

Benoit Hardy-Vallee and Pierre Poirier

Embodied Thoughts Concepts and compositionality without language

1. Introduction

Is thinking necessarily linguistic? Do we *think with words*, to use Bermudez's (2003) phrase? Or does thinking occur in some other, yet to be determined, representational format? Or again do we think in various formats, switching from one to the other as tasks demand? In virtue perhaps of the ambiguous nature of first-person introspective data on the matter, philosophers have traditionally disagreed on this question, some thinking that thought had to be pictorial, other insisting that it could not be but linguistic. When any problem divides a community of otherwise intelligent rational thinkers, one suspects some deep conceptual confusion is at play. Indeed, we believe that the conceptual categories used to frame these and related questions are so hopelessly muddled that one could honestly answer "both simultaneously", or "neither", depending on what is meant by the alternatives. But let's get our priorities straight. This paper first and foremost aims at defending what we believe to be a step in that direction of the proper view of thinking, a view we call the spatial-motor view. In order to do so, however, we have found it essential to start by addressing the conceptual confusion just alluded to. Accordingly, the paper proceeds in two steps. First a conceptual step, in which we reconsider some of the traditional categories brought into play when thinking about thinking. Then an empirical step, in which we offer empirical evidence for one of the views conceptually isolated during the first part of the work. Future version of this collaborative work will include a speculative step in which we spin out an evolutionary and developmental scenario whose function is to justify the spatial-motor view by showing how it fits into current evolutionary and developmental theories.

2. Conceptual step: Models of thoughts

2.1 Thinking in cognitive science

The first question to address to start clearing up the conceptual mess is how thinking is related to cognition generally. Conceptually, there seems to be three possible positions (and there are philosophers or cognitive scientists occupying each). First one can take the two terms to refer to the same phenomenon or capacity, either because the two are confused or identified. Simply confusing thinking with cognition is a bad thing, we take it, but identifying the two can be a viable position, provided the identification thesis is justified. There are two ways to identify thinking with cognition. The first position is rationalism wherein one develops a model of rational (and usually conscious) thinking and applies it to the whole of cognition. Jerry Fodor, for instance, has been defending (justification in hand) a rationalist position for 30 years which goes something like this: the Turing Machine, which is a model of thinking, is the only model we currently have of a rational process. The second position identifying thinking with cognition we call nihilism: develop models of cognition that make no space for rational deliberative thinking as a distinct process. *Nihilists* are usually neuroscientists, roboticists and psychologists working on lower level cognition or within the so-called micro-cognition perspective. One can be a nihilist either because one eliminates thinking altogether (the term is a throwback to an obsolete conception of the mind) or because one thinks that thinking is essentially no different from other low-level cognitive process, such as perception and categorization, and thus needs no special explanation of thinking over and above that of these processes.

We believe that these two avenues are reductive dead ends: the first because it reduces the whole of cognition to one of its aspects or function and the second because it neglects an important aspect of cognition. We believe that thinking is a genuine phenomenon, in need of explanation, and that that explanation cannot be extended to cover the whole of cognition. This is the third possible position on the conceptual landscape: thinking is a special kind of cognitive process, a process that shares important aspects of cognition, because it is a part of cognition, but at the same time a process that is markedly different from the rest of cognition in certain crucial ways. An explanation of thinking must therefore show how thinking emerges (computationally, developmentally, evolutionarily as a special kind of cognition.

We shall therefore aim to develop a model of cognition that stays clear from the Scylla of rationalism and the Charybdis of nihilism, but this is a job of the

next section and the next (empirical) step in our defence. The remainder of the present section attempts to conceptually isolate the kind of model of thinking we believe is apt for the job.

2.2. We Are Family

All the models of thought that we discuss in this paper agree on a number of key issues. Actually, they agree on much more than they disagree on and, in that sense, they form a *family of models*. For reasons that will become clear, we'll call this family "compositionalist models of thinking", and we'll label "compositionalist" anyone who puts forward or defends such a model. Our point in the paper will be claim that compositionalists are wrong, not in being compositionalists, since we'll end up defending a compositionalist view ourselves, but that they are wrong in the kind of compositionalist model that adequately accounts for thinking. Hence, the paper is a family dispute of sorts, aired in public! The present section aims at

(1) at setting out the common conceptual landscape that underlines how compositionalists think about thinking (subsection 2.1) and then (2) explaining how various compositionalist models differ (subsection 2.2).

We said that compositionalist models of thinking agree on much. First, but this by no means sets compositionalist models apart from other models, recent or historical, they all agree that thinking is a *representational* process, that is, a process that takes representations as inputs, outputs and intermediaries. Thoughts, accordingly, are the representations over which thinking occurs.

(TI) Thinking is a representational process; thoughts are the representations over which the process is defined (*Representationalism*)

We shall not discuss representationalism further here. Apart from Ryleans, Heideggerians, extreme system dynamicists and some fans of new-AI or embodied cognition, no one will contest representationalism. We do wish to make one point, however, since we *do* see ourselves as fans of new-AI and embodied cognition. This paper is about thinking, not about cognition generally. Thinking, as we emphasized, is one cognitive capacity; not the whole of cognition. A common error made by a generation of cognitive scientists was to develop models of thinking, and then mistake those for models of cognition generally. Some cognitive capacities are, we believe, non-representational (Brooks 1991), while others are minimally representational (Clark 1997) and still others are fully representational. Thinking, we believe, is among those. Be that as it may, we'll strive to show below that thoughts, even though they *are* representations, are much more embodied than most representationalists believed.

