowers AI to adapt to possible winters, by promoting cross-fertilization between the highly developed and successful frameworks of the empiricist party, and the rich repertoire of ideas and models on offer from the rationalist ally.

Let me give just few examples. *PathNet*, jointly developed by DeepMind and David Ha of Google Brain (Fernando et al., 2017), addresses the critical issue of reusing large neural networks for different tasks. The strategy is loosely borrowed from the society of agents notion, popular in AI. In this case agents are embedded in the neural network, with the task of discovering which parts of the network to reuse for new purposes. David Ha moved furthr with Jürgen Schmidhuber, the inventor of long-short-term memory networks, to propose *World Models*, targeting a concept well established in cognitive science, that of "mental models" of the world (Ha & Schmidhuber, 2018). They combined the reinforcement learning framework with recurrent neural networks and variational autoencoders. The resulting model is capable of learning a variety of virtual worlds by unsupervised observations. A differnt direction in integrating the rationalist and empiricist souls is pursued, again by a collaboration between DeepMind and Google Brain (Santoro et al., 2017), with *Relational Networks*. These peculiar networks are endowed with a structure primed for relational reasoning, without needing to be learned, and are integrated with other deep learning components.

References

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (Tech. Rep. No. 24001). National Bureau of Economic Research.

Chollet, F. (2018). *Deep learning with python*. Shelter Island (NY): Manning.

Fernando, C., Banarse, D., Blundell, C., Zwols, Y., Ha, D., Rusu, A. A., ... Wierstra, D. (2017). PathNet: Evolution channels gradient descent in super neural networks. *CoRR*, *abs/1701.08734*.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, *28*, 3–71.

Ha, D., & Schmidhuber, J. (2018). World models. *CoRR*, *abs/1803.10122*.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *28*, 504–507.

Kotseruba, I., & Tsotsos, J. K. (2018). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 1–78.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Science*, *40*, 1–72.

Marcus, G. (2018). Deep learning: A critical appraisal. *CoRR*, *abs/1801.00631*.

Santoro, A., Raposo, D., Barrett, D. G., Malinowski, M., Pascanu, R., & Battaglia, P. (2017). A simple neural network module for relational reasoning. In *Advances in neural information processing systems* (pp. 802–810).

Schwab, K. (2016). *Fourth industrial revolution*. Geneva: World Economic Forum.

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. *CoRR*, *abs/1708.06374*.

Explaining the behavior of learning classification systems: a black-box approach

R. Prevete, A. Apicella, F. Isgrò, G.Tamburrini
DIETI, University of Naples Federico II

Introduction

Many Machine Learning (ML) techniques give rise to automatic systems whose behavior is often hard to interpret. Various approaches to the problem of overcoming their opaqueness are now being pursued (Lipton Z. C., 2018, Montavon, G., Samek, W., & Müller, K. R., 2017). This problem is theoretically challenging and practically important. From a practical standpoint, interpreting the behavior of learning systems is crucial to prevent or fix errors in high- risk or ethically sensitive ML applications.
In the context of this multifaceted interpretability problem, we focus on the issue of what it is to explain the behavior of ML systems for which only Input-Output relationships can be computed (black-box). In literature, this type of explanation is known as *model agnostic explanation* (Ribeiro, M. T., Singh, S., & Guestrin, C., 2016). Model agnostic explanations are usually given by exhibiting a class *prototype* which the input data were associated to (Ribeiro, M. T., Singh, S., & Guestrin, C., 2016; Montavon, G., Samek, W., & Müller, K. R., 2017), and explanation requests are usually expressed as why-questions: "Why were input data associated to class C?". However, prototypes often have a poor explanatory power. For example, if an ML system responds that an image x given as input belongs to the class "cat", by providing as explanation a cat-prototype, one would not be put in the position to understand what features (parts) of the prototype are associated to what features (parts) of x. We propose to mitigate this poor level of understanding by producing explanations on the basis of dictionaries of local regions (atoms) of the input spaces, which are computed by unsupervised ML approaches (Tessitore, G., & Prevete, R., 2011). These atoms represent humanly interpretable features of the input data, and both prototypes and input can be reconstructed as linear combinations of atoms. Moreover, we provide both atoms that contribute to clarify the system's answer and atoms related to discarded classification options. Discarded options are needed to cope with the contrastive character of explanation requests, usually interpreted against the background of alternative and actually discarded classification possibilities. The model is exemplified by a discussion of experimental results concerning model agnostic explanations for a specific Deep Learning classification system.

Proposed approach

Our proposed model can be functionally described in terms of a three-entity framework: 1) *Oracle*: a learning system classifying input data, whose inner workings are unknown (black-box approach); 2) *Interrogator*: usually a human being, requesting explanations about oracular answers; 3) *Mediator*: another algorithmic system helping Interrogator to understand the Oracle's behavior by providing explanations built from atoms of input spaces. The Mediator provides explanations such that both prototypes and input can be reconstructed as linear combinations of atoms. Atoms are stored in *dictionaries* which are computed using a sparse dictionary learning technique (Bao, C., Ji, H., Quan, Y., & Shen, Z., 2016) insofar as these techniques provide data representations that are often found to be accessible to human interpretation. To this end we use a Sparse Non-negative Matrix Factorization (NMF) method (Lee, D. D., & Seung, H. S., 2001; Hoyer, P. O., 2004). Once the dictionary is computed the Mediator uses an Activation-Maximization-based technique, combined with NMF to provide an explanation in terms of atoms of the dictionary.

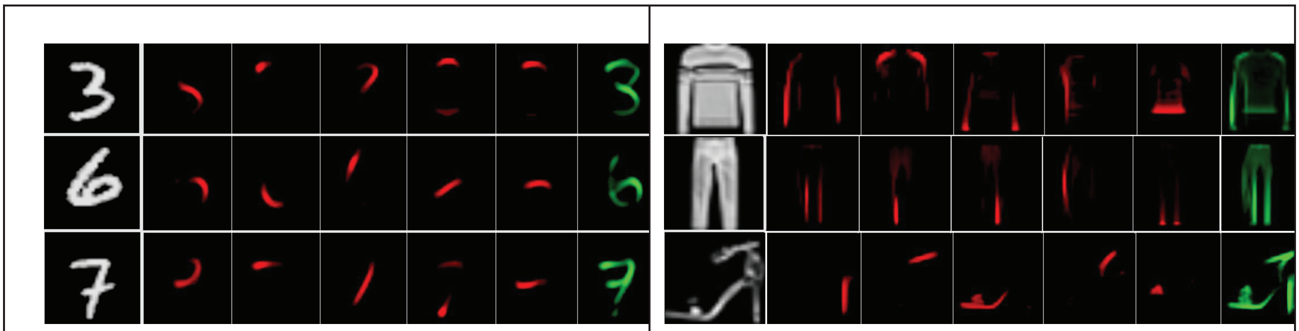


Figure 1: Examples of explanations. See text for further details.

Experiments and results

To test our proposed framework, we choose as Oracle a convolutional neural network architecture, LeNet-5 (LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., 1998). The network has been trained using two different datasets: MNIST and Fashion MNIST. Training and test set are composed of 50000 and 10000 images, respectively. To find the dictionaries, we use NMF with sparseness constraints. We set the number of atoms to 200 relying on PCA analysis. We choose the dictionaries having the best trade-off between sparsity level and reconstruction error.

In Figure 1 (a) we show the atoms with higher encoding values (i.e. that are more ``important" in the representation) on three inputs belonging to MNIST dataset, a '3', a '6' and a '7', on which the Oracle gives the correct answer. The chosen atoms seem to be very representative to describe the visual impact of the input digits providing elements that seems to be discriminative (the great presence of curved elements for the '3' and straight lines for the '7'). Similarly, in Figure 1 (b) we show atoms as explanations of the correct answers of the Oracle on Fashion MNIST input.

Conclusions

In this paper, we proposed a model-agnostic framework to explain the answers given by a classification system. In our approach explanations are provided in terms of atoms representing humanly interpretable features of the input data. The experiments conducted show encouraging results providing explanations that seem to be qualitatively significant.

References

Bao, C., Ji, H., Quan, Y., & Shen, Z. (2016). Dictionary learning for sparse coding: Algorithms and convergence analysis. *IEEE transactions on pattern analysis and machine intelligence*, *38*(7), 1356-1369.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, *5*(Nov), 1457-1469.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems* (pp. 556-562).

Lipton Z. C. (2018). The mythos of model interpretability. Queue, 16(3):30.

Montavon, G., Samek, W., & Müller, K. R. (2017). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*.

Tessitore, G., & Prevete, R. (2011). Designing structured sparse dictionaries for sparse representation modeling. In *Computer Recognition Systems 4* (pp. 157-166). Springer, Berlin, Heidelberg.

A Dynamic Cognitive Model of Bilingual Mental Lexicons

Qian Cheng
Université Paris Nanterre - UMR 7114 - Modèles, Dynamiques, Corpus (MoDyCo)

Commenting on insufficiently defined concepts in the realm of Cognitive Psychology and Cognitive neurosciences, Poldrack and Yarkony (2015)[1] bring ontologies into consideration as a tool for specifying knowledge. The point made by the authors is that ontologies are currently the best solution to perform adequate knowledge representation, not only from a formal- and computational-, but also from a cognition-centred perspective.

My research discusses ontology approaches aimed at the modelling of concept sharing in the bilingual mental lexicon. Being ER (Entity / Relation) graphs allowing for annotations, formal ontologies are prone to undergo systematic enrichment, which makes them flexible and robust over time, but also entitles them to support inferences about information not directly encoded in the graph, if an appropriate language, such as OWL or Topic Maps ISO/IEC 13250 or Cypher is used to describe them.

As for bilingual mental lexicon models, Pavlenko (2009)[2] underlines that current models of language processing posit cross-linguistic sharing of meanings and/or concepts. Focusing on a particular type of common mistakes made by mandarin native speakers at a beginner stage of French learning, (namely, the transfer of classifiers in discourse production), I take particular interest in the « Revised Hierarchical Model »[3] [4] because it explicitly states concept sharing at an *early* stage of second language learning. The specific status of early stage L2 learning has been put forward by examining control mechanisms at different proficiency levels: Abutalebi and Green[5] report studies showing that conceptual input increasingly drives L2 responses as the second language knowledge grows, while at an initial learning stage, linking to L1 word forms prevails.

« Concept sharing » being a quite elusive notion within bilingual Mental Lexicon approaches, my research revives the distinction between « semantic » and « conceptual » structure[6] as a means to appropriately identify possible connections of classifiers to both linguistic meaning and language-independent knowledge. As mandarin classifiers specify inherent meaning features of nouns but can also bare a discriminating value when used with verbs, they can be thought to affect both the conceptual *and* the semantic level of lexical entry representations.

Regarding noun classifiers, Jin[7] establishes that their rôle is to provide a semantical partition unit allowing numerical counting, as plural inflection does not exist in Mandarin Chinese and nouns are inherently neutral with respect to the singular / plural distinction. I am here proposing, on grounds of mistakes occurring within discourse production of early sinophone French learners, viewed in the light of the Revised Hierarchical Model, that the information provided by classifiers should be attached to the the meaning component of lexical entries.

