Unconscious representations 2:

Towards an integrated cognitive architecture.

Luís M. Augusto

Institute of Philosophy, University of Porto / FCT

Axiomathes. February 2013. DOI: 10.1007/s10516-012-9207-y

(This is the author's pre-print version; the published paper is available at

http://rd.springer.com/article/10.1007/s10516-012-9207-y)

Abstract: The representational nature of human cognition and thought in general has been a source of controversies. This is particularly so in the context of studies of unconscious cognition, in which representations tend to be ontologically and structurally segregated with regard to their conscious status. However, it appears evolutionarily and developmentally unwarranted to posit such segregations, as, otherwise, artifact structures and ontologies must be concocted to explain them from the viewpoint of the human cognitive architecture. Here, from a by-and-large Classical cognitivist viewpoint, I show why this segregation is wrong, and elaborate on the need to postulate an ontological and structural continuity between unconscious and conscious representations. Specifically, I hypothesize that this continuity is to be found in the symbolic-based interplay between the syntax and the semantics of thought, and I propose a model of human information processing characterized by the integration of syntactic and semantic representations.

Keywords: human cognitive architecture; Classical cognitivism; conscious/unconscious mental representations; syntax & semantics of thought

1. Introduction

Since at least the end of the 19th century, it has become a central issue to explain in which way(s) human cognition can take place unconsciously, or without awareness. Its understanding appears to be fundamental, as, for all we currently know, we are beings such that we can end up with an unconscious mental life alone (e.g., in coma and in the permanent vegetative state), but never with a conscious mind only. The question has become more pressing with the advent of artificial intelligence (AI) and its attempt to produce implementable models of the human cognitive architecture, which is now widely believed to be composed of both conscious and unconscious processes and representations (e.g., Anderson, 2005; Newell, 1990; Sun, 2002).

Unconscious cognition appears to be ontogenetically and perhaps also phylogenetically prior to consciousness, this being perhaps a transformation of, or addition to, the former (e.g., Reber, 1992a; 1992b). It is thus justified to believe that they must retain some ontological and/or structural continuity. But in the academic quarters where unconscious cognition is now widely accepted, it is still more often than not seen as basically distinct from conscious mental processing, with differences rather than continuities being emphasized (see Augusto, 2012; see Part 1 of this article, Augusto, 2013). The main tenet that the abundant bibliography on unconscious mentation tends to support is that unconscious cognition is *qualitatively* differential with relation to conscious modes of processing, being by and large more automatic (less controlled), more bottom-up (less top-down), less flexible, more durable, and altogether more procedural (less declarative) (for details, see Augusto, 2010; Dienes & Berry, 1997). These dissociations are well established, but if strictly or narrowly taken, are highly problematic (see Augusto, 2012) and they need not, or should not, reflect an ontological and/or structural gap or segregation.

This is notably concretized in the traditional model of human cognition (TMHC), which postulates strict segregations between the levels of information processing, the entities implicated, and the (un)consciousness status of the processes/representations. In particular, the TMHC sees higher-level cognitive processes as calling for conceptual/symbolic representations and, therefore, consciousness, whereas lower-level processes are seen as typically unconscious and non-conceptual/sub-symbolic (see Part 1 of this article, Augusto, 2013). However, the direct consequence of this segregation is that many psychological phenomena remain opaque to our understanding. For instance, the TMHC cannot account for the following Cases I-IV:

I. Infants as young as 3 months old can learn to control contingent events (e.g., by repeating certain actions) and to make predictions, as suggested by their emotional reactions to the success or failure of the supposed predictions (see, e.g., Papoušek, 1967).

II. Experimentally conditioned subjects can accurately predict a specific stimulus without being aware of that or even against their conscious expectations (e.g., Perruchet, 1985).

III. People with cortical blindness over their entire visual fields can navigate physical obstacles in a wholly new environment, an action that requires good planning and predicting, unaided (e.g., de Gelder et al., 2008).

IV. We more often than not make judgments about people, i.e., we make decisions regarding them, based on stereotypes and attitudes of which we are wholly unaware (see, e.g., Dion et al., 1972).

These and similar[1] Cases suggest, contrary to the TMHC, that the human cognitive architecture is integrated in the sense that conscious and unconscious processes/representations tango in normal cognition. From a contemporary perspective on cognitive architectures, specifying requirements such as ecological, bio-evolutionary, and cognitive realism (see Sun, 2004), it thus appears essential to elaborate on a continuity between unconscious and conscious mental processes/representations. This paper is an elaboration on one such ontological and structural continuity. Assuming a by-and-large Classical cognitivist perspective, I show that strict ontological and/or structural distinctions between conscious and unconscious representations are unwarranted, and hypothesize that the continuity between conscious and unconscious representations resides in computational aspects, where by “computation” I mean the symbolic-based (i.e., with, from, and into symbols) interplay between the syntax and the

[1] See Augusto (2010); Shanks (2005).

semantics of thought.[2] Finally, I assess the explanatory power of this approach by applying it to Cases I-IV above.

2. Cognition as integrated information processing

2.1. Preparing the theoretical and terminological terrain

It seems to us that our minds work by picking things from the environment and the self (e.g., this is fire; that is a tree; the lion is running towards me; I'm thirsty; etc.) and by instructing us as to the actions these things motivate (e.g., don't touch it; seek shadow or food; run!; search for water and drink; etc.) in such a way that when this goes right, we secure our well-being and survival, but when this goes wrong, both well-being and survival are at risk. Overall, this goes right, and we may believe this is so because it has evolved in that sense.

---

[2] I do not elaborate on a complete cognitive architecture, which must contemplate, among others, memory as an essential constituent; here, I focus almost exclusively on the representational, symbol processing component of the human cognitive architecture (see Figs. 3 and 5 below for schematic views). The aim is to elaborate on a theory of human cognition, and thus possible physical (i.e., machine) implementations are not primarily considered. Nevertheless, features of the symbol processing system here described are in principle compatible with implementable cognitive architectures such as Soar (Laird, 2012; Newell, 1990) and ACT-R (e.g., Anderson, 1996; 2005; see also Anderson, 1983), which are based on symbolic approaches and classify themselves as production systems (vs. schema systems, for example). Most importantly, in these architectures both procedural/implicit and declarative/explicit processes or levels are symbolic and expressed in a symbolic form. An example of a cognitive architecture not compatible in principle with the symbol processing system here described is CLARION, because of its dual representational structure, with procedural/implicit processes captured by sub-symbolic, distributed representations (i.e., a connectionist model) and declarative/explicit processes seen as symbolic in nature (a cognitivist or computational model) (see, e.g., Sun, 2002).

Millennia (at least two and a half) of speculative and empirical efforts have not assured us that, and as to how, this seemingness corresponds to the facts, and we suspect that terms such as “picking”, “instructing”, and “things” are inadequate to describe and/or explain how our minds succeed in keeping us alive and kicking in a highly dynamic environment with which we are up to a point continuous and which can change in unpredictable and often threatening ways. Let us accept that the ‘things’ the human mind ‘picks’ and further ‘transforms’ from the environment and the self are *information*, thus making it an *information processing system*. Then, an epistemologically informed way to say the above is that the human processing of information yields knowledge; a more psychologically informed way is that human minds are information processing systems dedicated to cognition.

I focus here on the latter perspective in such a way that cognition and knowledge can — with Neisser (1967; 1976) and Newell (1980; 1990) — be seen as more or less synonymous. In this perspective — *cognitivism* (see, e.g., Haugeland, 1981) —, processing of information can be understood as *manipulation of symbols*. I call this *computation*, too (without necessarily making reference to the digital computer). A controversial but still useful way to refer to this whole picture is *representation* and to see cognition as a representational affair. In this view, one aspect matters in a prominent way: to explain how extracted information becomes representations in the mind, i.e., how it becomes meaningful for us.

I shall argue that the answer resides in the interplay between syntax and semantics, believed here to be the two faces of human information processing in the following sense:

Human information processing = (symbolic representation = manipulation of symbols = computation) = syntax-semantics interplay

A fundamental aspect to take into consideration in a scientific approach to the mind from the viewpoint of cognition is that diverse structures, factors, and/or levels all work together for the same ends, i.e., cognition and behavior at large. Cognition is thus a unified activity. As Newell (1990, p. 17) put it, “a single system (mind) produces all aspects of behavior. It is one mind that minds them all [parts, modules, components, whatever].” With this in mind, from the more restricted viewpoint of information processing as computation or symbolic representation, normal human cognition can be seen as the integration of syntactic and semantic processes and properties. In other words, both kinds of processes and properties must contribute equally importantly to a normal, integral flow in the information processing (see below for details). But if cognition is an integrated, unified processing in the senses above, then it appears reasonable to postulate that both conscious and unconscious representations must share (parts of) the same syntax and semantics. This simplifies the approach to the dissociation between both kinds of representations, leaving us with the task of solving the many problems posed by this dissociation (see, e.g., Augusto, 2012) by looking for the computational correlates of (un)consciousness. In the next paragraphs, I first elaborate on this notion of information processing as symbolic representation, and then settle the meaning of *syntax*, *semantics*, and *syntax-semantics interplay* in this paper.

2.2. Symbols and symbolic representation

Soon after Shannon (1948) inaugurated the field of information theory proper, Weaver (1949) advocated that this mathematical viewpoint of information should be complemented by a more strictly semantic analysis concerning meaning and truth, as well as by a psychological approach studying the way information and its processing impacts on human behavior. Newell (1980) unified these three aspects through the notions that human information processing is representational, and that representing is tantamount to manipulating symbols, meaningful or designating entities. According to him, for some entity *X* to be a *symbol*, i.e., to have meaning, means that it designates or represents (also: refers to, denotes, is about, intends, stands for, etc.) an entity *Y*, or aspects thereof, with relation to a specific process or behavior *P*. What is the same: *P* depends on *Y*, but it takes *X* as mental input. For instance, the feeding behavior (*P*) of a human depends on taking (i.e., inputting) a distal stimulus, say, an apple (*Y*) as a proximal stimulus indicating a source of food (*X*) — and this independently of whether *Y* is present or not (for humans, as well as for many — most? — other animals, to wait for the presence of the stimulus to feed would soon lead to starvation). This entails that for a process or behavior *P* that depends on *Y* to be carried out, representing *X* is tantamount to having *Y*.

In this sense, symbols are synonymous with — internal tokens of — concepts, images, ideas, words, etc., and their manipulation, or computation, just is the rule-governed ways in which symbols can be activated/inhibited, associated to, and combined with, other symbols, it so being that these processes are determined by the very nature of the entities they operate on. In other words, representational processes are symbolic processes. Thus, by saying that a process is *symbolic*, it is meant that it operates, or is carried out, *on* and *with* symbols, i.e., it takes symbols as inputs, it processes symbols by means of other (perhaps more basic, or lower-level) symbols, and it outputs symbols. For instance, and going back to the introductory examples above, inputting the symbols "lion" and "the lion is running towards me", one should output something like "run for dear life!" (where the quotation marks isolate specific mentioned symbols from the rest of the symbols that compose this paragraph), if not consciously — for example, for time constraints —, then as commands to one's leg muscles.

The above entails the presupposition that symbols come in many formats, from the mere reflex command to a muscular nerve to a full-fledged concept such as LION, where by *full-fledged* I mean that a concept can be stored in a lexical and/or semantic memory base and allows (but does not necessitate) conscious recall and/or recognition, and by *reflex* one should understand that a symbol is in principle not apt for conscious manipulation.[3]

2.3. Syntax and semantics

Whether mental (e.g., concepts, images, etc.) or also physically realized (e.g., written/spoken words or sentences, traffic signs, gestures, etc.), symbols have both rule-governed form and content. Traditionally, the rule-governed way symbols are formed

[3] There may be kinds of reflex behavior that are not symbolic in essence; for instance, brainstem reflexes and motor responses preserved in the vegetative state. This might mean that "reflex" is not an all-or-none category, with visceral and somatic reflexes preserved in a state of deep or total unconsciousness being 'more reflex' than those that are preserved only in some minimal degree of awareness, such as in the minimally conscious state. However, this minimal degree of awareness does not entail necessarily some degree of conscious information processing (see the discussion motivated by Merker, 2007, in Part 1 of this article, Augusto, 2013).

and manipulated has been defined as the syntactic aspect of symbolic systems, while the ways in which these symbols become meaningful (i.e., have content) are seen as their semantic side. This has been particularly so for linguistics, logic, and semiotics, which all emphasize, on the one hand, the syntactic property of well-formedness (well-formed sentences, propositions, formulas, signs, etc.), and, on the other hand, the semantic properties of (well-formed) symbols and symbol constructions (e.g., truth; denotation; etc.).

But, surely, our thoughts in general are also subject to rules of well-formedness and interpretation, and we even have a plethora of labels for ‘conditions’ falling under the category “thought disorders,” in which one or both kinds of rules are disrespected or suspended, with schizophrenia high on the list as a disorder of both the form and content of thought. In fact, it is now fairly well established that anomalies in the syntactic and semantic processing of symbols in general appear to be intimately connected in schizophrenia: patients make grammatical mistakes, err in making inferences, and fail to interpret correctly social and cultural signs in significantly higher degree and frequency compared to the non-clinical population (see, e.g., McKenna & Oh, 2005). It seems thus plausible that not only are there a *syntax* and *semantics of thought processes* that underlie the processing of natural language, reasoning, and behavior at large, but they harmoniously tango in the ‘normal’ mind. This is precisely the main thesis of cognitivism, concretized in the postulation of a ‘language of thought’ (see especially Fodor, 1975; Fodor & Pylyshyn, 1988), a sort of ‘Mentalese’ with a combinatorial syntax and a semantic structure similar, but not identical to, natural language (that is, a human not endowed with or exhibiting diminished verbal skills can still ‘speak Mentalese,’ or think ‘orderly’).

From a by-and-large Classical cognitivist viewpoint to be explained below (but see, e.g., Hanna, 2006, Chapter 4, for an introduction), in this paper I shall use *syntax* and *semantics* as referring to thought, or cognitive processes at large. Because we appear to think with concepts or words as basic units, these, as well as sentences and propositions expressed in natural language, will often illustrate the ideas elaborated on, but the reader is asked to bear in mind that *thoughts*, or simply *representations* or *symbols*, rather than strictly natural language objects and constructions, are primarily meant.

3. An integrated model of human cognition

3.1. The ontological continuity: Syntax, semantics, and symbols

In Part 1 of this article (Augusto, 2013), I showed that the TMHC, as well as its escorting vehicle and process theories, is contradicted by the experimental evidence. This suggests that conceptual representations can be processed both with and without consciousness; furthermore, higher-level processes can be carried out outside the grasp of consciousness, and lower-level, basically sensorimotor, processes may attain consciousness. The way to get out of this apparent impasse lies in an analytically-motivated artificial separation of the two otherwise inseparable faces (rather than levels) of the human processing of information, to wit, syntax and semantics. For the sake of the argument, let me start with a somewhat unorthodox illustration.

In an episode of the cartoon series *South Park*, in order to probe her mother's (supposedly feeble) intelligence, a young girl asked her what eight times nine (or any two other numbers) was. Her mother told her not to be silly, as eight and nine are two totally different numbers. The young girl sighed, but: are not eight and nine really two totally different numbers? So, how come one gets seventy-two, another totally different number, out of them by multiplying them by means of something that can be graphically represented as "x", if anything even more different from these numbers than they themselves are with relation to each other? What, indeed, unites these entities ontologically and/or structurally, so that they relate to each other in this puzzling yet efficient way?

This undoubtedly appears to call for a lengthy philosophical discussion, but for the ends of this paper I propose a positive answer: firstly, what *unites* them is that they all are *symbols*, where symbols are to be understood in the sense exposed above; secondly, they are symbols that can co-occur in a meaningful way (here: "8 x 9 is 72"). What *separates* them is that "x" is a symbol that is interpreted as suggesting or denoting a rule (i.e., it triggers an operation), while "8" (or "9") is a symbol that is interpreted as denoting or referring to some sort of entity; more generally, we might also say that while "x" suggests or calls forth a process, "8" suggests or refers to some content.

Thus, by *co-occurring in a meaningful way*, I mean that in the co-occurrence "8 x 9", "x" triggers a rule that operates on the atomic structures "8" and "9" in such a way that another structure, "72", occurs as the output. This is to say that information is processed both *syntactically* and *semantically*, with syntax and semantics being like the two sides of a sheet of paper. This analogy means simply that syntax and semantics are the two sides or aspects of the same phenomenon: one cannot think in the absence of any of these two. More specifically, both syntax and semantics operate *on* and *with* symbols, which entails that whatever it is that distinguishes syntax and semantics must reside in how they relate to the symbols the human mind operates on and with. According to this, syntax is the set of *symbol rules*, and semantics is the interpretation of symbols into *symbol structures*, i.e., their becoming what we commonly call concepts and propositions.

As seen, this entails of necessity a kinship. Otherwise, if one postulates that rules (syntax) are wholly distinct from the structures (semantics) they operate on, then one falls back into the vehicle vs. process disputes, as positing an ontological distinction between them leads inevitably to the question of how then two sorts of entities that are believed to be *so* different (after all, it is at the ontological level that this distinction is postulated) can 'work it out' together. The fact is, symbol rules have themselves meaning, or, to put it differently, are themselves symbol structures of some sort ("x" is a symbol rule that 'means' multiplication, i.e., it operates on specific symbol structures, numbers, by scaling one by another), which is to say that syntax is proto-semantic; on the other hand, specific symbol structures call for specific symbol rules (as seen, "x" (or "+", etc.) is a symbol rule that is 'called forth' by symbol structures such as "8" and "9", but not by, say, "CAT" and "§"), which means that semantics must be in some way proto-syntactic. It is in this mutual *proto-ness* that they unite into a single phenomenon: *symbolic representation*.

This entails that, in human cognition at large, syntax and semantics are constitutive of each other, and that each is a condition not only necessary, but also sufficient, of the other. This is perhaps going farther than Classical, or standard, cognitivism is willing to go, and at least part of this claim has been notoriously argued against by Searle (1980; 1990) with what is known as the Chinese room argument.

Although Searle targets first and foremost strong AI, i.e., the position that sees the correct computer simulation of a mind as a mind de facto (vs. a model of the mind, as for weak AI), his argument actually draws on the view that posits a "gulf between form and content, between syntax and semantics" (Searle, 1990, p. 28). Briefly, the argument runs like this: let us imagine a room in which there are sheets of paper, pencils, files, and whatever else that is necessary for inputting and outputting written Chinese symbols; let someone who is an English speaker and knows no Chinese have in the room the right instructions in English on how to answer input questions in Chinese. While it is obviously possible that this person can, by following the right instructions, simulate an understanding of Chinese, in fact s/he does not understand this language, as the mere manipulation of formal symbols[4] (syntax) is not enough to generate meaning (semantics). In other words, manipulating symbols does not equate with understanding, thinking, reasoning, etc., or human cognition at large. Most importantly, this argument has the more or less implicit premise that syntax is wholly unconscious while semantics is, or can be, conscious, with the conclusion in mind being, of course, that computers cannot be conscious, because they are limited to syntax.

Attempting to refute this argument has already become a philosophical 'tradition' (see, e.g., Preston & Bishop, 2002), and I am not sure this paper fits into it in some way, though it is contrary to Searle's position. Searle's main premise is that symbols have no intrinsic semantics. Noticeable is the fact that this is a general premise, i.e., it does not distinguish between symbols in human minds from symbols in other symbol processing systems. It appears, in fact, that by "symbols" Searle means only either 0's and 1's, or squiggles (what Chinese characters are for him). This is a very limited conception of symbol, and a remarkably poor notion of syntax is one of its consequences. According to him, humans (or their brains) do, too, manipulate symbols, the difference with respect to the digital computer being that the human brain attributes meaning to the symbols it manipulates, which is only in accord with his premise that

[4] Note that Searle uses *symbols* and *formal symbols* as synonymous expressions; see next footnote.

symbols, per se, have no intrinsic semantics. This attribution of meaning (semantics) is for him "the mental contents biologically produced by the brain" (Searle, 1990, p. 30), an affirmation that, though obscure or opaque (perhaps because left wholly unexplained), leaves other thinking machines out by definition, and strong AI appears to be successfully falsified.

For a symbol to be attributed meaning, it has to be on (vs. off), and this is syntax at its most basic operation; in turn, a symbol cannot be on without having some sort of meaning, because it is part of the definition of a symbol that it has meaning, i.e., it intrinsically denotes, refers to, etc. *something* (see above). This is so no matter how 'formal' a symbol might be: for instance, the 'formal' symbol "x" in mathematics denotes (also: calls forth, carries out, motivates, etc.) (the operation of) multiplication. Translate "x" into so many 0's and 1's arranged in a specific, unique sequence, implement it in a computer program, and you still get multiplication. So we have it that symbols or, what is the same, symbol structures, are *by definition semantic*. But since meaning just is the operation or behavior that a symbol motivates or supports (see above), symbols are *by definition syntactic*, too. This suggests that a segregation between symbol rules and symbol structures is analytical rather than veridical, which weakens Searle's argument against computationalism in general, if not also his position against what he calls strong AI.[5]

---

[5] Searle makes a significant lapse in his (Searle, 1990), when, on page 27, he writes: "… symbols [in a computer program] are manipulated without reference to any meanings. The symbols of the program can stand for anything the programmer or user wants. In this sense the program has syntax but no semantics." The contradiction is obvious: how can symbols be purely syntactic if they always stand for something (i.e., have some meaning), be that at the will of the programmer or user? Note that this is what Searle appears to want to mean when he claims that symbols have no *intrinsic* semantics: they can be attributed

Let it be granted that I managed to settle satisfactorily — if only provisionally — the problem of the relation between the syntax and the semantics of thought; the $1,000,000-question now is: how does this bear on the distinction being discussed between conscious and unconscious mental representations?

3.2. The structural continuity: Syntax, semantics, and form

While Neisser (1967) propelled the advance of cognitive psychology to (in part) the expense of behaviorism in great measure because he postulated that human processing of information is representational, he did not specify the nature of the representations involved in cognitive processes. Fodor (1975) postulated that thought processes, or mental representations, are carried out in a 'language of thought,' or 'mental language' (a.k.a. 'Mentalese') that, like natural language, possesses a combinatorial syntax and semantics; this latter is said to be conceptual, the former logical in nature, and the relation between both is specified as "the semantic content of a (molecular)

---

any meanings whatsoever. But *some* meaning it must be, else there can be no talk of symbols, and the point is that while it then can be indeed said that symbols have no *intrinsic* semantics, being arbitrary/conventional as they are, it cannot be said that symbols have *no* semantics. Therefore, it cannot be counter-argued that Searle means that *for the program* the symbols have no meaning: the expression "symbols have no meaning" is senseless. Importantly, Searle (1) strictly narrows the meaning of *formal* so that this adjective is interchangeable with *meaningless*, and (2) he takes the expression *formal symbol manipulation* with *formal* qualifying *symbol*, whereas for computationalism, or Classical cognitivism, in general *formal* (1*) has to do with structure and (2*) it qualifies *manipulation* (i.e., symbols are manipulated in a structured way) (see below). From the perspective defended in this paper, it can be said that Searle confuses ontology (*formal symbol*) with structure (*formal manipulation*) of symbols and symbolic expressions.

representation is a function of the semantic contents of its syntactic parts, together with its constituent structure" (Fodor & Pylyshyn, 1988, p. 12).

By and large, the supporters of this language of thought hypothesis see it as distinct from its major rival approach, connectionism, in being "committed to 'complex' mental representations or to 'symbol structures'" (Fodor & Pylyshyn, 1988, p. 13), where the notion of *structure* is fundamental: mental processes operate upon (are sensitive to) mental representations that satisfy specific structural descriptions; these descriptions follow the formalism of symbolic logic. For example, the mental representation of the *form* "*P* & *Q*" can be operated upon in order to be transformed into the representation of the form "*P*" (or "*Q*"), by applying the inference rule of conjunction elimination (&E). Thus, a thought expressing the (complex) proposition

(1) Mary went to the library and brought home a few books.

can be transformed into the thought expressing the (atomic) propositions

(2) Mary went to the library.

and

(3) Mary brought home a few books.

This is so regardless of the number of atomic or complex constituents; the same inferential operation (conjunction elimination) can be carried out upon expressions like "*A* & *B* & *C*" or "(*P* & *Q*) & (*R* & *S*)", as represented in the following tree:

$$\cfrac{\cfrac{(P \mathbin{\&} Q) \mathbin{\&} (R \mathbin{\&} S)}{(P \mathbin{\&} Q)}\ \&\mathrm{E}}{P}\ \&\mathrm{E}$$

By the above, the standard, or Classical — as Fodor and Pylyshyn (1988) put it — cognitive science view of cognitive architectures attributes a central role to *expressions* (rather than — neural — nodes). The role of these structured symbolic expressions is actually even physically causal (the behavior of the system, realized in the brain, depends upon the properties of the symbolic expressions), but what interests us here is the notion of *symbolic structure* explored by the authors.[6] In order for a mental representation of the form, say, "$P \rightarrow Q$" to be structurally, and hence semantically, meaningful, both "*P*" and "*Q*" must be in a relation of *constituency*, in this case syntactically expressed by the logical connective for implication (→). Their position in the complex symbol "$P \rightarrow Q$" constitutes a syntactically defined specific structure that entails a specific, unique, semantic relation, i.e., "$P \rightarrow Q$" is not the same as "$Q \rightarrow P$", even though the atomic symbols are the same. This is to say that "*P*", "*Q*", and "→" are

[6] I am assuming that Fodor and Phylyshyn's conception of symbolic structure is not only compatible with, but actually complementary of, Newell's (1980; 1990) symbol ontology. It is in their unification that, as I see it, cognitivism can be a powerful explanatory approach to human cognition.

not only *activated* in a thought expressable as (of the form) "$P \rightarrow Q$"; they are also *in construction* in "$P \rightarrow Q$" in the sense that they are constituents of the complex symbol "$P \rightarrow Q$" with a specific place assignment geometrically representable as a tree. Put differently, the Classical representational view of cognition postulates that humans think by means of atomic and/or complex (also: molecular) symbols that are structured together in specific constitutionally-relevant and uniquely-analyzable complex ways that are meaningful. These complex symbols can be, for instance, propositions, or judgments, and the way they are structured determines the transformations (e.g., inferences) they can undergo.

Constituency entails further important features of human information processing. In Fodor and Pylyshyn (1988), it is postulated that to have a Classical cognitive architecture is also to have the ability that, once one acquires the rule "P & Q", one will consequently know that "P" and "Q", separately, by the inference rule of conjunction elimination (&E). This is by and large what can be called the *productivity* or *generativity* of the 'language of thought,' the ability of producing or generating an unbound corpus of symbolic structures, atomic or molecular, from a finite set of symbols.[7] For instance, from the representation above, "(P & Q) & (R & S)", which is obtained from the atomic representations "P", "Q", "R" and "S", one can obtain "P & (R & S)", "(P & Q) $\rightarrow$ (R & S)", "(P & Q) $\rightarrow$ (P & R)", etc. This can be represented (non-canonically) as in Figure 1.

[7] The notion of generativity here explored, in which introduction and elimination of logical operators play fundamental roles, is more immediately related to natural deduction in logical reasoning (see, for instance, Prawitz, 1965), rather than to Chomskyan generative grammar (e.g., Chomsky, 1957; 1959). This said, they are clearly kindred theories.

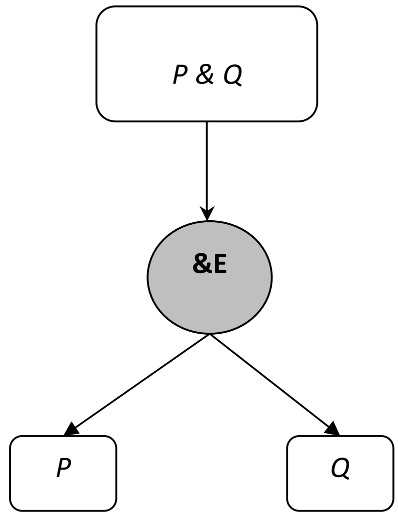

Figure 1. Symbolic constituency (internal structure) and productivity: the inference rule of conjunction elimination (&E) allows the production of “*P*” and “*Q*”, separately, from “*P* & *Q*”.

Productivity is connected to another essential property of Classical cognitive architectures, to wit, *systematicity*, in the sense that once one acquires the rule for conjunction elimination, one consequently acquires the rule for conjunction introduction (&I): from “*P*” and “*Q*”, one can infer “*P* & *Q*”. With other examples, one cannot understand a statement such as “John loves the girl” without understanding the statement “the girl loves John,” or understand “*aRb*” (read “*a* is *R-related* to *b*”) and not understand “*bRa*”. In sum, systematicity is the property of the ‘language of thought’ that allows those endowed with it — humans, but hypothetically other animals — to use it integrally, rather than just parts of it. Figure 2 illustrates this systematicity of the ‘language of thought.’

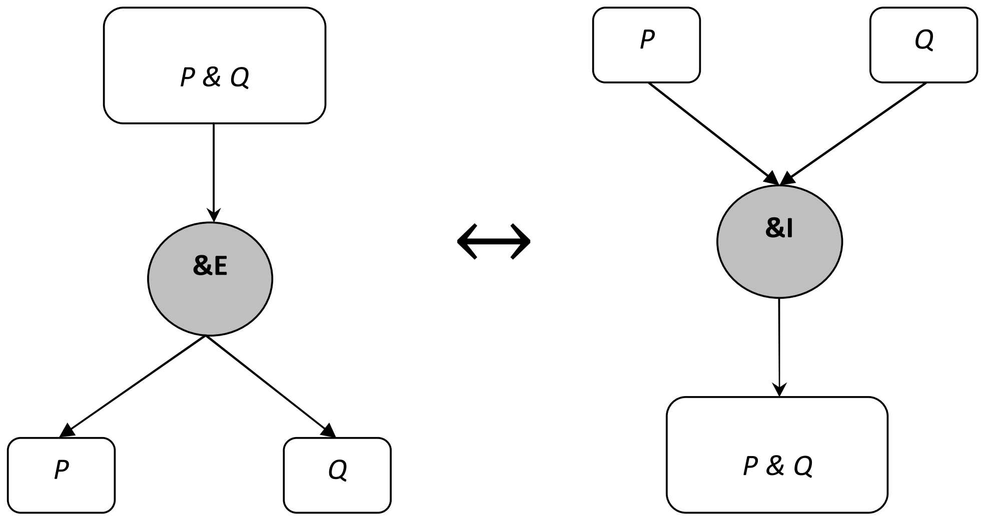

Figure 2. Systematicity of the 'language of thought.'

Two other fundamental features of mental representations according to the Classical cognitivist perspective, *compositionality* and *inferential coherence*, have already been touched upon, because, as a matter of fact, all four essential features are interconnected in ways that can be seen even in terms of identical phenomena. Compositionality, particularly connected to the feature of systematicity, just is the property that if one is able to understand, say, a proposition such as "John loves the girl" (expressing the thought that John loves the girl), then one must be able to understand a proposition like "the girl loves John" (expressing the thought that the girl loves John), because the constituents "John", "loves", and "the girl" make more or less the same semantic contribution to the two propositions (noun subject, verb, and noun object). In other words, "systematically related sentences are always semantically related" (Fodor &

Pylyshym, 1988, p. 43). This property becomes clearer if further complexity is introduced: for instance, the argument "(i) Turtles are slower than rabbits. (ii) Rabbits are slower than Ferraris. (iii) ∴ Turtles are slower than Ferraris." is valid not because of transitivity, primarily, but for the fact that the expression "are slower than" makes the same semantic contribution in the two premises and in the conclusion.

As for inferential coherence (or inferential systematicity), I touched upon it in the tree above depicting the processing from "(P & Q) & (R & S)" to "P". In cognitive terms, this means that if one has the ability to infer "P" from "P & Q", then one must have the same ability to infer "P" from a similar logical form, such as "P & (Q & R)", and vice-versa, from "P" and "Q" one can infer "P & Q". That is to say that all inferential processes that involve the conjunction of symbols or symbolic structures (i.e., conjunction elimination/introduction) are carried out by means of the same computational processes that equate, in the case of conjunction, with joining or separating representations in one's mind. For instance, given two representations, say, "CAT" and "MOUSE", one can associate them ("CAT & MOUSE"), and given the representation "CAT & MOUSE", one can dissociate its constituents. The association and dissociation of mental representations are thus the psychological processes that correspond to the inferential processes in which conjunction is present. Put differently, to logically homogeneous classes of inferences homogeneous psychological, computational, mechanisms or processes must correspond. If one can separate "CAT & MOUSE" into "CAT" and "MOUSE", then by means of the same processes one can separate "(CAT & MOUSE) & DOG" into, say, "CAT & DOG" and "MOUSE", or into "CAT", "DOG" and "MOUSE", and one can carry out the same process with any representations whatsoever.

The above can be summarized in Fig. 3, in which it is shown how symbol rules operate on symbols or symbol structures, and these call forth or motivate symbol rules. The aspects above of productivity, systematicity, compositionality, and inferential coherence arise from the right interaction or combination of symbol rules and symbol structures, and this interaction or combination comes in degrees of complexity, from lower-level, perceptive, representational processes involving the formation and categorization of representations, to higher-level, reasoning representational processes involving the association and further combination of representations.

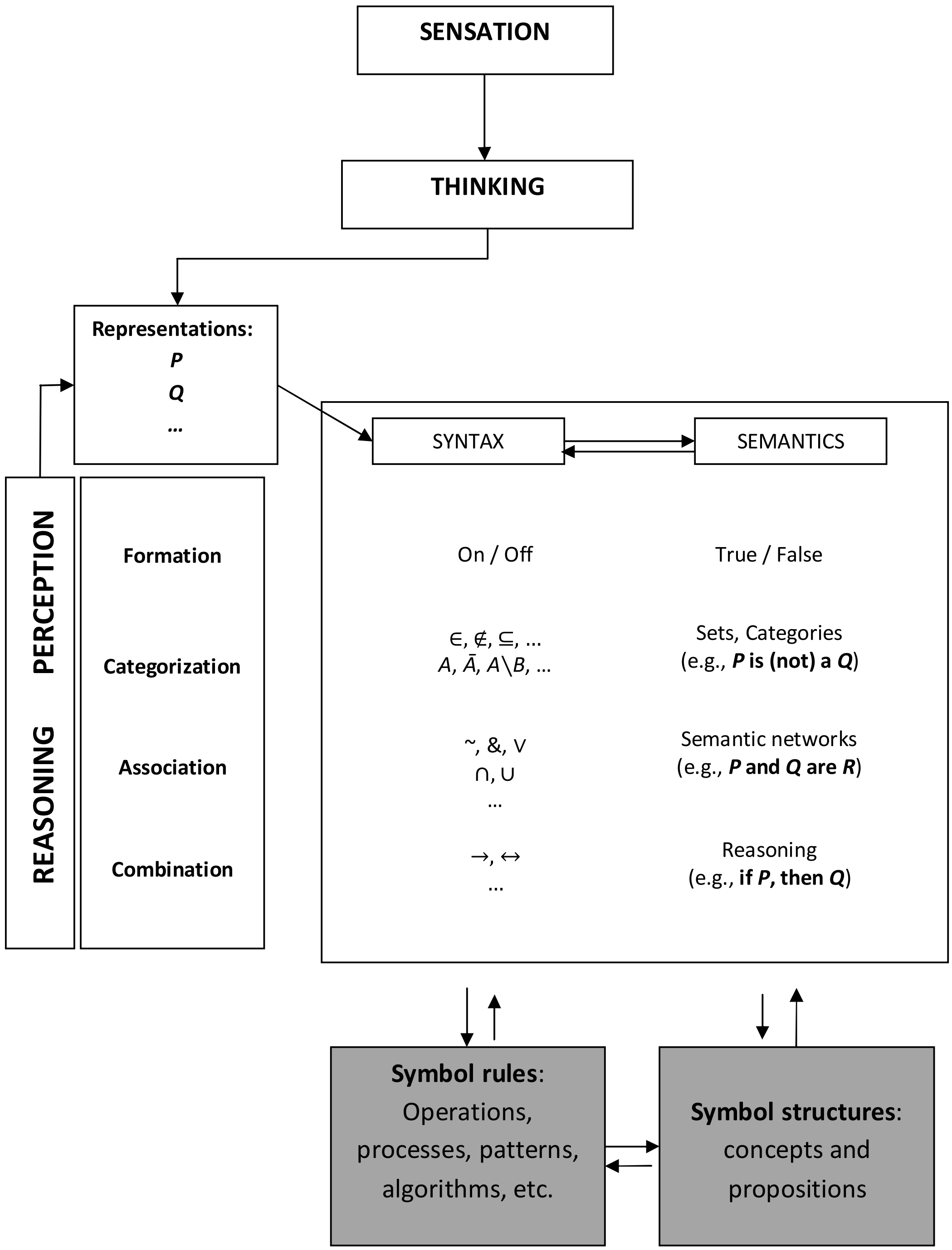

Figure 3. The syntax and the semantics of thought.

Fodor and Pylyshyn see it as possible (though unlikely) that one may acquire the rule "*P* & *Q*" holistically or as purely atomic, i.e., indecomposable into its structural constituents, but this seems to fit a connectionist approach rather than a cognitivist one (see Fodor & Pylyshyn, 1988, for this and further differences between cognitivism and connectionism). Roberts and MacLeod (1995), adopting a cognitivist stance, nevertheless elaborate on an 'atomic representation hypothesis' (ARH), seeing in it a powerful tool to explain many results in studies on non-strategic, unconscious learning involving rules or sequences in paradigms such as artificial grammars and simulated complex systems (cf. Dienes & Berry, 1997). However, as they put it, this is indeed a "curious property" (Roberts & MacLeod, 1995, p. 300), and the question is to what extent, if any, such a property would be evolutionarily and cognitively accountable for. The way the ARH is formulated, it is difficult to see how unconscious cognition contributes to conscious mentation and behavior. In fact, the ARH claims that the cognitive architecture of unconscious information processing is not compatible with that of conscious processing in that there is a structural gap between both that actually entails an ontological segregation: while unconscious representations are hypothesized to be indecomposable albeit compositional, which entails non-systematicity, conscious representations are said to be both compositional and decomposable, and therefore systematic in nature. This (systematicity) is, as seen, in turn connected to other features such as generativity and productivity, which might be connected to modularity, all central postulations of standard cognitivist models (see Fodor & Pylyshyn, 1988; see also Hanna, 2006, Chapter 4). The ARH likewise collides with these. In this scenario, the results Roberts and MacLeod report are simply mysterious, as they presuppose an obscure architectural transition or metamorphosis between the unconscious and the conscious mind*s*.

However, I take it that for talk of "mind" to make scientific sense, there is one single mind for each individual; this is presupposed when one speaks of an individual's mental life (vs. the mind as a general cognitive architecture). To avoid a 'dual mind' or 'hybrid architecture,' there is actually nothing that hinders us from hypothesizing that a complex representation like "*P* & *Q*" can be (indeed, must be) represented as decomposable *also* at an unconscious level, at which, however, it resists decomposition as it commonly takes place in conscious cognitive processing.

In this paper, this resistance to or, on the contrary, the favoring of, conscious processing are hypothesized to reside in the computational correlates (vs. the processes involved and/or the representations themselves), which, in turn, must per force have to do with the symbols processed. But the only *partial* difference here postulated in the realm of mental symbols is one between symbol rules and symbol structures. Let me begin by establishing a purely ad-hoc graphical strategy: symbol rules shall be represented between ⟦ ·⟧, while symbol structures shall be so between ⟪· ⟫. It can further be hypothesized that while a representation "⟦*P* & *Q*⟧" is in principle (but not necessarily) indecomposable at the conscious level, contrary to the *same* representation, "⟪*P* & *Q*⟫", it is not so at the unconscious level (see Fig. 4).

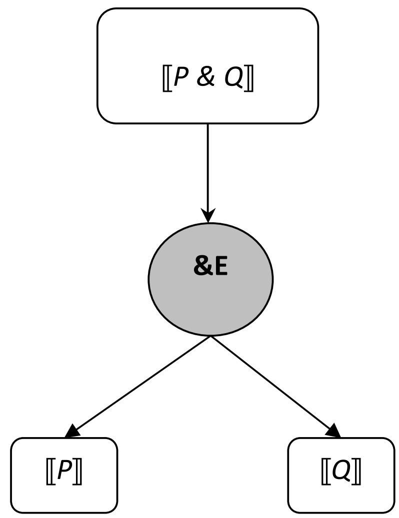

Figure 4. Graphical strategy to indicate an unconscious representation and its processing.

Furthermore, "⟪*P* & *Q*⟫", though in principle capable of being consciously processed, need not be so, in accord with the fact that concepts (i.e., symbol structures) can be processed in an unconscious way (see Part 1, Augusto, 2013). See Fig. 5 for a summary of the above.

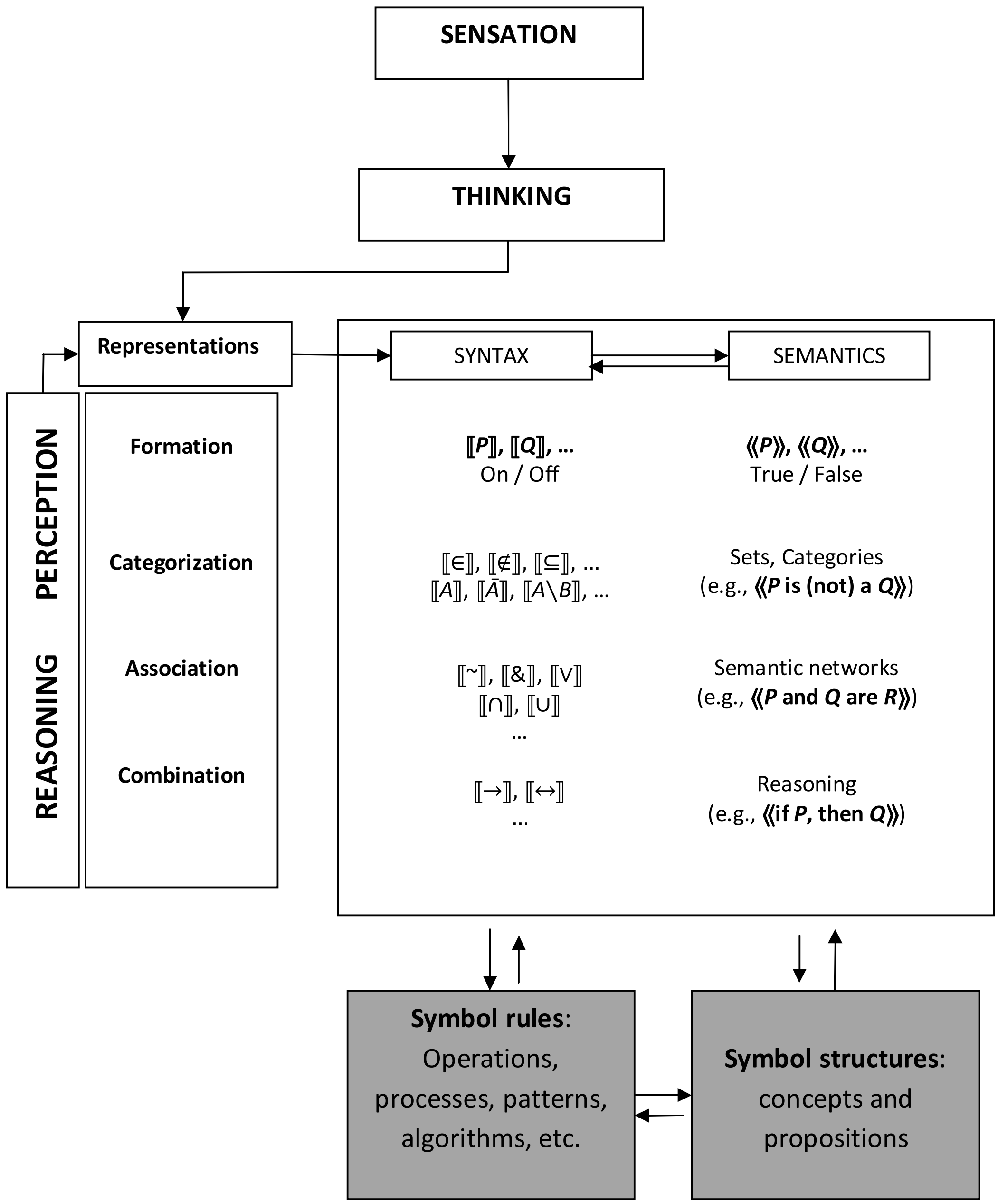

Figure 5. The human information processing system. (Note the graphical strategy to indicate primarily unconscious and possible conscious processing modes, ⟦P & Q⟧ and ⟪P & Q⟫, respectively; compare with Fig. 3.)