Hence, the present debate about thinking is among friends of representationalism. However, the process that transfers digital pictures from my

camera to my computer is also, in a way, a process defined over representations and yet no one would claim that that process is actually a kind of thinking. Thinking is a special kind of process defined over representations: a process that takes representations as inputs and transforms them into output representations only (1) if the input and output representations bear some appropriate epistemic relation (such as truth-preservation) and (2) if the process by which the input representations are transformed into the output representation can itself be given an epistemic interpretation (such as inference). Given the epistemic relation between input and output representations and the epistemic interpretation of the process itself, the input representations can be seen as offering a rationale (Cummins 1983) for the output representations and the input representations transfer the justificational status to the output representations. Note that many cognitive processes defined over representations do not thereby count as thinking. The process that maps retinal representations of the visual scene onto perceived 3-D representations does not thereby count as thinking, and it shouldn't: no one would want to count vision as a form of thinking¹.

(T2) Thinking is an inferential representational process (Inferentialism).

Up until recently, T2 (inferentialism) would have sufficed to characterize the family of models we are about to discuss. With the advent of connectionism, however, a number of new conceptual positions have opened up. In neural networks (ANNs), number vectors can be interpreted as vehicles for representations and the vector transformations effected by ANNs have been qualified by some as a kind of *statistical inference*. Statistical inference is a process that transforms input representations into output representations in virtue of the statistical properties of input vectors. For instance, some ANN learning algorithms extract the principal components present in input vectors such that a trained networks map input representations onto its principal components. To properly count as a form of *inference*, the input and output representations must bear some relevant epistemic relation. If the principal components are thought of as a kind of prototypical representation of a category and the input vectors are thought of as instances of that category, then the vector transformation can be seen as a form of categorization, which could be described thus: given what I know about my perceptual world (as encoded in my weight matrix), there is a 0,85 probability that the object causing my current input is an apple. Hence, the representational process effected by ANNs may be understood as a kind of inferential process. But is it thinking?

The categorization example rehearsed above provides a nice case. Note the difference between two kinds of classification procedures. 1) Someone looks at

¹ Anyone who did would have to invent a new word to label the process someone engages in when she says, for instance, "I'm thinking about it!", say in response to a marriage proposal.

an apple (retinal representations of the apple are thereby generated) and recognizes it as an apple. 2) Someone looks at an apple (retinal representations of the apple are thereby generated) and does not recognize it as an apple right away (maybe it's a genetically modified apple that's blue). Intrigued, that person looks around to see that it's in the apple section of the supermarket and that, looking closer, that it does have the skin-texture and consistence of other apples she has previously encountered. On the basis of that evidence, without *thinking about it further*, she takes the blue thing in the apple section to be an apple. We claim that these two cases of classification were done in radically different ways. The first case was presumably the kind of (unconscious) Bayesian inference that might have been realized by a neural network vector transformation: given how that thing looks and what I know about my visual world, I'm pretty confident in my belief that it is an apple (e.g., if I were hungry, I would eat it). Again, we believe that no one would want to qualify this process as thinking, although, like thinking, it is a representational inferential process. However, most would qualify the second process as thinking (as we did in describing it): not recognizing the thing, she looked around, thought about it and inferred that it was an apple. Maybe the (unconscious) inference went like this: given that this thing is in the apple section and that I know that the supermarket groups produce according to type, and given that the thing does have the skin texture and consistence of apples, I conclude that it is an apple. As a process, what sets this inference apart from the first one? Following Fodor and Smolensky, we say that the first kind of inference is defined over non-representational constituents of representations, whereas, in the second case, the inference is defined over the representational constituents of representations. The first kind of process is sensitive to the statistical structure of its input representation and the second kind is sensitive to the representational structure of its input representations (the representations that *compose* them). We'll call the first "statistical structure-sensitive inference" and the second, following Fodor, "constituent structure-sensitive inference".

We can thus state what holds compositionlists together as a family (T3) Thinking is an inferential process sensitive to the constituent structure of representations; thoughts possess the appropriate compositional structure to sustain structure sensitive inference (*Compositionlistmf*)

Note that T3 includes T2, which itself includes T1. Hence, we'll lake T3 to succinctly state the position we're after.

2.3. Family Feud

Compositionlists agree that thinking is an inferential process sensitive to the constituent structure of representations. What tears the family apart is the

source of the relevant representations' constituent structure. To understand the conceptual space in which the family dispute is framed, it is important to consider additional properties representations may have as thoughts. According to compositionalist models, these properties come organized in two neat orthogonal dichotomies:

- linguistic vs. non-linguistic, and
- digital vs. analog.

Linguistic vs. non-linguistic representations. Jerry Fodor published an important book in 1975 in which he claimed that thought is linguistic, a book aptly titled *The Language of Thought*.² In that book, he defended the idea that the representations over which thinking was defined had to be language-like in nature; that is, they had to be language-like for the cognitive science of his day to have any chance of being true. By "language-like", Fodor meant that mental representations had the kind of structure that public language has according to Chomskyan linguistic. Like public language, mental representations would be the result of a chomskyan generative process, a process which automatically gives mental representations the appropriate constituent structure. Indeed, according to Chomsky, public sentences are constructed from primitives by means of processes like concatenation and permutation that preserve the original integrity of the primitives. If the representations over which thinking is defined are similarly constructed, then they cannot but have the kind of constituent structure that compositionalist think they have.