On the one hand, this modelling perspective could explain why, in discourse production, novice Mandarin-speaking learners, while accessing meaning of L2 lexical entries via their mother tongue equivalents (as suggested by the RHM), sometimes replicate the attached classifier literally. On the other, it accounts for a compositional view of meaning, including, in particular, the partition units mentioned in Jin's work.

As the same classifier can be shared by different lexical entries, network representations (such as ontologies) appear as the most appropriate tool for dynamical lexicon modelling: not only they allow to express component sharing, but also to semantically qualify the relation between nodes, a crucial issue in concept categorisation.

Figure 1 shows an excerpt of an ontology sample containing some action and object concepts relating to verbs and nouns. Conceptual structure components appear in squares (generic categories in dashed-lined ones, attributes of them in solid-lined ones). Language-specific Semantic Structure components, linked to lexemes, appear in ellipses (« * » stands for quantifiable). Only classifiers (in the range of given attributes of concept types) are shown, italics featuring lexeme-driven instances.

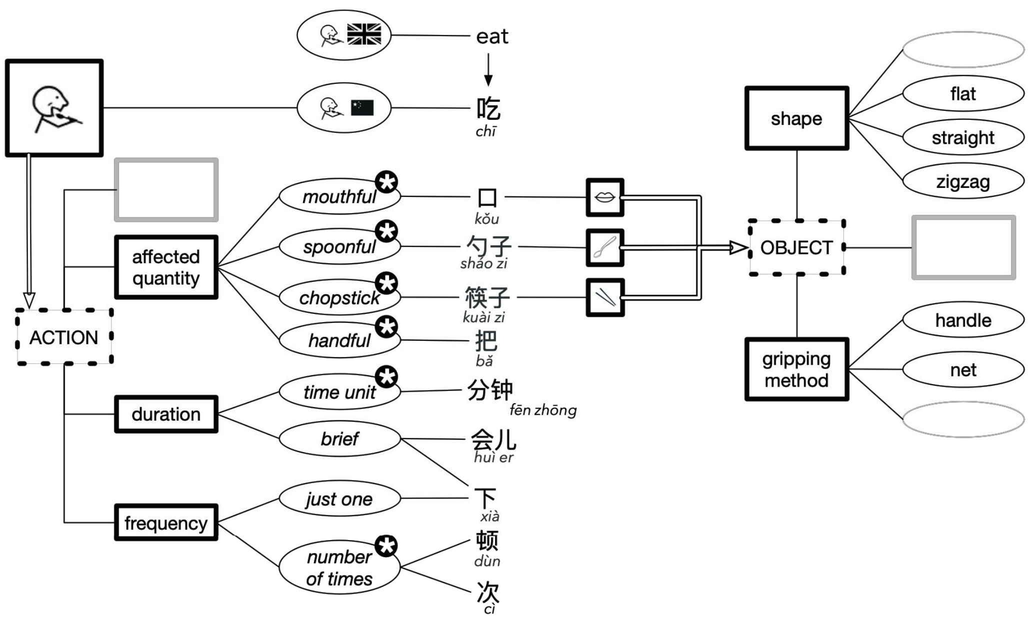


Figure 1 : A sample bilingual mental lexicon ontology excerpt

References

1 Poldrack, R., & Yarkony, T. (2015). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual Review of Psychology*, 67, 587–612. https://doi.org/10.1146/annurev-psych-122414-033729

2 Pavlenko, A. (2009). Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In Pavlenko, A. (Ed.), *The Bilingual Mental Lexicon Interdisciplinary Approaches* (pp. 125-159). Bristol, Buffalo, Toronto: Multilingual Matters

3 Kroll, J., & Steward, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language,* 33, 149-174. https://doi.org/10.1006/jmla.1994.1008

4 Kroll, J., Van Hell, J., Tokowicz, N., Green, D. (2010) The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition,* 13(3), 373–381.https://doi.org/10.1017/S136672891000009X

5 Abutalebi, J., Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics* 20(3), 242-275. https://doi.org/10.1016/j.jneuroling.2006.10.003

6 Evans, V. (2009). Semantic structure vs . conceptual structure : The nature of lexical concepts in a simulation-based account of language understanding. https://www.semanticscholar.org/paper/Semantic-structure-vs-.-conceptual-structure-%3A-The-Evans-Evans/4dbdbbe38bf63f78536d3c2ef4798d4d2663cb43

7 Jin, J. (2013). Functions of Chinese Classifiers: A Syntax-Semantics interface Account. Ph.D Thesis. The Hong Kong Polytechnic University. http://ira.lib.polyu.edu.hk/bitstream/10397/6480/2/b26527686_ir.pdf

Can the Human Brain be replicated?

Steven S. Gouveia[1]

[1]*PhD Student Philosophy, University of Minho (Portugal); Researcher Mind, Language and Action Group – Institute of Philosophy, University of Porto; Researcher Mind & Reasoning Group – NOVA University of Lisbon; Visiting Researcher Mind, Brain Imaging and Neuroethics of the Royal Institute of Mental Health, University of Ottawa.*