In other words, it can be postulated that while unconscious mentation preserves the overall structure of representations based on symbolic constituency *syntactically*, it does not do so *semantically*, namely to the same degree that it would do so if consciousness were to participate in the information processing, without for that requiring different representations, i.e., mental representations that are ontologically and overall structurally distinct. This is a reasonable explanation if one agrees that syntax has a semantic face, and semantics has a syntactic face, as proposed above. In this view, there is no risk of hypothesizing either purely syntactic representations/processes or purely semantic representations/processes, nor is there any risk of postulating exclusively unconscious (or procedural) or exclusively conscious (or declarative) representations/processes, though one can (indeed, must) see some representations/processes as *primarily* syntactic, and therefore more procedural or implementable than declarative, or *sufficiently* semantic, and thus more declarative.[8]

This contradicts both vehicle and process theories: here, neither the representations determine the kind of processing, nor vice-versa. A primarily syntactic symbol structure is more procedural or implementable, rather than declarative, but this does not entail that it has to be processed exclusively at an unconscious level; nor does a sufficiently semantic structure necessitate conscious processing. This explains why, for instance, in a dichotic listening task, highly meaningful stimuli (e.g., the subject's name) presented in the unattended channel can cause the processing of auditory material to become conscious — although not necessarily (see, e.g., Moray, 1959; Wood &

[8] To make a non-reductive use of the influential dissociation between declarative and non-declarative/procedural kinds of knowledge and memory, which appears to be both functionally and anatomically plausible (see Augusto, 2010) (but see Part 1, Augusto, 2013).

Cowan, 1995). It can be hypothesized that the subjects' name immersed in an auditory message that is unattended constitutes a symbol structure that is sufficiently semantic for conscious processing once attention has been diverted to it. Because attention can be diverted to virtually anything, this — instead of, or in addition to, invoking attention solely — seems a plausible explanation for this and similar phenomena. In fact, material that is sufficiently semantically structured can be processed in a wholly unconscious way, as the application of indirect tests often show (e.g., Eich, 1984; see also Szymanski & MacLeod, 1996); in this case, it can be hypothesized that, without attention, the structure of the material tested is processed in a primarily syntactic mode, which can be — not entirely correctly, or even wrongly — interpreted as a non-analytic, or holistic, mode (e.g., Anooshian, 1989; Roberts & MacLeod, 1995).

One thus avoids the ontological problems discussed above brought about by an opposition between vehicle and process theories, as well as the structural divide postulated by the ARH, which all entail that the cognitive architecture of the conscious and unconscious mind*s* must differ in such significant ways so as to render their coexistence mysterious. Further, the evolutionary continuity between the unconscious and the conscious 'modes' of information processing is preserved: if unconscious mental processing has evolutionary significance in face of conscious mentation (evolutionary principles of success and stability; see Reber, 1992a, p. 38), then the latter has to build upon the former (principle of conservation; cf. *ibid.*) in ways that preserve ontological and structural continuity. Regarding this latter aspect, the principle of commonality (cf. *ibid.*) can here be applied with some adjustments; in fact, this ontological and structural continuity is not only what allows us to (re)interpret information processed in an unconscious way, but it permits us to learn about ourselves by observing the cognitive behavior of other animals, namely of other mammals. In this case, a plausible hypothesis is that while humans — and perhaps (some) other primates and also (some) cetacea — can represent information in a fully semantic, conceptual way, many other animals, especially mammals and also birds, possessing a neocortex or an homologous structure (e.g., the avian Wulst), share (parts of) the same syntax of thought with us, even if perhaps a relatively simpler one, which might account for the hypothesized 'poorer' semantics. Behaviorism, not so much wholly dismissed but rather largely extended by cognitivism, implicitly relied on this assumption.

4. A 'reality test'

Summing up: Let one entertain a mental representation, say, "*P* & *Q*". I argue that this representation can be primarily syntactic — in which case it is graphically represented as "⟦*P* & *Q*⟧" —, or sufficiently semantic (graphically: "⟪*P* & *Q*⟫"), without for that being a different representation either in ontology or structure. However, I hypothesize that the former is not easily, or at all, accessible to conscious mentation, while the latter can more easily, or even entirely, be so. I base this hypothesis on the theory that there are two non-dissociable faces to human information processing, or cognition, to wit, syntax and semantics, and that these, though the two sides or faces of the same cognitive phenomenon, nevertheless entail diverse computational correlates of consciousness, with syntax being constituted by symbol rules, essentially procedural or implementable (on symbol structures), and semantics constituted by symbol structures, more declarative-like or 'declarable,' but nevertheless requiring the application of symbol rules (see above; see also Fig. 5).

I support these conjectures on how well they fare in the face of (i.e., can be applied to) experimental evidence. Let us then revisit the four cases above (for a discussion of each Case, see Part 1, Augusto, 2013).

4.1. Case I. Higher-level cognition in the absence of consciousness and concepts: Infant cognition

Infants as young as 3 months old appear to be capable of higher-level cognition, exhibiting goal-controlled/directed action; they show pleasure when their expectations or anticipations are met and displeasure when the contrary is the case (e.g., Papoušek, 1967). While this would be difficult to account for were one to insist on a full-fledged semantic processing by the infants (i.e., processing with full-fledged concepts), because, among other reasons (see Part 1, Augusto, 2013), they have no observable developed verbal skills involving concepts and semantic networks that require long-term memory (see, e.g., Collins & Loftus, 1975; Rosch, 1978), the hypothesis of a first and foremost syntactic processing does not appear unwarranted. If Hanna (2006) is right, humans are endowed with an innate logical cognitive faculty, a sort of proto-logic. It can be speculated that while this is not identical with the logic(s) logicians have acquainted us with, it may very well be a primitive or foundational scaffolding upon which the syntax and semantics of logic proper gradually develop and mature into adult thought (cf., for instance, Inhelder & Piaget, 1958; Piaget, 1953). One can further speculate that Fodor's (1975) 'language of thought' and this proto-logic are essentially the same.[9]

---

[9] Although Fodor's 'language of thought' appears to be (implausibly) richer in semantic terms (see Fodor, 1975).

Such a proto-logical scaffolding, massively syntactic in comparison to the incipient semantics, may be speculated to contain, along with the sensorimotor commands that young infants use in their first explorations of the environment and of the self, incipient symbol structures providing them with a basic representational, largely or even entirely proto-conceptual (rather than non-conceptual), symbolic system. In fact, on the one hand, it seems unwarranted to postulate a semantically Lockean-like blank slate upon which concepts would be ‘spontaneously’ originated, as, for instance, Thelen and Smith (1994) do. It appears evolutionarily and developmentally more reasonable to hypothesize that innate proto-conceptual structures must be present at birth. This, also against those who propose innate primitive or core adult-like concepts (e.g., Spelke, 1994; Spelke et al., 1992), because, on the other hand, concepts are both historical and cultural, i.e., they must be acquired in time and in context, and are subject to radical alterations (Quine, 1960; Rosch, 1978).

In the view defended here, an innate (proto-) conceptual structure is hypothesized to be provided to a great extent by the syntactic structure that supports all human thought as the processing or computation of representations. If one sees this — i.e., both syntax and semantics — as symbolic in essence, then the problem becomes a more clear-cut one: how “$[\![P \,\&\, Q]\!]$” and “$\langle\!\langle P \,\&\, Q \rangle\!\rangle$” are the *same* representation.

We do not know the answer — actually, we have not yet started looking for it in these terms —, but it can be hypothesized that, endowed with such an overwhelmingly syntactic symbolic system, a young infant can represent the conjunction of a (planned) action with a (predicted or expected) reward from the environment as “$[\![P \,\&\, Q]\!]$”, where, say, “$[\![P]\!]$” stands for the action, and “$[\![Q]\!]$” for the reward. In this way, the infant can conjoin “$[\![P]\!]$” and “$[\![Q]\!]$” as “$[\![P \,\&\, Q]\!]$” (inference rule of conjunction introduction) in the absence of full awareness of what the thought process entails in terms of meaning. This can, in fact, further explain why this kind of information processing can still be available as the individual develops into adulthood and into old age: it simply is the basic representational faculty at the basis of all further cognitive developments, very likely the last mental competence or faculty to be suspended (e.g., in coma) or turned off (with the death of the individual). It probably requires the minimal viable amount of brain (roughly, the brainstem and immediately adjacent subcortical areas, i.e., the upper brainstem or the midbrain, which adjoins rostrally the diencephalon; see Part 1, Augusto, 2013, Fig. 1) and because it allows for surprisingly complex emotional behavior, probably because it spares central organs or areas of the limbic system, it may be — very likely wrongly — connected to conscious processing (e.g., Merker, 2007).[10]

Indeed, it appears more realistic to hypothesize that these are largely unconscious representations whose processing is made possible by the fact that a syntax or innate proto-logical competence, intimately connected with basic sensorimotor competences, can be supported by this 'minimal brain.' This — and setting up the bridge between computational and neuronal correlates of consciousness — is supported by theories that 'put' consciousness in the cortex, namely in the mammalian neocortex (e.g., Eccles, 1992; Koch, 2004), as well as by empirical evidence indicating that implicit or unconscious forms of learning and memory depend heavily on subcortical structures in the midbrain: for example, Reiss and colleagues (Reiss et al., 2005) have found a significantly higher activation of the ventral striatum in tasks of implicit as compared to tasks of non-implicit learning (see also Uddén et al., 2010, for the

[10] Note the central importance of this 'minimal' brain for creature consciousness, or wakefulness (see, e.g., Schiff, 2007). See the discussion motivated by Merker (2007) in Part 1 (Augusto, 2013).

implication of the striatum in implicit learning), and the striatum, as well as the basal ganglia, of which the former is a part, has been convincingly linked to procedural forms of implicit memory (e.g., Packard & Knowlton, 2002; Knowlton et al., 1996). Relevant is also the existence of a subcortical, colliculus-pulvinar visual pathway to both the amygdala and the dorsal visual stream that appears to explain the anomalous phenomena of (affective) blindsight, prosopagnosia, and left visuo-spatial neglect (see Augusto, 2010, for the joint discussion of these three conditions in light of the dual visual stream hypothesis).