Non-linguistic representations are those representations that cannot be represented as the result of a Chomskyan generative process.

Digital vs. analogical representations. There is an old opposition, in philosophy and cognitive science, between analogical and digital representations. Analogical representations are said to be modal, continuous, particular, iconic and holistic while digital representations are thought to be amodal, discrete, general, symbolic and structured. Following Dretske's lead, we say that "a representation that *s* is *F* is digital insofar as it carries nothing else than *s*'s being *F*". (Dretske, 1981). By contrast, a representation that *s* is *F* is analog if it carries other information besides *s*'s being *F*. Dretske's paradigmatic examples draw on the distinction between a statement and a picture. « The cup has coffee in it » doesn't tell us how much coffee there is, what kind of cup it is, but a picture of the cup with the coffee would represent all those details. The idea here is that mental representations (and

² In retrospect however, it seems that the title was not so apt, since the book was about linguistic nature of cognition generally. If one thinks, as Fodor probably does, that thinking is a good model of thinking, confusing thought with cognition is somewhat inconsequential (although it may lead to confusion in others).

not only public representations like pictures and statements) also come in two kinds, mental statements (say, in mentalese) and mental images.

Whereas linguistic representations, conceived on the Chomskyan model, are automatically compositional, advocates of digital representations still need to explain what makes digital representations compositional. This is not a project we address here. The important point for us is that, in the context of the present debate, it is usually understood that only digital representations can be compositional (an important exception is Cummins 1996). Analog representations are simply not the kind of things that can have compositional structure. This is the idea we attack in the next section.

To end the conceptual step, let's characterize the three compositionalist models of thinking in terms of what distinguishes them. We saw that all compositionalist models agree that thinking is an inferential process sensitive to the constituent structure of representations (T3) and that they disagree about the source of the necessary constituent structure.

- (*Linguistic-T3*). Thinking is an inferential process sensitive to the constituent structure of *linguistic* representations (all linguistic representations possess constituent structure).
- (*Digital-T3*). Thinking is an inferential process sensitive to the constituent structure of *digital* representations (digital representations can possess constituent structure - through some process yet to be explained).
- (*Analog-T3*). Thinking is an inferential process sensitive to the constituent structure of *analog* representations (analog representations can possess constituent structure).

It is the purpose of the next section to explain how analogical representations can possess constituent compositional structure.

3. Empirical Step: thinking with analogical representations

We saw that partisans of *linguistic-T3* automatically get an account of compositional constituent structure. This affords them the option of posing a challenge to their opponents within the compositionalist family: Can there be compositionality, hence thinking (inference sensitive to constituent structure), without language? Their answer is: No. No language, no compositional thoughts. Partisans of *digital-T3* take a more liberal view that is ready to answer positively: babies and animals do think when, *but only when*, they manipulate digital representations. We too wish to adopt a more liberal view but disagree with partisans of *digital-T3* that thinking has to be digital. We argue here that empirical knowledge in cognitive science shows that some kinds of analog representations can be compositional. We begin by outlining an account of analog representation

(3.1), and then argue that analogical representational systems can exhibit constituent structure (3.2), which can support a kind of inference sensitive to constituent structure (3.3). Together, these findings support our Spatial-motor conception of thoughts.

3.1. Analogical representations

After an initial anti-representationalist stance, ncw-AI roboticists began to design *representational* control architectures for autonomous robots. Cognitive robotics (Clark & Grush, 1996) thus replaced (or complemented) reactive robotics. However, cognitive roboticists did not see representational activity as a symbolic re-description of the world, but as a sensorimotor simulation of possible behaviours. A *cognitive robot* “doesn’t have to jump off a cliff before discovering that this is dangerous; it can recognize the affordance and let its hypothesis about moving toward the cliff action die in its place (...)” (MacDorman, 1999, p. 21). Lynn Andrea Stein’s *MetaToto* (Stein, 1994) is an example of such “neo- representationalist” control architecture: her robot uses its knowledge of the external world to build a map of its environment, and then consults its map either on-line, to guide its behaviour, or offline, to try out behaviours. As Grush (2003, p. 85) puts it: “this model creates a ‘virtual reality’ with which [its] basic reactive navigational routines can interface in order to imaginatively explore the environment”. We agree but we believe that the expression *virtual reality* should not be put inside quotation marks: in non-linguistic beings, cognitive representation *is* virtual reality (VR). Virtuality is not about “not being *really* real”, but about *possibility*: organisms that represent the world explore the possibilities of action, as players in VR environments do.

On this view, (non-linguistic) representation consists in the reactivation of sensorimotor experiences, either in the presence or absence of the objects or causes of the original experiences, in order to simulate the temporal evolution of the experience and thus predict its outcome or consequence.³ A representational system is a system that can plan ahead (McFarland & Bossert, 1993; Cruse, 2003), one that is able to *pre-select* among possible behaviours (Dennett, 1994). There can also be representations, not of things in the world, but of the body, that is,

³ We are prepared to agree that human representation is too complex to be restricted to that definition, but our point here is to describe non-linguistic representations. In what follows, we stick with a minimal theory of representation, which takes the epistemological dimension as its primitive (instead of a view of representation which takes its metaphysical dimension, that is, reference, as its fundamental aspect.).

body models (or schemata). Maravita and Iriki (2004) define a body model as a constantly updated status of the body shape and posture. According to Cruse (1999, 2003) and Damasio (1999), body models are the basis of world models. Neuroscientists have begun to unravel the neural implementation of body models (Berlucchi & Aglioti, 1997; Maravita A. et al., 2003), their role in body posture (Morasso, Baratto, Capra & Spada, 1999) and in the phantom limbs phenomenon (Melzack, 1999). Grush (2003) analyses body models, or motor emulators as he calls them, in a control-theoretical framework: motor emulators are neuronal forward models of the musculoskeletal system (MSS) that implement the same input-output operation of the MSS. Emulators behave as internal predictive models.

We will frame the distinction between the body model and the world model in terms of emulation and simulation⁴. On this view, the brain *emulates* the body (motor emulators reproduce in parallel the body's behaviour and generate feedback, like real perception and action) and *simulates* the external world (it reproduces possible things, agents, and events). There may be no strict distinction between emulation and simulation, since they are, after all, two kinds of sensorimotor representations. All that applies to emulation applies also to simulation, except that what is simulated is outside the body.

Barsalou's theory of Perceptual Symbols Systems (1999) offers a detailed account of simulation. Its foundational thesis is that perception and cognition recruit the same brain resources, a fact already acknowledged by cognitive neuroscience (Grezes & Decety, 2001). Semantic memory representations are not language-like, amodal representations such as frames, schemata, predicate calculus expression, or a feature list. A perceptual symbol is a "record of the neural activation that arises during perception" (Barsalou, 1999, p. 583). After every perceptual experience, the brain records sensorimotor aspects of the experience, and integrates it in multimodal frames. Only some aspects are thus recorded: selective attention focuses on some features (e.g., movements, vertices, edges, colors, spatial relations, heat, etc.) that are encoded in long-term memory encodes and recreated when needed. Hence, only schematic representations are available for processing. On Barsalou's view, each frame is not only an aggregate, but also a *simulator* that, after perception or in linguistic processing, may generate a simulation of categories of objects or events stored in the frame by reactivating some perceptual symbols. Perceptual symbols intervene in categorization and

⁴ These terms are getting to be quite popular, and they are sometimes used to mean various things. Before it gets to the point where we cannot understand each other, we propose to use their computer science acceptance (grabbed from FOLDOC, The Free On-line Dictionary of Computing <URL:<http://www.foldoc.org/>>): *Emulation*: 'When one system performs in exactly the same way as another [...] *Simulation*: Attempting to predict aspects of the behaviour of some system by creating an approximate [...] model of it.

prediction: for instance, once the object seen is categorized as an X, the simulation of schematic properties of X allows the system to predict X's behaviour, or its own behaviour toward X. These simulations are the human equivalent of MetaToto's control architecture: a sensorimotor VR that helps planning ahead. Many simulations can be generated simultaneously, and they can be combined into complex simulations. We won't defend this account as an exhaustive theory of human cognition; we are sympathetic to Barsalou's model to the extent it provides a model of analogical representation (at least for animals and babies):

many animals have perceptual symbol systems that allow them to simulate entities and events in their environment. [...] Because many animals have attention, working memory, and long-term memory, they could readily extract elements of perception analytically, integrate them in long-term memory to form simulators, and construct specific simulations in working memory. (Barsalou, 1999, p. 606)

Similar to nonhumans, infants may develop simulators and map them into their immediate world. During early development, infants focus attention selectively on aspects of experience, integrate them in memory, and construct simulators to represent entities and events [...] (Ibid.)

We have described what, on a minimal theory of representation, we take representations to be, how they are produced and what they're for:

1. Representations are dynamical re-productions of something else.
2. Re-production is achieved by reactivating neural marks of sensorimotor experiences.
3. Representation helps prediction and control.

However, this is only a general theory of *representation*, not of *thoughts*. Thoughts, as we saw, are representations that have constituent structure. Our current account will deserve to be seen as a general theory of thinking provided (1) we can show that analogical representations can possess compositional structure and (2) we can show that inference can use this constituent structure. The next two sections address these points respectively.

3.2. Structured analog representations

Hypothesis formation, exposition and evaluation in science is often based on analogy or metaphor (Holyoak & Thagard, 1995). Cognitive science is no exception, especially when the object is to define abstract and complex concepts such as representation and its digital and analogical subspecies. Dretske's conception of analogical representations, for instance, relies on a *photographic* metaphor. Analogical representations are compared to photographs and they

inherit most of their properties (but not all, which is why it is a metaphor). A photograph of a coffee cup can be interpreted as “this coffee is espresso” or “the cup is black”, and so on. Since photographs are patently non-compositional, it followed that analog representations were not compositionally structured (and this held for public as well as for mental analog representations). But in 1981, when Dretske first published *Knowledge and the Flow of Information*, 3D software, video games and Virtual Reality were as not common as they are today and, consequently, his metaphor could not be informed by these technologies. But ours can. VR representations (simulation, emulation) are analog. If we can show that VR representations can be compositionally structured, then we have showed that some analog representations could be structured.

