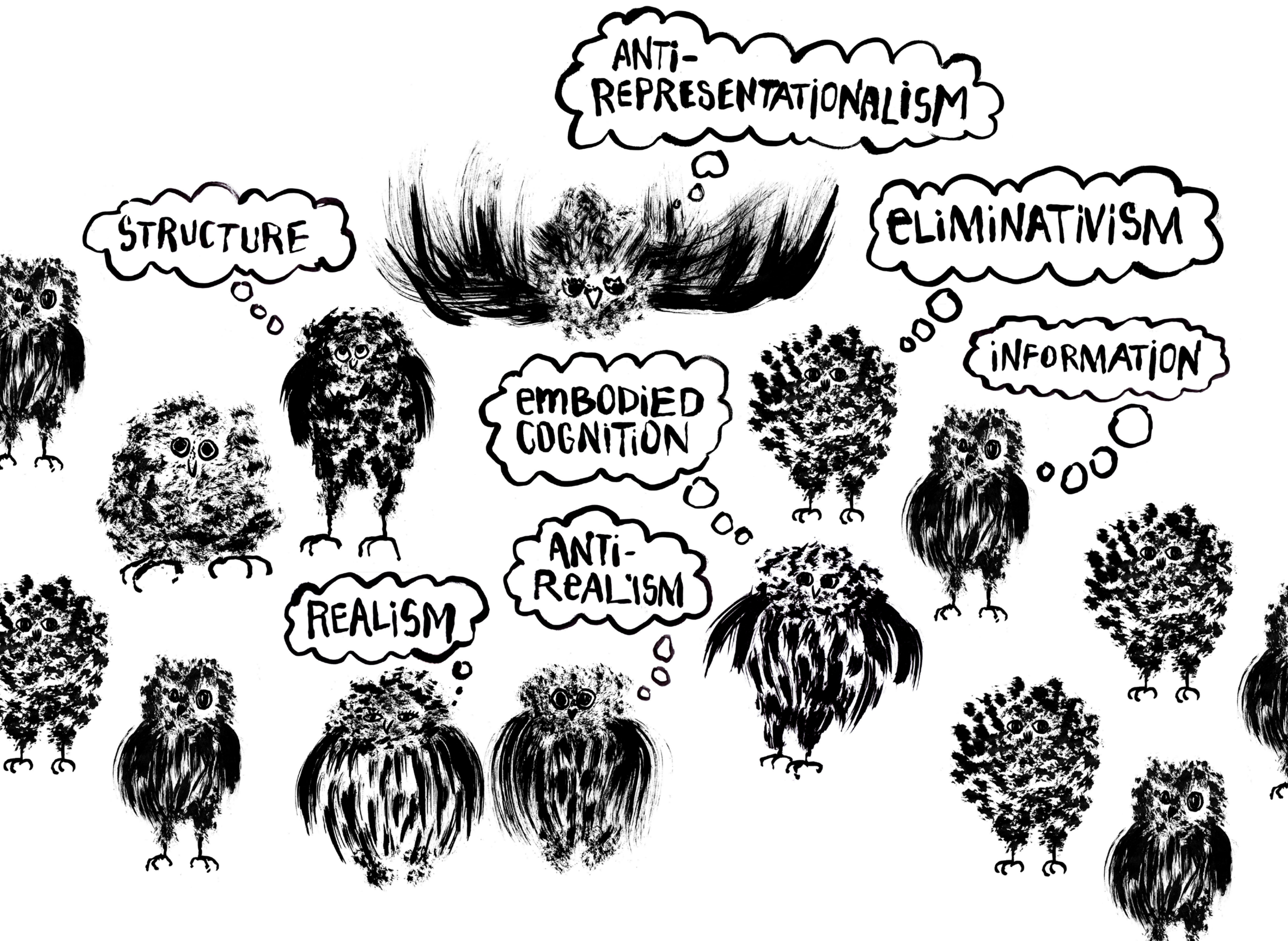
# Representations in Cognitive Science

*A "Varieties of Information" Workshop*




**September 8-9, 2022** University of Barcelona

# SCHEDULE

## THURSDAY 8

11:30 - 12:50
### A SYSTEMS NOTION OF VEHICLE
Manolo Martínez
Universitat de Barcelona

13:00 - 14:45 Lunch

15:00 - 16:20
### THE VECTOR GROUNDING PROBLEM
Dimitri Coelho Mollo
Umeå University

16:30 - 17:50
### A BETTER WAY TO NATURALIZE INTENTIONALITY
Frances Egan
Rutgers University

18:00 - 19:20
### VARIETIES OF REALISM AND ANTI-REALISM ABOUT MENTAL REPRESENTATION
Zoe Drayson
University of California, Davis