Abstract: Since the birth of Artificial Intelligence as a coherent field of research, the main goal has been to try to replicate the miracle of evolution, that is, to create a conscious human brain from a non-organic substrate. Some proposals (e.g. Mind-Uploading, cf. Gouveia, Steven & Neiva, Diana (2017), Chalmers (2014)) focuses more on the idea of transferring our mental life from an organic basis to a non-organic but equivalent in term of causal powers. Others try to focus on the replication of natural selection itself (cf. Biota.org). Digital Biota is a kind of autonomous software that self-replicates, and which is embodied in viruses, genetic algorithms, and general adaptive networks. The objects of the software interact with their environment, they are able to multiply, learn and change, being affected by "natural selection" whose rules have been programmed to evolve. On the other hand, some researchers think that both simulation or emulation of the brain (or the evolution itself) cannot be achieved (cf. Searle (1984), Nicolelis (2015)) since the brain has something special that cannot be replicated by any non-organic substrate. In this talk, we will evaluate if there are really some principles or meta-caveats that prevents Artificial Intelligence researchers to develop a full and functioning human brain. Our main goal is to understand if those caveats are real obstacles or just circumstantial walls that can be surpassed through the development of new and better technology. In particular, we will argue that some predictions made by Artificial Intelligence researchers and transhumanists in general seems to be more motivated by their own bias than by rigorous ideas and investigation. As we will show (Armstrong et all, 2014), the consequence of this attitude is the following: there is no difference in the predictions made by experts and non-experts about the future of Artificial Intelligence. We will finish the talk raising questions about the ethical problems that can arise if a real brain-emulation happens (cf. Gouveia & Teixeira, forthcoming).

References:

Chalmers, David (2014). Uploading: a Philosophical Analysis. *Inteligence Unbound: the Future of Uploaded and Machine Minds* (ed. Russell Blackford and Damien Broderick). Oxford: Wiley Blackwell.

Gouveia, Steven & Teixeira, João de Fernandes (eds.) (forthcoming). *Artificial Intelligence: a Mutlidisciplinary Perspective.* Wilmington: Vernon Press. Contributors: Eray Orzkul, Ben Goertzel, Alex Montes, Roman V. Yampolskiy, Susan Schneider, Frederic Gilbert, Stevan Harnad, Davide Pearce, Natasha Vita-More, Nicole Hall, Daniel Dennett and many more.

Gouveia, Steven & Neiva, Diana (2017). The Problem of Consciousness on the Mind Uploading Hypothesis. *Philosophy of Mind: Contemporary Perspectives*. Newcastle: Cambridge Scholars Publishing.

Nicolelis, Miguel & Cicurel, Ronald (2015). *The Relativistic Brain*. São Paulo: Kios Press.

Searle, John (1984). *Minds, Brains and Science*. Cambridge (MA): Harvard University Press.

Stuart Armstrong, Kaj Sotala & Seán Ó hÉigeartaigh (2014). The errors, insights and lessons of famous AI predictions – and what they mean for the future. *Journal of Experimental & Theoretical Artificial Intelligence*, 26:3, 317-342.

The effects of stochastic environmental variations and plasticity on evolutionary adaptation and developmental memory

Stefano Tiso[1],Nicola Milano[2],Stefano Nolfi[2]
[1]MEME Erasmus Mundus master in evolutionary biology,
[2] CNR Istituto di Scienze e Tecnologie della Cognizione

With this study we investigated the role played by environmental variations and plasticity in the adaptation of artificial evolving agents controlled by neural networks. This dynamic might indeed reveal how the evolutionary algorithm naturally favours and nurtures an embedded learning capability of genomes. Watson and Szathmáry in their review from 2016 list a consistent body of literature claiming that evolution strictly resembles a learning process. The two processes, for certain aspects, are indeed mathematically equivalent(Harper, 2009; Shalizi, 2009) and bear considerable conceptual similarities(Valiant, 2013; Watson, Wagner, Pavlicev, Weinreich, & Mills, 2014). They show as well, through a model made by Watson et al. in 2014, how evolution brings to a genomic architecture displaying a form of memory of past (evolutionary) events. This made these genomes able to re-adopt previous successful solutions, obtained during their evolution, in new and never-experienced situations where such solutions could be useful. Therefore, these evolved genomes were able to generalize, using previously acquired knowledge/adaptation to tackle new and unknown problems/selective pressures, much like what happens with learning in cognition. Watson called this observed phenomenon “developmental memory”, and compare it with “associative memory” displayed by simple cognitive systems. Coming from a separate line of research in evolutionary biology, in her book “Developmental plasticity and evolution” from 2003 (in particular in chapter 12) Marie Jane West-Eberhard states that shifting environment can induce changes in the phenotype as readily as mutation, and that this two kinds of variation are equally important for evolution and the process of adaptation. West-Eberhard’s explains how this happens thanks to two processes linked to plasticity, which she defines “phenotypic accommodation” and “genetic assimilation”. Phenotypic accommodation is the capability of an organism to operate changes at the phenotypic level during its lifetime following environmental clues and without altering its genome. In successive generations these changes can then be fixated in the genome, through the process of “genetic assimilation”. In this case, as the same environmental clue is presented multiple times through generations, the temporarily induced modifications become permanent (and so also independent from the environmental stimulus). Therefore, West-Eberhard accords to the environment a crucial role in both eliciting new adaptations and in (by repeatedly inducing the same change) making them permanent. This is a concept very seldom considered in evolutionary biology, where usually, the environment is seen only as a filter for novelties and not an elicitor of novelties itself. These two points (Watson & Szathmáry’s and West-Eberhards’) come from different areas of biology and stem from different reasoning, nonetheless, they could be brought together in a very interesting new line of inquiry. The capacity of organism of promptly react to environmental changes and eventually hardwire them into their genome, would naturally bring to the evolution of organism who are able to both retain past experience and generalize from them finding solutions to unknown challenges, especially so, if such organisms were evolved in a constantly changing environment. This would happen because these organisms would have to constantly shift their phenotype towards a new optimum while the environment keeps changing. Such conditions would create a selective pressure not only to obtain the best possible adaptation in the short term, but also to possess a genomic architecture that would be capable of ensuring efficient and consistent adaptation constantly to new situations in the long term. Therefore, a genetic line that is capable of “retaining” past experiences and rapidly re-deploy solutions achieved previously without the need to go through a process of trial and error every time to find them again, would have a clear advantage in the evolutionary run. More so, if it could use these past solutions to tackle new problems, therefore "generalizing" the evolutionary challenges thanks to their previous experience, as Watson and Szathmáry's article sustains. If these predictions reveal themselves to be true, we could affirm that variation in the environment is not only shifting the evolutionary optimum, leading the phenotype to adapt/assume a new form, but is also pushing the genome to be more efficient in finding these solutions, or in other terms it is facilitating its "learning/evolutionary" process through generations. This objective will be pursued by verifying whether and in which conditions artificial agents evolving in varying environments achieve a higher adaptation level than agents evolved in stationary environments, and if this make them better as well in never-experienced environments. We will as well see if plasticity, and in particular "phenotypic accommodation" plays an important role as we expect in this evolutionary process. Initial studies from Laral laboratory in the ISTC center of the CNR in Rome show how this direction might be an interesting path to follow(Milano, Carvalho, & Nolfi, 2018). In this study we will also bring the model closer to a biological reality by making the morphology of the robot and not its behaviour controlled by the evolving neural network, and using soft material cells to build its body instead of the more classical, but less biologically accurate, rigid arms and joints. The agents will be then selected and evolved for the ability to perform a simple locomotory task and will be exposed to different environments which will affect their ability to perform their task (in our case we will repeatedly expose the genome to both a land and an acquatic environment). We will evolve two kind of populations, one (the treatment) which is able to perceive the external environmental conditions as it evolves, and therefore being potentially able to react consequently, and one which is not (the control). Our study shows interesting first results that might induce us to conclude that this hypothesis could be correct, indeed showing that plasticity plays a significant role in this process and is able to produce "generalizing genomes", bringing cognitive sciences and evolutionary biology a step closer.

Harper, M. (2009). The Replicator Equation as an Inference Dynamic. Retrieved from http://arxiv.org/abs/0911.1763

Milano, N., Carvalho, J. T., & Nolfi, S. (2018). Environmental variations promotes adaptation in artificial evolution. *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings*, *2018–January*, 1–7. https://doi.org/10.1109/SSCI.2017.8280909

Shalizi, C. R. (2009). Dynamics of Bayesian Updating with Dependent Data and Misspecified Models. https://doi.org/10.1214/09-EJS485

Valiant, L. (2013). Probably approximately correct: Nature's Algorithms for Leaning and Prospering in a Complex World.

Watson, R. A., & Szathmáry, E. (2016). How Can Evolution Learn? *Trends in Ecology & Evolution*, *31*(2), 147–157. https://doi.org/10.1016/j.tree.2015.11.009

Watson, R. A., Wagner, G. P., Pavlicev, M., Weinreich, D. M., & Mills, R. (2014). The evolution of phenotypic correlations and "developmental memory". *Evolution; International Journal of Organic Evolution*, *68*(4), 1124–38. https://doi.org/10.1111/evo.12337

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press. Retrieved from https://books.google.it/books?hl=en&lr=&id=7DQNTPYaHlYC&oi=fnd&pg=PR15&dq=developmental+plasticity+west+ebherard&ots=X-0UgQR5kE&sig=hbBzEXLSdeMgiPqY52sdAPTKvcA&redir_esc=y#v=onepage&q=developmental plasticity west ebherard&f=false

Linguistic, social and interoceptive experiences during the processing of abstract concepts

Caterina Villani[1], Luisa Lugli[1], Fabrizio Campanile, Marco Tullio Liuzza[2], Anna M. Borghi[3,4]

[1]Department of Philosophy and Communication, University of Bologna;[2]Department of Medical and Surgical Sciences, "Magna Graecia" University of Catanzaro; [3]*Department of Dynamic and Clinical Psychology, Sapienza University of Rome;* [4]*Institute of Cognitive Sciences and Technologies, Italian National Research Council, Rome.*

Introduction

The topic of the representation of abstract concepts, i.e. of concepts that do not possess a single, concrete referent, like "freedom", is widely debated in cognitive (neuro) science (see special issues: 1; 2). According to multiple representation theories, abstract concepts are not only grounded in sensorimotor system but evoke more linguistic, social and inner experiences than concrete ones. Within these views, the Word As social Tools (WAT) view (3; 4) proposes that the activation of linguistic experience during abstract concepts processing leads to the involvement of the mouth motor system. However, abstract concepts are quite heterogeneous and can substantially vary in the experiences they re-enact (5; 6). Our study aims to test the differences within abstract concepts investigating the extent of involvement of linguistic, social and interoceptive experiences during their processing. More specifically, we hypothesized that the gum, social and inner condition interfere with the activation of the linguistic (mouth involvement), social and interoceptive experience, respectively. If this is true, then the pleasantness scores for abstract concepts should decrease, while difficulty scores should increase as compared to the ball condition. The interference in the mouth condition should be stronger for abstract concepts that obtained high scores in dimensions relevant for language (Age of Acquisition (AoA) and Modality of Acquisition (MoA) (7); the interference in the inner condition should be more pronounced in the concepts with high scores of emotionality (8) and interoception (9), while the interference in the social condition should be more pronounced for concepts scoring high in social valence (10) and social meta-cognition (1).

Method and Results

Participants

One hundred twenty-nine students (102 female, 18 left handed, Mean age: 24.2; SD age: 3.7) from the University of Bologna participated in the study.

Procedure

We adopted the paradigm from a previous study (11) in which we found that participants evaluated abstract concepts less pleasant and more difficult than concrete ones when actively moving the mouth (chewing a gum) but not the hand, likely because this interferes with the mouth activation during abstract concepts processing.