4.2. Case II. Conceptual representations processed unconsciously in higher-level cognition: The Perruchet Effect

Eye blink conditioning in humans appears to implicate conceptual representations, but the subjects process them wholly unconsciously, contrary to what is believed according to the TMHC. This contradiction is provided by the experimental results reported by Perruchet (1985): conditioned in a 50% partial reinforcement schedule by pairings of puffs of nitrogen (US) and 70dB, 1 sec tones (CS), the subjects were reported to blink despite consciously not expecting the US; they appeared to have unconsciously acquired an accurate long-term memory of the timing of the presentation of the CS.

It is reasonable to hypothesize that propositional representations (see below) of the kind "$P \rightarrow Q$" are implicated, expressing representations like "if CS is due to be presented now, then do CR" (e.g., "if there is a tone, then blink"), or "$P \rightarrow (R \;\&\; Q)$" ("if US is present, then expect CS and do CR"). These appear to be processed without a conscious representation being attained. Using the ad-hoc symbolism above, this processing can be expressed as "$[\![P \rightarrow Q]\!]$" and "$[\![P \rightarrow (R \;\&\; Q)]\!]$". The relevant aspect suggested by Perruchet's (1985) results is that there seems to be a dual, dissociable processing being carried out, one conscious in which the subject represents to her-/himself something of the kind "if there has been repeated puffs for a while, expect them to be absent now (i.e., don't blink)" (the gambler effect), a representation of the kind "⟪$P \rightarrow \sim Q$⟫" (where "~" is the symbol for negation), that is nevertheless wholly disregarded by the unconscious processing taking place that appears to allow the learning and memorizing of the accurate length and nature of the trials, too complex to be learned and memorized consciously. To sum up: while consciously not expecting the US, the subject does unconsciously expect the CS and behaves accordingly.

The overruling of the conscious processing by the unconscious representations — here, the classical conditioning effect — might be explained by an evolutionary supremacy and a developmental primacy of unconscious over conscious processing in terms of action.[11] However, in spite of the first-and-foremost rule-based, syntactic processing closely connected to sensorimotor processes hypothesized to be involved in this case, it seems that a conceptual level, likely supportive of a propositional level, must be attained, too, as it would otherwise be difficult to explain the eye blink conditioning, supposed to be supported by the unconscious conceptual/propositional representation of the expectation of the CS, that is, the tone (see Delamater, 2012; Mitchell et al., 2009).[12] Agreeing with the characteristics above for representations that

[11] These are supported by the implication of the brainstem and cerebellum, structures believed to be more primitive in the evolution of the brain, in delay eye blink conditioning (see Part 1, Augusto, 2013). Again, subcortical structures are strongly implicated; as seen above, in the perspective here adopted, it may be speculated that these can mostly support a first-and-foremost syntactical processing of representations.

[12] See Part 1 (Augusto, 2013), footnote 11 for a note concerning conceptual representations supported by subcortical brain areas or structures. Incidentally, and interestingly, Delamater (2012) is a connectionist approach.

define a Classical cognitive architecture, the internal structure of the representations conditions the interpretation carried out (in this case, "$P \rightarrow Q$", and not, say, "$Q \rightarrow P$"), which entails that the syntactic process constrains the semantic interpretation, and vice-versa.

4.3. Case III. Higher-level cognition without consciousness and concepts: Blindsight

Patients with blindsight seem to be capable of higher-level cognition based on visual stimuli that they claim not to perceive, indicating, in a different way from that of infant cognition (Case I above) but nevertheless as suggestively, that neither conceptual representations nor consciousness are required for high-level cognitive processes. A particularly interesting study (de Gelder et al., 2008) revealed that a patient with no cortical visual abilities can successfully navigate obstacles in the workplace. It is believed that the patient resorted exclusively to visual information provided by processing in subcortical alternative paths (see Part 1, Augusto, 2013).

At the strictly representational level, it may be hypothesized that the patient studied by de Gelder and colleagues represents 'visually' the environment in, say, spatial maps indicating location of self (proprioception) and objects (exteroception) by means of conjunctions (e.g., "there is an object here on my left and another here in front of me and etc.", expressible as "P & Q & R & …"). Such spatial maps might provide information not only on the location, but also on the size and kinetic state of objects. All these may allow state and/or action-directed propositional representations of the kind "$P \rightarrow Q$" (e.g., "if there is a large static object on the left, move to the right"), "$R \leftrightarrow S$" (e.g., "move forward if and only if there is no obstacle in front"), or "$P \vee Q$" ("move forward or turn left"), etc., without, however, allowing the subject to become aware of why s/he acts the way s/he does.[13] In the ad-hoc symbolism here proposed, the representations in play are of the sort "$[\![P \rightarrow Q]\!]$", "$[\![R \leftrightarrow S]\!]$", etc.

Of import is the fact that these propositional representations must be overwhelmingly syntactic, given that conceptual representations appear to be safely ruled out. Indeed, the patients claim either not to perceive the visual stimuli (blindsight Type 1) or to perceive them in modalities other than the visual one (blindsight Type 2). Nevertheless, low-spatial-frequency representations of the objects respecting their location, size, kinetic state, etc., must be possible in face of their above-chance performance in visual perception tasks. Afforded by the visual pathways indicated above, without being properly conceptual they must be in some way semantic, providing the patient with a sort of 'action semantics,' as opposed to an 'object semantics' supporting object identification and storage in semantic memory (see Hodges et al., 1999).

The representational processing hypothesized to be carried out in blindsight possesses all the features above discriminated as belonging to a Classical cognitive architecture, all based on constituency. For instance, the patient analyses a representation of the form "$P \rightarrow Q$" uniquely, i.e., as "$P \rightarrow Q$" (e.g., "if there is a large static object in front, then move to the right") and not as "$Q \rightarrow P$" (e.g., "if move to the

---

[13] Such representations affording an also (proto-)conceptual level, in this and the other cases analyzed, would in principle require, for their formal expression, predicate rather than only propositional logic, given that judgments of existence ("There is an x and it is such and such", $P(x)$, or x is P) appear to be involved (e.g., "there is an object and it is on the right" is expressible formally as "$\exists x(Ox \; \& \; Rx)$"), but we may in fact see the latter logic as primitive in relation to the former: a well-formed formula such as "$P(x) \; \& \; Q(x)$" is clearly secondary in relation to "$P \; \& \; Q$". In other words, predicate, or first-order logic extends propositional logic.

right, then there is a large static object in front"), because this interpretation is dependent on the internal structure of "$P \rightarrow Q$", which, in turn, is conditioned or constrained by the representations "P", "Q", and "→".

4.4. Case IV. Conceptual representations processed unconsciously in higher-level cognition: The halo effect

We tend to associate prominent properties or qualities perceived in self or other individuals with properties that have objectively nothing to do with them; for instance, attractive individuals are typically attributed higher ratings in career success and personality (see, e.g., Dion et al., 1972). While conceptual representations are implicated, these associations are carried out largely or wholly unconsciously, creating a puzzle for the TMCH.

The tell-tale signs that these operations are carried out unconsciously are the spontaneity and automaticity, with extremely short reaction times (for instance, 100ms or less is the time it takes for subjects to make value judgments — form an impression — from unfamiliar, neutral, faces; see, e.g., Willis & Todorov, 2006), in contrast with the subjects' inability to say what rules they draw on in order both to make initial judgments regarding a trait of a face (e.g., attractiveness) and to make — often reliable — inferences from that trait. The high complexity of such a task is obvious in face perception: there are several organs involved and many factors are analyzed with respect to each, individually (e.g., size, shape, color, etc.) and in relation to the others (e.g., distance, match, symmetry, etc.). Again, subcortical structures are believed to be strongly implicated: for instance, the striatum has been verified to be activated when a neutral stimulus was paired with an attractive face (see Bray & O'Doherty, 2007), and the amygdala is 'automatically' activated by a face judged as untrustworthy (e.g., Engell et al., 2007). All this supports a processing of information massively syntactic, given that subcortical structures appear to be more directly implicated in unconscious processing (see above).

This apparently simple phenomenon, known as the halo effect, might actually take up all our syntactic resources. In fact, other people are categorized as possessing certain properties, which, in turn, entail associated categorizations, for which effect we carry out operations having to do with sets and membership in sets (e.g., x is an A, formally: $x \in A$; all As are Cs, formally expressible as $A \cup C = C$ and $A \cap C = A$; x is not a B, formally: $x \notin B$; no Bs are Ds, formally $B \cap D = \emptyset$; etc.), describable by set theory. The same operations can be described by logic: for instance, if someone is categorized as having property A, and there is an unconscious rule or pattern (an attitude or a stereotype, in social psychology jargon) according to which all As are Cs, then this equates with a representation of the kind "$[\![A \rightarrow C]\!]$". But the halo effect can occur in the two senses, that is, from A to C ("$[\![A \rightarrow C]\!]$") and from C to A ("$[\![C \rightarrow A]\!]$"), reason why it might be best expressed by the operation of material equivalence, i.e., "$[\![A \leftrightarrow C]\!]$", at least in some instances.[14]

5. Conclusions

Are unconscious mental representations different from conscious ones? Indeed they are, and in the most obvious, trivial sense: we are not aware of the former, we are so of the

[14] For instance, there seems to be a bidirectional intelligence ↔ attractiveness halo effect for some people/communities/age groups, which might be up to a certain point biologically and/or evolutionarily, if not purely culturally, accountable for, as Zebrowitz and colleagues discuss (Zebrowitz et al., 2002).

latter. In a less obvious or trivial sense, they are phenomenally, qualitatively different in that the one processing is reportable (the subject can talk about her/his experiences or refer directly to them in some other way), while the other is not directly so, though it can be 'reported' in indirect ways (e.g., covert behavior, revealed by, for instance, electroencephalogram readings and skin conductance responses). In this paper, I defend the view that this phenomenal, qualitative difference rests on the prevalence of the side or aspect of a cognitive process being carried out, syntactic or semantic, upon what must be essentially the same representations in structure and in nature.

Mental representations *are* already, and are further carried out *upon*, something: *symbols*, i.e., data as mediated by sensation, perception, and thinking in general. In order to be manipulated, symbols require rules and interpretations, namely rules *for* interpretations. This is so at least in the case of humans. So, there is, in the human case, no syntax without semantics, as little or incomplete as the semantic expression (vs. content) might be in face of the more pressing procedural requirements that have to do with, for instance, time constraints. This fact, i.e., that both syntax and semantics are symbolic in essence, sharing many, most, or even all the symbolic features that define human cognition as information processing and the mind/brain as a symbol processing system, entails a continuity between conscious and unconscious cognition.

By positing this continuity between unconscious and conscious representations, the former being evolutionarily and developmentally earlier in relation to the latter, the structural and ontological unities of the human cognitive architecture are safeguarded. This is formally expressible by applying the ad-hoc symbolism that consists in distinguishing, say, "⟦*P* & *Q*⟧" from "⟪*P* & *Q*⟫", where the first is supposed to be computed at a more syntactic, procedural level, while the second attains a more fully conceptual expression, i.e., not only can one consciously represent a symbol such as ⟪a cat and her kittens⟫, but one can also represent oneself as representing ⟪I am seeing/imagining/… a cat and her kittens⟫. But this is the only structural difference between these unconscious (the first) and conscious (the second) forms of processing: in terms of internal structure, there is no difference, as, for instance, both "⟦*P* & *Q*⟧" and "⟪*P* & *Q*⟫" are to be analyzed into "*P*" and "*Q*" separately (i.e., into "⟦*P*⟧" and "⟦*Q*⟧" in the first, into "⟪*P*⟫" and "⟪*Q*⟫" in the second 'mode' of processing).

By making the unconscious processing of information primary — indeed foundational — with relation to consciousness, one is establishing that the study of the latter should start in that of the former. Because unconscious cognition is primarily syntactic, this equates with studying how first and foremost syntactic processing of information (from sensorimotor commands to complex structures representations) allows for conscious meaning extraction and interpretation. That Searle's syntax (see above) cannot originate semantic expression is simply because it is not the *right* (human-like) syntax. As I see it, *the hard problem of consciousness* lies in explaining how a largely unconscious syntax of thought, implementable in a physical, biological mechanism or machine (the brain), computationally supports a semantic expression that is adequate for consciousness, without the need to segregate between both, syntax and semantics.

Acknowledgments: I wish to thank the Portuguese Foundation for Science and Technology (FCT) and the European Social Fund for the post-doctoral fellowship that funded, among others, this work.

References

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355-365.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *25*, 313-341.

Anooshian, L. J. (1989). Effects of attentive encoding on analytic and nonanalytic processing in implicit and explicit retrieval tasks. *Bulletin of the Psychonomic Society*, *27*, 5-8.

Augusto, L. M. (2010). Unconscious knowledge: A survey. *Advances in Cognitive Psychology*, *6*, 116-141.

Augusto, L. M. (2012). Lost in dissociation: Behavioral-phenomenological paradigms in unconscious cognition. [forthcoming]

Augusto, L. M. (2013). Unconscious representations 1: Belying the traditional model of human cognition. *Axiomathes*. DOI: 10.1007/s10516-012-9206-z. [in press]

Bray, S. & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology*, *97*, 3036-3045.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language*, *35*, 26-58.

Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.

de Gelder, B., Tamietto, M., van Boxtel, G., Goebel, R., Sahraie, A., van den Stock, J., et al. (2008). Intact navigation skills after bilateral loss of striate cortex. *Current Biology*, *18*, 28-29.

Delamater, A. R. (2012). On the nature of CS and US representations in Pavlovian learning. *Learning & Behavior*, *40*, 1-23.

Dienes, Z. & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*, 3-23.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.

Eccles, E. C. (1992). Evolution of consciousness. *Proceedings of the National Academy of Sciences*, *89*, 7320-7324.

Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory & Cognition*, *12*, 105-111.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*, 1508-1519.

Fodor, J. A. (1975). *The Language of Thought*. Sussex: Harvester Press.

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

Hanna, R. (2006). *Rationality and Logic*. Cambridge, Massachusetts: MIT Press.

Haugeland, J. (1981). The nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, *1*, 215-226.

Hodges, J. R., Spatt, J., & Patterson, K. (1999). “What” and “how”: Evidence for the dissociation of object knowledge and mechanical problem-solving skills in the human brain. *Proceedings of the National Academy of Sciences U.S.A.*, *96*, 9444-9448.

Inhelder, B. & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts & Company.

Knowlton, B. J., Mangles, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399-1402.

Laird, J. E. (2012). *The Soar Cognitive Architecture.* Cambridge, MA: MIT Press.

McKenna, P., & Oh, T. (2005). *Schizophrenic Speech. Making Sense of Bathroots and Ponds that Fall in Doorways.* Cambridge: Cambridge University Press.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for science and medicine. *Behavioral and Brain Sciences*, *30*, 63-134.

Mitchell, C. J., De Houwer, J., Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-246.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56-60.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and Reality*. San Francisco: W. H. Freeman.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135-183.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Packard, M. G. & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, *25*, 563-593.

Papoušek, H. (1967). Experimental studies of appetitional behavior in human newborns and infants. In H. W. Stevenson, E. H. Hess, & H. L. Rheingold (eds.), *Early Behavior: Comparative and Developmental Approaches* (pp. 249-277). New York: John Wiley.

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Pavlovian Journal of Biological Science*, *20*, 163-170.

Piaget, J. (1953). *Logic and Psychology*. Manchester: Manchester University Press.

Prawitz, D. (1965/2006). *Natural Deduction: A Proof-Theoretical Study*. Mineola, NY: Dover.

Preston, J. & Bishop, M. (2002). *Views into the Chinese Room. New Essays on Searle and Artificial Intelligence*. Oxford: Oxford University Press.

Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Reber, A. S. (1992a). An evolutionary context for the cognitive unconscious. *Philosophical Psychology*, *5*, 33-38.

Reber, A. S. (1992b). The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition*, *1*, 93-133.

Reiss, J. P., Campbell, D. W., Leslie, W. D., Paulus, M. P., Stroman, P. W., Polimeni, J. O., Malcolmson, K. A., & Sareen, J. (2005). The role of the striatum in implicit learning: a functional magnetic resonance imaging study. *NeuroReport*, *16*, 1291-1295.

Roberts, P. L., & MacLeod, C. (1995). Representational consequences of two modes of learning. *Quarterly Journal of Experimental Psychology, Section A*, *48*, 296-319.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Burns (eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum.

Schiff, N. D. (2007). Global disorders of consciousness. In M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness* (pp. 589-604). Oxford: Blackwell.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417-457.

Searle, J. (1990). Is the brain's mind a computer program? *Scientific American*, *262*, 26-31.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (eds.), *Handbook of Cognition* (pp. 202-220). London: Sage.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379-423, 623-656.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*, 431-445.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605-632.

Sun, R. (2002). *Duality of the Mind: A Bottom-up Approach toward Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, *17,* 341-373.

Szymanski, F. K. & MacLeod, C. M. (1996). Manipulation of attention at study affects an explicit but not an implicit test of memory. *Consciousness & Cognition*, *5*, 165-175.

Thelen, E. & Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge MA: MIT Press/Bradford.

Uddén, J., Folia, V., & Petersson, K. M. (2010). The neuropharmacology of implicit learning. *Current Neuropharmacology*, *8*, 367-381.

Weaver, W. (1949). The mathematics of communication. *Scientific American*, *181*, 11-15.

Willis, J. & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592-598.

Wood, N. & Cowan, N. (1995). The cocktail party phenomenon revisited: How frequent are attention shifts to one's name in an irrelevant auditory channel. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 255-260.

Zebrowitz, L. A., Hall, J. A., Murphy, N. A., & Rhodes, G. (2002). Looking smart and looking good: Facial cues to intelligence and their origin. *Personality and Social Psychology Bulletin*, *28*, 238-249.