## FRIDAY 9

10:00 - 11:20
### CORRESPONDENCE THEORY OF SEMANTIC INFORMATION AND REPRESENTATIONAL MECHANISMS
Marcin Miłkowski
Polish Academy of Science

11:30 - 12:50
### STRUCTURAL SIMILARITY AS THE PROPER BASE FOR BAYESIAN MODELS
Nina Poth
Ruhr-Universität Bochum

13:00 - 14:45 Lunch

15:00 - 16:20
### THE ROLE OF COMMUNICATION THEORY IN THEORIES OF REPRESENTATION
Stephen Mann
Max Planck Institute for Evolutionary Anthropology

16:30 - 17:50
### STRUCTURAL REPRESENTATIONS: WHAT ARE THEY? WHAT CAN THEY DO FOR US?
Marc Artiga
Universitat de València

# **ABSTRACTS** THURSDAY 8

11:30 - 12:50

Manolo Martínez

## A SYSTEMS NOTION OF VEHICLE

*Vehicles* are a central theoretical posit in cognitive science (Bechtel 2009). It is almost universally assumed that representations can be analyzed in terms of their *content* (the entities those representations stand for) on the one hand; and the physical structures that bear that content, on the other—the representational vehicles.

Cognitive scientists have paid much attention to which kinds of entities can act as vehicles for which specific kinds of representations (say, sentences in a language of thought, (Fodor 1980, 2008); or "clusters in the state space of a hidden layer [in an artificial neural network]", (Shea 2007)). The problem of providing a general characterization of vehiclehood—that is to say, of what it is that makes some physical entity a vehicle—has received much less attention. This is the task I will approach in this piece: I sketch a *systems* notion of vehicle—a high-level description of the kinds of processes that need to be in place for vehicles to emerge, and to be maintained. The literature on vehicles often gives the impression that any physical entity could potentially act as a vehicle. The systems perspective shows that making and maintaining vehicles is a relatively complex engineering task.

In particular, I will argue that vehiclehood is tied to *channel coding* in the sense that information theory gives to this notion (Cover and Thomas 2006; MacKay 2003). One popular idea about vehicles is that they are "picked out in terms of intrinsic *processing-relevant* non-semantic properties" (Shea 2018, 39, my emphasis). How does the brain (a very noisy environment, Faisal, Selen, and Wolpert 2008) ensure that different tokens of the same vehicle type have the same processing-relevant properties? The answer suggested by the theory of channel coding is that this is done by *identifying vehicles with regions of activation space, such that the probability of overlap between these regions is suitably low*. This identification provides several theoretical advantages:

1. It offers a graded notion of vehicle (indexed by the probability of overlap);

2. It suggests concrete strategies for the generation of vehicles, such as introducing redundancy into signals, or leaving regions of activation space unexploited;

3. It offers the possibility of quantifying the cost of maintaining a system of vehicles, in terms of the trade-off between available rate of communication and probability of error.

I end by interpreting two widespread operations in the brain as cases of vehicle creation, along the lines just sketched: sparse coding (Perez-Orive et al. 2002) and neural oscillations (Buzsáki 2006). I will argue that much of the discussion on sparse coding (e.g. Spanne and Jörntell 2015) and oscillatory behavior is vitiated by a conflation of vehicle maintenance (error management) with other information-theoretic operations, such as compression, which are not non-semantic.

**References**

Bechtel, William. 2009. "Constructing a Philosophy of Science of Cognitive Science." *Topics in Cognitive Science* 1 (3): 548–69. https://doi.org/10.111 1/j.1756-8765.2009.01039.x.

Buzsáki, György. 2006. *Rhythms of the Brain*. Oxford University Press.

Cover, T. M., and J. A. Thomas. 2006. *Elements of Information Theory*. New York: Wiley.

Faisal, A. Aldo, Luc P. J. Selen, and Daniel M. Wolpert. 2008. "Noise in the Nervous System." *Nature Reviews Neuroscience* 9 (4): 292–303. https: //doi.org/10.1038/nrn2258.

Fodor, Jerry A. 1980. *The Language of Thought*. 1 edition. Cambridge, Mass: Harvard University Press.

—. 2008. LOT 2: *The Language of Thought Revisited*. Oxford Clarendon Press.

MacKay, David JC. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Perez-Orive, Javier, Ofer Mazor, Glenn C. Turner, Stijn Cassenaer, Rachel I. Wilson, and Gilles Laurent. 2002. "Oscillations and Sparsening of Odor Representations in the Mushroom Body." *Science* 297 (5580): 359–65. https: //doi.org/10.1126/science.1070502.

Shea, Nicholas. 2007. "Content and Its Vehicles in Connectionist Systems." *Mind & Language* 22 (3): 246–69.

—. 2018. *Representation in Cognitive Science*. Oxford: Oxford University Press.