We asked participants to evaluate difficulty and pleasantness of abstract words, touching a 5-points-scale on the computer screen. We designed four conditions manipulated between participants: during the rating task participants were required either to squeeze a ball (ball condition), to chew a gum (gum condition), to hold the hand of their desk mate (social condition) or to hold a hot/cold bag (inner condition). Each experimental condition is tailored at producing selective interference in processing distinct abstract categories.

Materials

Stimuli were 60 abstract concepts, selected from a database in which different kinds of abstract concepts were scored on a variety of dimensions and then divided into the following subgroups: philosophical and spiritual (PS) (e.g. mystery), self and sociality (SS) (e.g. ability), emotive/inner states (EI) (e.g. anger), physical, spatio-temporal and quantitative concepts (PSTQ) (e.g. addition). (5; 6).

Results

Generalized estimated equations (GEE) models with logistic links function were applied and the factors taken into consideration were Rating (Complexity, Pleasantness) as within participants factors and Condition (ball, gum, social, inner) as between participants factor. As predicted, abstract words in gum condition were rated more complex compared to the ball one (p <.001), while the rating of abstract concepts in the gum condition and those of social and inner conditions did not differ (p_s >.221). Conversely, in pleasantness rating abstract word were rated less pleasant in gum condition than in ball one (M = 3.2; M = 3.3). Abstract concepts were valued more pleasant in the inner condition than in social one (p =.077).
Since we are interested in the contribution of different types of experience in representing specific abstract concepts, we applied a further GEE analysis, taking into consideration the four different kinds of abstract concepts (PS, SS, EI, PSTQ) as within participants factor. PS concepts were significantly rated as more difficult in gum condition compared to the other kinds of concepts in all conditions (p_s <.001); while pleasantness scores decrease as compared to SS concepts in all conditions (p_s <.003).
Furthermore, SS and EI concepts were significantly rated ad more difficult in the inner condition compared to ball one (p_s <.001); while pleasantness scores in social and inner condition decrease significantly for EI concepts compared to SS one (p_s =.003).

Discussion

The study reveals interference effects on abstract concepts processing. Overall, the active involvement of the mouth motor system during processing of abstract words slightly increases perceived difficulty and lowers perceived pleasantness of words. This result support multiple representation views, according to which linguistic, social and inner experience is pivotal for representation of abstract concepts. More crucially, multiple dimensions concur in representing specific abstract concepts: linguistic experience is more pronounced for PS concepts, instead social and inner experiences is more pronounced for SS and EI concepts. This finding highlights the importance to investigate the varieties of abstract concepts.

References

1. Borghi, A.M., Barca, L., Binkofski, F., & Tummolini, L. (2018). Abstract concepts, language and sociality: from acquisition to inner speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0134
2. Bolognesi, M., & Steen, G. (2018). Editors' Introduction: Abstract Concepts: Structure, Processing, and Modeling. *Topics in Cognitive Science*.
3. Borghi, A.M., & Cimatti, F. (2009). Words as tools and the problem of abstract word meanings (Vol. 31). Retrieved from http://escholarship.org/uc/item/58m9n8rp.pdf
4. Borghi, A.M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts*. Springer.
5. Villani, C., Lugli, L, Liuzza, M.T., Borghi, A.M. (in press). Different Kinds of Abstract concepts. *Sistemi Intelligenti.*
6. Villani, C., Lugli, L, Liuzza, M. T., Borghi, A. M. (under review). Varieties of abstract concepts and their multiple dimensions. *Psychological Research.*
7. Wauters, L. N., Tellings, A. E., Van Bon, W. H., & Van Haaften, A. W. (2003). Mode of acquisition of word meanings: The viability of a theoretical construct. *Applied Psycholinguistics*, *24*(3), 385–406.
8. Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology. General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446
9. Connell, L., Lynott, D., & Banks, B. (2018). Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0143
10. Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0144
11. Borghi, A. M., & Lugli, L. (in preparation). Abstract concepts and mouth motor system: evidence with an interference paradigm.

Levels of conceptual format: the challenge from supramodality

Fabrizio Calzavarini[1,2],
[1]*University of Turin, Italy,* [2]*Center for Logic, Language, and Cognition (LLC)*

What are concepts, and how are they represented in the human brain? For a long time, the debate on such issues has been dominated by two opposite conceptions. On the one hand, traditional theories in philosophy and cognitive science consider the conceptual system as functionally dissociated by the sensorimotor systems. According to such view, conceptual representations are amodal and arbitrary structures that bear no relationship to the physical and perceptual features of their referents (e.g., Fodor, 1981), and that are stored in the multimodal or associative brain regions (=*convergence zones*) in frontal, temporal, and parietal lobes (e.g., Binder, 2016). On the other hand, advocates of the so-called Simulation Framework have argued that concepts are sensorimotor representations. For instance, the concept CAT is a complex mental representation formed by visual information (how cats look like), auditory information (meow), tactile information (softness), and so one. According to such view, the conceptual system is distributed in brain regions that overlap or fall very close to the sensory primary and secondary regions (e.g., visual cortex, auditory cortex), and to motor cortex. In my talk, I will argue that this might be a false dichotomy. In recent years, many neuroscience data have undermined the classical distinction between sensorimotor and associative cortex, showing that multisensory convergence processes (bi- or three-modal) are already present at the level of traditionally unimodal areas (e.g., Ghazanfar e Schroeder 2006). I will argue that, in order to explain such data, one have to postulate the existence of various kinds of representations with a supramodal format, i.e. a format that is intermediate between amodal and perceptual (unisensory) representations. For instance, it is plausible that traditionally visual regions of the human ventral temporal lobes (LOtv, hMT), which surprisingly respond also to tactile stimuli, are functionally dedicated to integrating visual and tactile information in a more abstract, supramodal representation of object form and movement (e.g., Ricciardi et al., 2013). Similarly, it has been argued that certain regions of the motor cortex integrate visual, auditory, and proprioceptive information in a supramodal representation of the body (e.g., Holmes e Spence 2004). In my talk, I will argue that there are good reasons to believe that supramodal representations are exploited not only in perceptual functions, such as visual or tactile recognition, but also in cognitive functions that require conceptual knowledge, such as mental imagery and language. Nevertheless, the role of supramodality in conceptual competence is a largely unexplored topic (but see Fernandino et al. 2016). In the final section of my talk, I will consider some important implications these observations and these data have for the traditional and the Simulation accounts of conceptual competence.