Are VR images compositionally structured? First, look at the mechanisms for creating VR. VR and video games programming owe their development to the parallel development of computational geometry (CG), a field of computer science devoted to the „algorithmic study of geometric problems” (Mitchell, 1997, p. 1), and computer graphics. Computer scientists face several problems: how to (re)create 3D space and objects, and how to (re)create their evolution in time. These are exactly the problems faced by cognitive systems: how to predict, with internal machinery, the evolution of things and events. In CG as in nature, in order to predict, one needs to reproduce things/events and their evolution, what CGers call *modeling* and *simulation*. To model and simulate, computer scientists don’t rely on pictures databases, or on large amounts of unstructured images. They use combinatorial structures (Chazelle et al., 1999, p. 7). VR-production systems can recursively generate a potentially infinite number of complex combinations.

Brains are natural VR-production systems able to generate simulations of different objects, relations, events, etc, and combine them. In fig. 1 we see how object simulators in long-term memory (BALLOON, PLANE, and CLOUD) generate tokens of analog representations in working memory while combining with spatial relations simulators (ABOVE, LEFT).

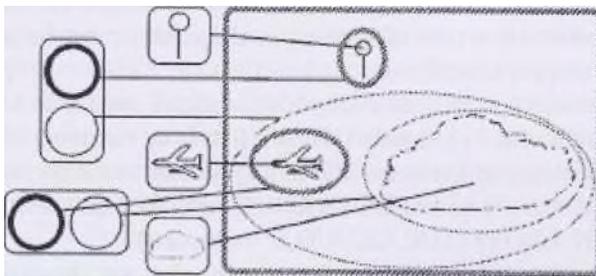


Figure 1. Complex simulation (from Barsalou, 1999) Simulators are represented on the left, simulations on the right.

Metaphors, however instructive, are by no means conclusive. We need to show that VR analog representations have compositional structure, or constituency. To do so, we first state the four criteria for compositionality, and then show that VR analog representations satisfy them. The four criteria for compositionality are mereology, combinatoriality, productivity and systematicity. The remainder of the present section is devoted to show that VR analogical representations are mereological and combinatorial. Section 3.3 will strive to show that they are productive and systematic when used to draw inferences.

According to compositionality, real constituency is about parts and wholes (Fodor & Pylyshyn, 1988). Then compositionality is, *stricto sensu*, a genuine mereological relation (like the relation between a wheel and a car). Compositionality is not a set-theoretic relation, such as the relation between a car and the set of cars in the parking. The basic predicate of mereology is proper parthood. Proper parthood has four individually necessary and jointly sufficient properties; it is an irreflexive, asymmetrical, transitive and supplemented relation (See Casati & Varzi, 1999). “Supplemented” means that if *x* is a proper part of *y*, then there is at least another part *z*, distinct from *x*. Mereological relations („is a part of”) are thus distinct from set-theoretical relations (“is a member of”): while there is nothing illogical in the idea of an empty set or a one-member set, a whole with only one part (or no parts) is logically impossible. There must be at least two parts in a genuine mereological whole; if thoughts are mereological wholes, they must be made out of at least two parts, which is precisely what functionalism supposes: that mental states are wholes made up of functions and arguments.

Simulations are mereological because they satisfy Casati and Varzi’s criteria for a genuine mereological whole. A simulator (the CLOUD simulator, for instance):

- is not a part of itself (irreflexivity)
- is a part of a thought, but the thought is not a part of the simulator (asymmetry)
- If the simulator is a part of a thought which is itself a part of another thought, then the simulator is also a part of the second thought (transitivity)
- If the simulator is a part of a thought, then there must be at least another simulator (supplementation).

Now, what about combinatoriality? Simulations are combinatorial because they satisfy Fodor and Pylyshyn’s (1988) criteria for combinatoriality:

- there is a distinction between structurally atomic and structurally molecular representations (BALLOON, ABOVE and CLOUD are atomic while BALLOON ABOVE THE CLOUD is molecular)
- molecular representations have constituents that are themselves molecular or atomic (BALLOON ABOVE THE CLOUD is made out of BALLOON, ABOVE and CLOUD)

— the semantic content of a (molecular) representation is a function (i) of the contents of its parts, together with (ii) its constituent structure. (The meaning of BALLOON ABOVE THE CLOUD is a function of the simulators for BALLOON, ABOVE and CLOUD)

Since they are both mereological and combinatorial VR, analog representation are structured representations. Therefore, we should not restrict discreteness and structure to digital representations (linguistic or not): simulators are structured analog representations. A remark about discreteness is in order. Because of a strong association between “digital” and “discrete” (as in Dretske’s account), analog representations have been conceived as *continuous* (a notable exception is Sloman, 1971). But “analogicalness” is not about being non discrete; it is about being non digital. Analog representations share similarity with their referent⁵, whether they are structured or not is another story.

3.3. Analogical inference

A compositional system is structured (mereologically and combinatorially), but mereology and combinatoriality are not sufficient. Compositional systems are also productive and systematic. A clear criterion for the systematicity and productivity of cognitive systems, which is also a strong constraint on the theory of thought, is inference. We argued that analogical representational systems can be both mereological and combinatorial. We now argue that analogical representational systems can produce inferences that are sensitive to the constituent structure of its representations and that such systems are systematic and productive (we don’t argue that the converse holds; whether it does or not is not important here). Analog representational systems will then have been shown to possess the four attributes we established as criteria for compositionality. Such systems, then, would truly think.

Thinking is an inferential activity that is sensitive to the constituent structure of the representations manipulated. Given a structured (mereologically and combinatorially) input representation, a system that thinks will output another structured representation according to rules that apply in virtue of the representation’s structure. These rules are transformation rules that specify how the transformation should be done. It is the conformity to these rules that makes transformations correct or not. Inference is thus an epistemic action that can be “kosher” if it conforms to some epistemic virtue.

⁵ Sharing similarity is neither necessary nor sufficient to qualify as a representation (see Dennett 1981, Putnam 1981): A counts as a representation of B insofar as it is used by a system as a representation for B.

On that view, thinking, constituent structure-sensitive inference, is a *deductive* process. Thinking systems generate new patterns of information that they don't have to gather from nature. They are looking for the consequences of the information they already possess, looking for what that information implies. To argue that non-linguistic analogical cognitive systems deductively infer, that their analogical simulations follow deductive patterns, a construal of "deduction" is needed, a construal that is slightly different and broader in sense. Accordingly, deduction will be construed here as the following:

a mental transformation from one representation to another according to transformation rules (1) that apply to the constituent structure of the representation and (2) respect some epistemic virtue.

For some philosophers, such as Bermudez (but also Frege, Davidson, Fodor, etc.) the epistemic virtue to which these transformation rules must conform can only be, *formal*, or, more precisely, logico-syntactic. A good inference is an inference whose structure conforms to classical logic. Such a stance makes non-linguistic inference a chimera, because non-linguistic representations cannot be syntactic:

We understand inference in formal terms — in terms of rules that operate upon representations in virtue of their structure [Note: more precisely, logico-syntactic structure]. But we have no theory at all of formal inferential transitions between thoughts which are not linguistically vehicled. Our models of formal inference are based squarely on transitions between natural language sentences (as codified in a suitable formal language) (Bermudez, 2003, p. III)

The problem is not that we have no theory "of *formal* inferential transitions between thoughts which are not linguistically vehicled", but that we have no theory of inferential transitions *tout court* between non linguistic thoughts. A formalist view of inference equates "good inference" with "formally valid inference". But formalism is not the only option. Non-formalists such as Brandom or Sellars consider the inference from "it's raining" to "the streets will be wet" as materially valid, where the transformation rules are not syntactic but semantic. *Semantic inferences* are drawn according to other rules, less general than logical rules: the particular rules for applying a certain word, dictated by socio-linguistic conventions.

Other non-formal rules of transformation exist. Suppose you are a bartender and a client, to whom you served a martini, says to you: "this glass is empty, captain!". He didn't only want to inform you of a physical fact, but was asking for another drink. This is an example of a *pragmatic inference* (Grice, 1989; Sperber & Wilson, 1986) in which the intentions of speakers are inferred from the illocutionary content (here, a request) of their utterances and the conversational context (the bar). Understanding the client's assertion implies that you understand

the structure of the speech act, in this case, a request, which is a kind of directive speech act (Bach & Hamish, 1979). Pragmatic inferences are genuine constituent structure-sensitive inferences: insofar as an utterance is categorized as a token of a particular speech act, the inferential understanding of the token is deduced from the condition of use of this type of structure. Syntactic, semantic and pragmatic inferences share a common feature: their transformation rules are linguistic (logic, linguistic conventions and illocutionary understanding). But are structure-sensitive transformation rules only linguistic?

Goodness of inference relies partly on the capacity to recognize relevant input. It begins in perception, where selective attention filters out irrelevant properties. To draw a syntactic, semantic or pragmatic inference, the inferential process must be sensitive to some relevant properties of the representation, be it its logical connective, the meaning of words or the type of the speech act involved, and draw a conclusion from these relevant structural properties.

According to Relevance Theory (Sperber & Wilson, 1986; Wilson & Sperber, 2004), looking for relevance is a fundamental characteristic of human cognition. An input is relevant insofar as it yields a *positive cognitive effect*, that is, “a worthwhile difference to the individual’s representation of the world - a true conclusion, for example” (Wilson & Sperber, 2004, p. 608). If an input, inferentially processed, yields relevant information, then it is a relevant input. Note that relevant *transformation rules* are also needed: even if one has relevant inputs, non relevant processing will not yield positive cognitive effects. Wilson and Sperber hold that the evolution of our cognitive system is geared toward efficient processing of relevance. The human cognitive system has evolved to make our perceptive, mnemonic and inferential processes geared toward the processing of relevant information.

Relevance Theory seems a little anthropomorphic. If we acknowledge, as we already did, that non-linguistic creatures can represent, then there is no reason to believe that non-conversational aspects of Relevance Theory don’t apply to non-linguistic cognition. Relevance is not only useful in linguistic settings: in perception or memory retrieval, relevance is beneficial for all representational creatures. Predators must rely on relevant cues to hunt their prey, and must process them relevantly. Relevance Theory already has a name for this basic kind of inference: *contextual implication*. A contextual implication is a conclusion drawn from the input and the context, but neither the input nor the context is sufficient for drawing the conclusion. Turning a lamp on and seeing that no light comes from the lamp, you infer that the light-bulb isn’t properly working anymore. The light-bulb’s defect is inferred from the input (looking at the lamp) and the context (your attempt to turn the light on).

Contextual implication is the kind of thinking accessible to non-linguistic organisms. Cognitive systems able to run Barsalou-style multimodal simulations

can draw deductive inferences, however limited they may be. When some perceived object, event or agent x triggers the X simulator, the X simulation is a categorical simulation: x is perceived as a token belonging to a certain type. Its belonging to that type makes deductive inference possible on the basis of the type's properties when the X simulator, together with other simulators, constitutes a complex representation. The X simulator can be transformed into another one. Thus there are two elements in the simulationist theory of representation: a claim about representation formation (compositional combination of simulators) and one about transformation (the temporal evolution of the simulations), and the second follows naturally from the first.

Analog compositionality differs trivially from digital compositionality in being applied to analog instead of digital representations. But the two also differ non-trivially. First, analog compositionality is restricted to categories and relations that can be simulated by multimodal simulators. With their mnemonic and sensorimotor apparatus, we could expect animals and non-linguistic human babies to only (or mostly) have basic-level categories (Rosch, 1978), and this is so for at least two reasons: (1) basic-level categories are recognizable on the basis of schematic analog properties and (2) because of their multimodal content (gestalt perception, imagery, etc.), basic-level categories are more easily simulated by multimodal brain reactivation. Another distinction between analogical and digital compositionality is generality. It is plausible to assert, on empirical ground, that non-linguistic minds mostly have no intermodular fluidity (Hauser & Spelke, forthcoming). In such beings, information hardly flows from one module to another, while this is something that easily occurs in us, enculturated apes. Hence, the generality constraint (GC, Evans 1982) may hold only inside modules: a module may satisfy GC - to think that Fa one must be capable of thinking Fb and Ga - but only if a , b , F and G are in the actual domain of the module. With language, material culture and science, humans have access to theories and to expert knowledge.

5. Conclusion

We argued for a view of thinking that lies between the classical or received view defended, in one form or another, by most researchers, scientists and philosophers alike, and another view which we find in Bermudez's latest book, *Thinking without words*. The classical view emphasizes language as the model for thought, and the „success semantics” view put forth by Bermudez (2003), which insists on the digital nature of thought. Our view stands with Bermudez's view being, we believe, more empirically founded, less species-specific and “adult-oriented” (meaning less geared towards human adult cognition) than the classical view.

It also stands with Bermudez's view in thinking that the digital versus analogical dimension is the most important dimension in which to think about thinking. This last attitude marks a major shift in contemporary thinking about thinking, which, in part because of its association with the philosophical school of analytical philosophy, a school that methodologically emphasized *a priori* linguistic analysis, has mostly seen thinking as an exclusively linguistic phenomenon. The shift in contemporary Anglo-American philosophy, away from linguistic analysis and towards philosophical naturalism, has left the dominant position of language unchanged, because, we believe, of the major influence of cognitive science in naturalistic philosophy and the central position of Chomskyan linguistics in the cognitive revolution (at least until recently). If thought *is* linguistic, then Chomskyan linguistics offers naturalistically inclined philosophers and researchers a powerful model, unrivalled in depth and breadth, with which to think about thinking. But, as emphasized by Bermudez, the linguistic model has the major drawback of making thought a uniquely human adult phenomenon⁶. We agree with Bermudez that this drawback is serious enough to demand a rethinking of the traditional position on thinking in contemporary Anglo-American philosophy. And with Bermudez, we strive to develop a view of thought that applies to babies and animals, making human adult thought a special case of the model rather than an unexplained cognitive innovation or a suspicious paradigm.

On the other hand, although, for lack of space, we did not argue for this here, our view also stands with the classical view in thinking that thoughts, when digital, are thus *because* they are linguistic; that it is the linguistic character of some thoughts that makes them digital. Bermudez's view, we believe, posits another dubious cognitive first: the appearance of digital thoughts. Unlike many naturalistic philosophers today, such as Bermudez, we do not look to mainstream cognitive science for our main source of empirical constraints on philosophical theory, but to an emerging sub-genre: evolutionary developmental cognitive neuroscience. Over and above well-heeded constraints of cognitive psychological and neurological plausibility, evolutionary developmental cognitive neuroscience demands that our philosophical understanding of thoughts lives up to current understanding of the ontogeny and phylogeny (especially not forgetting the important interplay between the two) of cognition. Since Bermudez strives to open up the conceptual space surrounding the notion of thinking to include non-

⁶ If the Language-of-Thought view is adopted, then this problem wanes, but another one appears: thought generally, including in animals and human babies, can only be studied through the prism of human adult thought. This truly seems to be putting the cart before the horse. Human adult thinking, perhaps the crowning achievement of hominid evolution, should not be used as the model of all thinking: it should instead be *explained* in terms of more basic forms of thinking (evolutionarily, developmentally).

human and non-adult thought, it might seem strange to accuse him of not paying proper attention to ontogeny and phylogeny. Indeed, we sided with him against the linguistic view of thought especially because of Bermudez's insistence that our view of thinking makes room for baby thoughts and monkey thoughts. But, by doing so, Bermudez opens Pandora's proverbial box: once the door to such forms of thinking is opened, naturalistic philosophers must pay attention to current evolutionary and developmental theory, and such a stance, we maintain, argues against the digital model as a view of thinking generally.

In short, we have defended a view of thinking that in some sense stands between the classical view and Bermudez's view. We supported this view, which we called the spatial-motor view, in two steps: a conceptual step, in which we presented the three views, insisting on their defining features. This step served to highlight what the views share and what distinguishes them, thereby painting the conceptual landscape in which people have thought about thinking. The second step was empirical: we defended the central claim of our view, that is, the idea that spatial-motor analogical representations can have compositional structure. Note that our philosophical inclination towards naturalism makes this the central step of our defence of the spatial-motor view. We are perfectly prepared to let our model stand or fall with relevant empirical data.

References

- Allen, C. & Bekoff, M. (1997). *Species of Mind. The Philosophy and Biology of Cognitive Ethology*. Cambridge, MA: MIT Press.
- Application Challenges to Computational Geometry. The Computational Geometry Impact Task Force Report, B. Chazelle et al., *Advances in Discrete and Computational Geometry, Contemporary Mathematics*, 223, AMS, Providence (1999), 407-463.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
- Barsalou, L.W., Simmons, W.K., Barbey, A., & Wilson, C.D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7, 84-91.
- Berlucchi G. and Aglioti S. (1997) The body in the brain: neural bases of corporeal awareness. *Trends Neurosci.*, 20:560-564.
- Bermudez, J.L.(2003) *Thinking Without Words*. Oxford University Press.
- Brooks, R. A., „Intelligence Without Representation”, *Artificial Intelligence Journal* (47), 1991, pp. 139-159.
- Carroll, L., „What the Tortoise Said to Achilles” *Mind* 4, No. 14 (April 1895): 278-280. Clark, A. (1997). *Being There*. Cambridge (Mass.): MIT Press.
- Cummins, R. (1996) Systematicity. *The Journal of Philosophy*, 93: 591-614.
- Dennett, D.C. (1995) The Role of Language in Intelligence in „What is Intelligence?”, *The Darwin College Lectures*, ed. Jean Khalfa, Cambridge, Cambridge Univ. Press.

- Dretske F. (1981), *Knowledge and Flow of Information*, Cambridge (Mass.): MIT Press.
- Forkman B. (2000) Domestic hens have declarative representations. *Animal Cognition* 3: 135-137
- Gallistel CR (1990) *The Organization of Learning*, Cambridge MA: The MIT Press
- Grice, H. Paul. 1989. *Studies in the Way of Words*. Harvard University Press, Cambridge MA .
- Grush, Rick (2003). In defence of some «Cartesian» assumptions concerning the brain and its operation. *Biology and Philosophy* 18:53-93
- Grush, Rick (2003,forthcoming). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*.
- Grush, Rick (to appear). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*
- Harman, G. (2002) „The Logic of Ordinary Language,” an attempt to distinguish logical principles from others, in *Common Sense, Reasoning, and Rationality*, edited by Renee Elio (Oxford: Oxford University Press, 2002), pp. 93-103.
- Kirsh, D. Today the Earwig Tomorrow Man? *Artificial Intelligence*. (1991). Reprinted in *Artificial Life*, ed M, Boden, OUP. 1996
- MacDorman, K. F. (1999). Grounding symbols through sensorimotor integration. *Journal of the Robotics Society of Japan*, 77(1), 20-24.
- Maravita A. et al. (2003) Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.*, 13:R531-R539
- Maravita, A. & Iriki,A. (2004) Tools for the body (schema) *Trends in Cognitive Sciences*, 8:79-86.
- Mitchell, J.S.B. (1997) Introduction: Special Issue of *Algorithmica* devoted to Computational Geometry and Manufacturing, *Algorithmica* 19: 1-3.
- Proust, J. (1999) Mind, space and objectivity in non-human animals. *Erkenntnis*, 51, L 41-58.
- Relevance Theory*. Wilson, Deirdre & Sperber, Dan. In Horn, L.R. & Ward, G. (eds.) (2004) *The Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Richard S. Sutton. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning*, Austin, TX, 1990. Morgan Kaufmann.
- Richard S. Sutton. Planning by incremental dynamic programming. In *Proceedings of the Eighth International Workshop on Machine Learning*, pages 353-357. Morgan Kaufmann, 1991.
- Roger C. Conant and W. Ross Ashby, „Every Good Regulator of a System Must be a Model of that System”, *International Journal of Systems Science*, Vol. 1, No. 2 (1970),
- Shaun Gallagher and Jonathan Cole. 1995. „Body Schema and Body Image in a Deafferented Subject,” *Journal of Mind and Behavior* 16 (1995), 369-390; reprinted „” in *Body and Flesh: A Philosophical Reader*, ed. Donn Welton (Oxford: Blackwell), pp. 131-147.
- Sperber, Dan & Wilson, Deirdre. 1986a. *Relevance: Communication and Cognition*. Blackwell, Oxford and Harvard University Press, Cambridge MA . (Second edition 1995. Blackwell, Oxford .)

- Stein, L. A., „Imagination and Situated Cognition,” *Journal of Experimental and Theoretical Artificial Intelligence*, 6, 393—407, 1994. Reprinted in *Android Epistemology*, K. M. Ford, C. Glymour, and P. J. Hayes, eds., AAI Press/The MIT Press, 1995, pp. 167—182
- Wimsatt, W.C. 1986. Forms of aggregativity. In Donagan/Perovich Jr./Wedin, *Human Nature and Natural Knowledge*, 259—291.