Spanne, Anton, and Henrik Jörntell. 2015. "Questioning the Role of Sparse Coding in the Brain." *Trends in Neurosciences* 38 (7): 417–27. https: //doi.org/10.1016/j. tins.2015.05.005.

---

15:00 - 16:20

Dimitri Coelho Mollo and Raphaël Millière

## The Vector Grounding Problem

The impressive performance of current artificial natural language processing systems (NLP) — such as GPT-3 — on complex linguistic tasks has generated considerable debate about how to understand their abilities. Are those outputs intrinsically meaningful, or are we simply projecting meaning onto AI systems that are nothing more than statistical parrots?

We argue that in order to shed light on this question we must first examine if and to what extent the internal representations and outputs produced by NLP systems are grounded in the world. Although the debate on AI grounding is relatively old, it has mostly focused on what Harnad called the Symbol Grounding Problem for classical AI systems: how do symbols manipulated by these systems acquire their meanings? The most successful NLP systems today, however, are large, deep neural networks employing the Transformer architecture, which compute over vectors and matrices, instead of symbols. Nonetheless, a similar problem arises for such systems: what we call the Vector Grounding Problem.

Our aim in this paper is two-fold. First, we distinguish different ways in which internal representations can be grounded in biological or artificial systems. We dub them referential, sensorimotor, communicative, and epistemic grounding, respectively. Unfortunately, these different senses of grounding are often confused or conflated, leading to misunderstandings. By clarifying the differences between them, we show that referential grounding is the most basic one, as well as the most relevant to the question of whether the representations and outputs of NLP systems can be intrinsically meaningful.

Second, taking our cue from theories of representational content in cognitive science, we argue that current NLP systems have the tools required to overcome the Vector Grounding Problem, lending force to the idea that they are more than stochastic parrots. We show that current artificial language systems can satisfy the minimal requirements for referential grounding, insofar as:

· they are trained on datasets whose implicit structure essentially depends on causal interactions between humans and the world;

· the tasks they are trained to solve require them to represent and exploit such structure to produce their outputs;

· when applied to concrete tasks, such as language translation or image generation, their outputs, and internal representations, have standards of correctness and satisfaction that depend on how the world is or could be.

As a corollary, the more human-dependent structure these systems are trained on, the stronger the grounds for seeing their representations as grounded. This suggests that multimodal AI systems with NLP components, such as image-generation systems like DALL-E, have a yet stronger claim to being referentially grounded.

Importantly, we are not claiming that current NLP systems have the tools to achieve language understanding, or to form beliefs and communicative intentions. Our point is more basic, and more modest: the representations and outputs of current NLP and multimodal systems can be grounded in the world by indirect means, i.e., through the mediation of our own interactions with the world. We conclude by discussing some implications of our view for neurobiologically plausible models of psychological semantics.

---

16.30 - 17.50

Frances Egan

## A BETTER WAY TO NATURALIZE INTENTIONALITY

Computational cognitive science (CCS) aims to provide a naturalistic foundation for theorizing about the mind. It has been widely assumed that internal states posited by CCS are representational and that their content can be naturalized by specifying non-intentional and non-semantic sufficient conditions for a state to have a particular content. So far, this *naturalization project* has not met with success. In a series of papers I have defended a *deflationary* construal of representation, arguing that representational content plays a merely heuristic role in CCS. In this talk I suggest that we reconceive the project of naturalizing intentionality. I sketch an alternative to the traditional naturalization project, one that computational theories are likely to satisfy.

---

18:00 - 19:20

Zoe Drayson

## VARIETIES OF REALISM AND ANTI-REALISM ABOUT MENTAL REPRESENTATION

In this paper I survey the plethora of positions in the debate between realists and anti-realists about representations in cognitive science, mapping out the logical space with a view to future discussion. I introduce the most robust form of realism, according to which our best theories in cognitive science commit us to discrete concrete vehicles of representation which objectively possess determinate content. I show how other positions in the debate (including those of Egan, Coelho Mollo, and Rescorla) can be characterized by which of the requirements (e.g. concreteness, objectivity, determinacy) they drop. I appeal to the more general debate in philosophy of science to further explore the viability of constructive empiricism and structural realism, and to reassess the arguments for eliminativism.

# **ABSTRACTS** FRIDAY 9

10:00 - 11:20

Marcin Miłkowski

## CORRESPONDENCE THEORY OF SEMANTIC INFORMATION AND REPRESENTATIONAL MECHANISMS

In my talk, I am going to show the place of correspondence theory of semantic information in my account of representational mechanisms. In fact, this theory was the implicit assumption of my previous joint work on structural representation with Paweł Gładziejewski. In this presentation, it will become explicit for the first time.

Overall, the account insists that there is more to representation than semantic information, and that pragmatics is essential to make it explanatorily relevant.