References

Barsalou, L.W. (1999) Perceptual symbol systems, The Behavioral and Brain Sciences, 22(4), pp.577-609; discussion 610-660.

Binder (2016) In defense of abstract conceptual representations, *Psych. Bulletin & Review*, 23, 4: 1096–1108.

Fernandino L1, Binder JR1, Desai RH2, Pendl SL3, Humphries CJ1, Gross WL4, Conant LL1, Seidenberg MS (2016), Concept Representation Reflects Multimodal Abstraction: A Framework for Embodied Semantics, *Cereb Cortex,* 26(5):2018-34.

Fodor JA. (1981) Representations: philosophical essays on the foundations of cognitive science. MIT, Brighton: Harvester.

Ghazanfar AA. & Schroeder CE. (2006) Is neocortex essentially multisensory?, *Trends Cogn Sci,* 10(6):278-85.

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014) Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture?, *Neuroscience and Biobehavioral Reviews*, *41*, pp.64–77.

Cognitive functions and neural structures: population-bounded mappings

Marco Viola

Russian Institute for Advanced Studies, Moscow State Pedagogical University

Since its early days, neuroimaging studies have been interested in localizing the neural underpinnings of several cognitive functions. However, several objections have been raised against the project to seek for systematic one-to-one mappings (1-1M) between cognitive functions and neural structures.

While renegotiating the taxonomy of either neural structure or cognitive functions might improve systematicity, brain structures seem inherently pluripotent (i.e. involved in multiple cognitive functions), and cognitive functions sometimes exhibit degeneracy (i.e. they might be implemented in distinct neural substrates. Viola 2017).

Thus, Anderson (2014) advocates abandoning 1-1M in favor of probabilistic many-to-many mappings. Taking advantage of a data-driven approach, his strategy promises to yield revolutionary insights to cognitive ontology.

However, even though probabilistic mapping is the best you can get when only two variables are considered (namely, function and structure), adding further variables might improve predictability. This is the idea behind the so-called contextualist approaches to brain mapping (Klein 2012; Burnston 2016; Khalidi 2017). In my talk I advance a new and important contextualist approach, that I dub populational contextualism.

It is widely acknowledged that neural structures acquire (and modify) their functional role due to several factors, e.g. ontogeny or pathology. Scientists used to side-step this variability by assuming a "normal brain" and studying only "normal" (i.e. allegedly universally) cognitive capacities. Barring this assumption, one might fear, neuroscience is doomed to become idiographic. However, in recent years scientists are developing tools and interest for a systematic study of these individual differences (de Schotten and Shallice 2017).

In fact, normality assumption can be relaxed by individuating populations with common patterns of function-structure mappings. Three examples: occipital lobe in congenitally blind (Bedny 2017); resting-state activity as a marker for pathology (Greicius 2016); the reinterpretation of fusiform face area as an area for top-down recognition of objects on which the subject has expertise (Bilalić 2016).

While supporting the viability approach, these case studies highlight the heterogeneity of populations. The challenge I set for future neuroscience is then the following one: defining an ontology of neurocognitive homogeneous populations thus become, along with that of neural structures and cognitive function.

References

Anderson, M. L. (2014). After phrenology: Neural reuse and the interactive brain. MIT Press.

Bedny, M. (2017). Evidence from blindness for a cognitively pluripotent cortex. Trends in cognitive sciences, 21(9), 637-648.

Bilalić, M. (2016). Revisiting the role of the fusiform face area in expertise. Journal of cognitive neuroscience, 28(9), 1345-1357.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. Biology & Philosophy, 31(4), 527-550.

de Schotten, M. T., & Shallice, T. (2016). Identical, similar or different? Is a single brain model sufficient?. Cortex, 86, 172-175.

Greicius, M. (2016). Resting-State Functional MRI: A Novel Tool for Understanding Brain Networks in Neuropsychiatric Disorders. In Lehner, T., Miller, B. L., & State, M. W. (Eds.). (2016). Genomics, Circuits, and Pathways in Clinical Neuropsychiatry.

Khalidi, M. A. (2017). Crosscutting psycho-neural taxonomies: the case of episodic memory. Philosophical Explorations, 20(2), 191-208.

Klein, C. (2012). Cognitive ontology and region-versus network-oriented analyses. Philosophy of Science, 79(5), 952-960.

Viola, M. (2017). Carving Mind at Brain's Joints. The Debate on Cognitive Ontology. Phenomenology and Mind, (12), 162-172.

Keyword Index

abilità verbali e non-verbali 3
abstract concepts 74
academic recruitment 21
ad hoc concept construction 27
agent based model 21
agent-based models 31
Alzheimer's Disease 14
amodal accounts of concepts 5
analogia 36
anticipatory internal representations 10
apprendimento sociale 49
Aristotle 39
artifacts 30
artificial agents 51
Artificial Intelligence 8, 37, 39, 41, 63, 69
artificial life 71
auditory perception 47
autismo 49
autonomous systems 65

behavioral economy 30
bilingualism 67
body representation 19

card game 51
categorization 41
causation of movement 53
cognitive modelling 57, 67
cognitive neuroscience 76
cognitive overload 59
cognitive systems 41
competenze narrative 3
concept combination 43
concepts 5
concettualizzazioni errate 45
connectionist representations 6
consciousness 28, 69
consciousness biological Theories of 28
consciousness hard problem of 28
contextualism 76
cooperation 31
core knowledge confusion 23

deep learning 63

description logics	43
developmental memory	71
diagnosi	36
dictionary learning	65
DSL	3
dyslexia	61
early multisensory phenomena	5
educational tool	51
embedded cognition	71
embodied semantics	5
emotions	57
emozioni	17
emozioni estetiche	17
empiricism	63
environmental variation	71
epistemic bias	21
epistemology of AI machine learning algorithms	55
ethics	69
evolutionary robotics	31
expectation-based derivation	12
expectation-driven semantic	10
explanation	65
extended self	19
fairness	25
fake news	59
finzione	17
functionalism	39
general intelligence	8
GloVe	55
handwriting	14
hiring	21
history of science	28
human brain	69
induzione	36
inferential competence	6
information ecology	59
intelligent tutoring system	51
intelligenza emotiva	17
intention	57
interaction child-robot	25
interaction human-computer	57
interference paradigm	74

joint's rotation 53

kinematic tree 53
knowledge representation 43

lambda calculus 53

Machine Learning 14, 55, 65
mappe concettuali 45
meaning postulates 6
mental lexicon 67
merge 10
metafora 37
metodo 1
misconceptions 45
moral action recognition 39
moral perception 39
morphosyntax 61
motion planning 53
multiple representation theories 74
music 61

neural network 6
non monotonic reasoning 43
nosologia 36
numerical cognition 51

object ownership 19
online disinformation 59
ontologie 45
ontology 67

paranormal beliefs 23
pedagogia naturale 49
perception action 34
perception motor 34
perception pictorial 34
perception picture 34
perception vision 34
philosophy of cognitive science 76
philosophy of mind 76
predicative structure 12
presupposition 27
problem solving 8
prototype theory 43
psychology of reasoning 59

realtà virtuale 37

reasoning 8
recursion 10
referential competence 6
relevance theory 27
research assessment 21
research policy 21
resonant states 12
retelling 3
risk 31
robotica 49
rubber hand illusion 19

school of thought 21
semantic biases in language 55
semantic knowledge organization 23
semantic network 6
simulazione 37
skeleton 53
sounds 47
structure-function mapping 76
studio della mente 1
supramodality 5
sviluppo 1
symbolic representations 6
symmetry 12

technological determinism 30
technological instrumentalism 30
Technology and cultural biases 55
tecnologie didattiche 45
Theory of Mind 25
thought experiment 47
trust 31

Author Index

Airenti, Gabriella 1
Apicella, Andrea 63
Arosio, Fabrizio 57

Beraldi, Francesca 3
Bianchi, Valentina 9
Blanzieri, Enrico 5
Borghi, Anna Maria 23, 71
Borgo, Stefano 5

Calzavarini, Fabrizio 6, 7
Campanile, Fabrizio 71
Cerrato, Antonio 50
Chesani, Federico 8
Chesi, Cristiano 9, 11
Cilia, Nicole Dalia 13
Consoli, Gianluca 14

De Bortoli Vizioli, Aurora 23
De Stefano, Claudio 13
Debernardi, Carlo 24
Dellantonio, Sara 26
Di Dio, Cinzia 27
Domaneschi, Filippo 28

Ellia, Francesco 29

Fasoli, Marco 30
Felletti, Silvia 31
Ferretti, Gabriele 35
Fontanella, Francesco 13

Gagliardi, Francesco 38
Gagliardi, Gloria 3
Gambetti, Elisa 8
Giusberti, Fiorella 8
Gola, Elisabetta 39, 54

Hämäläinen, Hasse 41

Innocenti, Milvia 3
Isgrò, Francesco 63
Ishiguro, Hiroshi 27
Itakura, Shoji 27

Kanda, Takayuki 27

Lieto, Antonio 7, 42, 44, 46
Lorenzi, Giulia 48
Loreti, Daniela 8
Loria, Emiliano 49
Lugli, Luisa 71
Luizza, Marco Tullio 71

Manzi, Federico 27
Marchetti, Antonella 27
Massaro, Davide 27
Mazzarella, Diana 28
Mello, Paola 8
Miglino, Orazio 50
Milano, Nicola 69
Moro, Andrea 11
Mulatti, Claudio 26

Nishina, Hiroyuki 51
Nolfi, Stefano 69
Numerico, Teresa 52

Ojha, Amitash 54

Paglieri, Fabio 55
Pagliuca, Paolo 31
Pastore, Luigi 26
Persici, Valentina 57
Pishyar-Dehkordi, Peiman 59
Plebe, Alessio 62
Ponticorvo, Michela 50
Pozzato, Gian Luca 44
Prevete, Roberto 63
Priori, Eleonora 24

Qian, Cheng 65

Rebuffo, Francesco Vittorio 46

S. Gouveia, Steven 66
Sapienza, Alessandro 31
Schembri, Massimiliano 50
Scotto Di Freca, Alessandra 13
Stucchi, Natale 57

Tagliabue, Jacopo 68

Tamburrini, Guglielmo 63
Tiso, Stefano 69
Togni, Andrea 70
Tummolini, Luca 23

Villani, Caterina 71
Vilone, Daniele 31
Viola, Marco 24, 72

Zhang, Xinlei 74