11:30 - 12:50

Nina Poth

## STRUCTURAL SIMILARITY AS THE PROPER BASE FOR BAYESIAN MODELS OF CONCEPTUAL THOUGHT

Generalisation is an ability most basic to conceptual thought (Evans 1982, Camp 2009). How should cognitive scientists explain it? Probabilistic models of cognition approach this problem typically as a task of Bayesian inference of concepts, where a rational agent infers whether an unknown instance falls under the concept that subsumes a known instance (Tenenbaum & Griffiths 2001). Despite their predictive success, it remains unclear how to interpret the representational content associated with probabilistic inferences in these models (Cao 2020; Sprevak 2020). While proponents often claim that probabilistic inferences are structured into informationally rich topical domains with causal representations at their core (Ullman & Tenenbaum 2020), they largely remain silent about what principles relate these structures to their underlying perceptual dimensions and resort to implausibly strong accounts of representation, such as a 'probabilistic language of thought' (Goodman, Tenenbaum & Gerstenberg 2015). I suggest a novel way to integrate Bayesian rational analyses of conceptual thought with less demanding information-theoretic analyses of mental representation (Isaac 2020, Martinez 2019, Artiga and Sebastián 2020). Specifically, I argue that probabilistic inferences of concepts originate in a biased relationship between contingent statistical features of the environment and mutual-informational relevance relations among perceptual representations. On this view, Bayesian inferences of concepts should be grounded in structural similarity representations, which partly provide the information-theoretic content of concepts and simultaneously justify perceptual belief. I contrast this view to recent discussions of conceptual thought in the predictive processing literature (Figdor 2021; Williams 2020).

**References**

Artiga, M., & Sebastián, M. Á. (2020). Informational theories of content and mental representation. *Review of Philosophy and Psychology,* 11(3), 613–627.

Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, 78(2), 275–311.

Cao, R. (2020). New labels for old ideas: Predictive processing and the interpretation of neural signals. *Review of Philosophy and Psychology*, 11(3), 517–546.

Evans, G. (1982). *The varieties of reference*. Oxford University Press.

Figdor, C. (2021). Shannon+ friston= content: Intentionality in predictive signaling systems. *Synthese*, 199(1), 2793–2816.

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought.

In *The conceptual mind: New directions in the study of concepts*. MIT Press.

Isaac, A. M. (2020). The semantics latent in shannon information. *The British Journal for the Philosophy of Science*, 70(1), 103-125.

Martínez, M. (2019). Representations are rate-distortion sweet spots. *Philosophy of Science*, 86(5), 1214–1226.

Sprevak, M. (2020). Two kinds of information processing in cognition. *Review of Philosophy and Psychology*, 11(3), 591–611.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, 24(4), 629–640.

Williams, D. (2020). Predictive coding and thought. *Synthese*, 197 (4), 1749–1775.

---

15:00 - 16:20

Stephen Mann

## THE ROLE OF COMMUNICATION THEORY IN THEORIES OF REPRESENTATION

Influential views about the relevance of information theory for naturalistic approaches to intentionality all agree on one premise: that the only relevance the theory could have is via measures of correlation it defines. Those views then diverge on the question whether measures of correlation really are relevant for naturalistic intentionality. Here I challenge that premise. In doing so I correct two fundamental misconceptions about information theory (or, more properly, communication theory). First, it is not true that the meanings of signals are irrelevant for the theory, nor did Shannon say they were (his oft-quoted warning was about something else, and has been widely misinterpreted). Second, although measures of correlation cannot distinguish between signals and natural signs, it is not true that the theory itself cannot distinguish them. Taken together, these two points challenge the contemporary practice of keeping communication theory at arm's length from naturalist theories of representation — especially teleosemantics, which is uniquely well suited to incorporating the formal theory.

16:30 - 17:50

Marc Artiga

# STRUCTURAL REPRESENTATIONS: WHAT ARE THEY? WHAT CAN THEY DO FOR US?

Abstract: The notion of `structural representation' has recently been suggested to play two key dialectical roles: on the one hand, against eliminativist challenges, it is supposed to capture some cognitive processes, in which a represenationalist explanation can earn its keep. On the other, against liberal views, it vindicates a concept of representation that avoids the problem of trivialization. A first goal of this paper is to show that the notion of 'structural representation' can be understood in different ways. Secondly, I will argue that probably no interpretation can underpin the theoretical role that this concept is meant to play in the recent literature.


## VENUE

Residència d'Investigadors
C/Hospital, 64 - 08001 Barcelona

## ONLINE

Please write to mail@
manolomartinez.net for the link

Research Group in Analytic Philosophy

BIAP
Barcelona Institute
of Analytic Philosop

VLC Philosophy LAB

GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACIÓN