ELSEVIER

Journal of Cognitive Systems Research 1 (2001) 241–249

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Computation, reduction, and teleology of consciousness

Action editor: Vasant Honavar

Ron Sun*

CECS Department, University of Missouri at Columbia, 201 Engineering Building West, Columbia, MO 65211, USA

Received 9 October 2000; accepted 9 October 2000

Abstract

This paper aims to explore mechanistic and teleological explanations of consciousness. In terms of mechanistic explanations, it critiques various existing views, especially those embodied by existing computational cognitive models. In this regard, the paper argues in favor of the explanation based on the distinction between localist (symbolic) representation and distributed representation (as formulated in the connectionist literature), which reduces the phenomenological difference to a mechanistic difference. Furthermore, to establish a teleological explanation of consciousness, the paper discusses the issue of the functional role of consciousness on the basis of the aforementioned mechanistic explanation. A proposal based on synergistic interaction between the conscious and the unconscious is advanced that encompasses various existing views concerning the functional role of consciousness. This two-step deepening explanation has some empirical support, in the form of a cognitive model and various cognitive data that it captures.

Keywords: Consciousness; Cognition; Qualia; Implicit learning; Computation; Reduction; Teleology

The importance of consciousness to cognitive science cannot be over-estimated. Studying only brain neurobiology or only computational models (e.g., neural networks or production rules), or even both, will not resolve the fundamental question of human cognition — human consciousness (Lloyd, 1995). Directly confronting this issue, in this paper I will argue for mechanistic and, furthermore, teleological explanations of consciousness, in contrast to many existing theories that focus on what appears to be tangential aspects of consciousness.

1. Mechanistic views of consciousness

Can we explain consciousness in mechanistic terms? In science, we generally assume the sufficiency and necessity of mechanistic explanations (if we are not going to be dualists). By mechanistic explanation, it is meant any concrete physical processes, that is computational processes in the broadest sense of the term ‘computational’. In general,

*Tel.: + 1-573-884-7662; fax: + 1-573-884-8318.
E-mail address: rsun@cecs.missouri.edu (R. Sun).

1389-0417/01/$ – see front matter
PII: S1389-0417(00)00013-9

'computation' is a broad term that can be used to denote any process that can be realized computationally, ranging from chaotic dynamics (Freeman, 1995) and 'Darwinian' competition (Edelman, 1989), to quantum mechanics (Penrose, 1994). In terms of the sufficiency of mechanistic explanations, Jackendoff (1987) suggested the following hypothesis: ''Every phenomenological distinction is caused by/supported by/projected from a corresponding computational distinction.'' Due to the lack of a clearly better alternative, this hypothesis remains a viable working hypothesis, despite various criticisms of it. These criticisms (e.g., Edelman, 1989; Freeman, 1995; Damasio, 1994; Penrose, 1994; Searle, 1980) failed to show that computation, in general, cannot account for the nature of consciousness, although they had some legitimate complains about specific computational approaches and models. We can use the term 'mechanism' (or 'mechanistic processes') interchangeably with the term 'computation', avoiding some of the negative connotations associated with the latter. In addition to the sufficiency, the necessity of mechanistic explanations is also self-evident: it is obvious to anyone who is not a dualist that the foregoing definition of mechanistic processes has to include the necessary condition for consciousness, for the physical basis of mental activities and phenomenal experience cannot be anything else but such mechanistic processes.

An explanation of the mechanistic basis of consciousness and its mechanistic roles (or 'teleology') in the human mind is needed: what kind of mechanism leads to the conscious, and what kind of mechanism leads to the unconscious? What is the functional role of the conscious? What is the functional role of the unconscious? Such issues are highly relevant, because they provide necessary theoretical frameworks for further empirical (scientific) work.

The two important premises for the subsequent discussion are: (1) the direct accessibility of conscious processes; and (2) the direct inaccessibility of unconscious processes. Conscious processes should be directly accessible (e.g., directly verbally expressible), without involving intermediate interpretive or transformational steps, which is a requirement prescribed and/or accepted by many others (see, e.g., Clark, 1992; Hadley, 1995)[1]. Unconscious processes should be, in contrast, inaccessible directly, thus exhibiting different properties (see, e.g., Heidegger, 1927; Dreyfus & Dreyfus, 1987; Berry & Broadbent, 1988; Reber, 1989).

There have been a variety of explanations. I will categorize existing mechanistic (computational) explanations of the distinction between the conscious and the unconscious, especially as embodied in existing cognitive models, on the basis of the following types of emphases: (1) differences in knowledge organization (e.g., between two subsystems), (2) differences in knowledge content (between two subsystems), (3) differences in knowledge representation (between two subsystems), (4) differences in knowledge processing mechanisms (between two subsystems), or (5) difference in processing modes (of the same system). Among them, some explanations are based on recognizing that there are two separate subsystems in the mind. The difference between the two subsystems can be explained in terms of differences in either knowledge processing mechanisms, knowledge organization, knowledge content, or knowledge representation. In contrast to these two-system views, there are also views that insist on the unitary nature of the conscious and the unconscious; that is, they hold that the conscious and the unconscious are different manifestations of the same underlying system or process (Dennett, 1991). The difference is thus that of different processing modes in the same system.

For instance, Anderson (1983) posits in his ACT* model that there are two types of knowledge; declarative knowledge is represented by semantic networks, and it is consciously accessible; procedural knowledge is represented by production rules, and it is inaccessible. The difference lies in the different ways of organizing knowledge: whether the knowl-

[1][Accessibility is solely concerned with surface syntactic structures of objects being accessed (at the level of outcomes or processes, to be explained later), not their semantic meanings. Thus, for example, a LISP expression is directly accessible, even though one may not fully understand its meaning. The internal working of a neural network may be inaccessible even though one may know what the network essentially does (through an interpretive process)].

edge is organized in an action-centered way (procedural knowledge) or in an action-independent way (declarative knowledge). Both types of knowledge are implemented symbolically (using either symbolic semantic networks or symbolic production rules).

The model, unfortunately, does not explain the fundamental, qualitative phenomenological differences between the conscious and the unconscious (e.g., in terms of conscious accessibility). Although the knowledge organization is apparently different between semantic networks and production rules (with different degrees of action-centeredness), the difference is insufficient to account for the qualitative phenomenological difference, since both are symbolically represented and fundamentally the same. The difference in conscious accessibility is assumed in the model rather than intrinsic. Thus, there is no theoretical reduction of the phenomenology to any fundamental mechanistic notion (e.g., representation).

Hunt and Lansman's (1986) model is almost the exact opposite of Anderson's model, although the emphasis is on knowledge access (as opposed to knowledge organization). In their model, the 'deliberate' process of production matching and firing, which is serial, is assumed to be a conscious process, while the spreading activation (Collins & Loftus, 1975) in semantic networks, which is massively parallel, is assumed to be an unconscious process. Despite the different emphasis, the problem with this model is the same: the difference between symbolic rule matching and firing and symbolic spreading activation is minor, and insufficient to account for the qualitative phenomenological difference between the conscious and the unconscious. (A number of other views had the same problem, such as Anderson, 1993; Bower, 1996; Logan, 1988; etc.)

There have also been various proposals in neurobiology that there are different processing pathways in the brain, some of which lead to conscious awareness while others do not. For example, Milner and Goodale (1995), Damasio (1994) and LeDoux (1992) proposed various versions of this view. Likewise, Schacter (1990) and Revonsuo (1993) suggested, based on neuropsychological data, that multiple modules co-exist in the brain, each of which performs specialized processing (without incurring conscious awareness), with the exception of one module that is solely responsible for conscious awareness. Each of the specialized modules sends its output to the conscious module and thus makes the output consciously accessible. The problem of these biologically motivated two-system views is that, although there is biological evidence that indicates the existence of multiple pathways (in visual, language, and other processing modes), some of which are correlated with conscious awareness while some others are not, there is no explanation of why some result in consciousness while others do not, that is, what is different, mechanistically, between these different pathways. Mere association with different neural pathways does not constitute an explanation.

Yet another two-system view is the representational difference view. As proposed in Sun (1994, 1995, 1999), different representational forms (in different subsystems) may be used to explain the qualitative phenomenological difference between the conscious and the unconscious. In localist (symbolic) representation, one distinct entity (e.g., a node in a connectionist model) represents a concept. Therefore, the representation is easily accessible. In distributed representation, a non-exclusive set of entities (e.g., a set of nodes in a connectionist network) are used for representing one concept and the representations of different concepts overlap each other; that is, a concept is represented as a pattern of activations over a set of entities (a set of nodes). Therefore, the representation is not easily accessible (relatively speaking). The mechanistic difference in accessibility between the two types of representation accounts for the phenomenological difference in accessibility between the conscious and the unconscious.

Turning to one-system views, as suggested by Baars (1988) and many others, some sort of coherence in the activities of the mind gives rise to consciousness. The emphasis is on internal consistency, which supposedly produces consciousness. The distinction of the conscious and the unconscious is linked to the distinction between coherent and incoherent activities in the mind. For example, Mathis and Mozer (1996) suggested that being in a stable attractor of a dynamic system (a neural network in particular) leads to consciousness. Crick

and Koch (1990), on the other hand, suggested that synchronous firing of neurons (at 35–75 Hz in particular) leads to conscious awareness. There is also the variety that insists on the reverberation of information flows in various cortical and sub-cortical areas that leads to consciousness, as suggested by Damasio (1994). The difficulty with these views is that there is no explanation why coherence (whether it is in the form of attractors, reverberation, or synchronous firing) leads to consciousness; that is, what is qualitatively different about coherence (in any of these above forms) that can account for the qualitative, phenomenological difference in consciousness.

Notably, some of these views are based on biological data (i.e., concerning biological mechanisms, processes or systems; for example, the synchronous firing view, the reverberation view, and the two-pathway view). However, being biological does not automatically constitute an explanatory sufficiency, especially not in terms of underlying mechanisms, although it does lend more credibility, in a subjective sense.

What are the key issues involved in establishing mechanistic explanations of consciousness? First of all, we want a mechanistic (computational) explanation of the conscious/unconscious distinction, which is the bottom line. Without mechanistic explanations, we can never claim to have achieved a true understanding of the nature of consciousness. Only through contrasting the conscious and the unconscious may we understand the different characteristics of the two and thus the nature of consciousness. Furthermore, mechanistic explanations should account for the qualitative phenomenological difference between the conscious and the unconscious, in mechanistic terms. The explanatory sufficiency, generality, and parsimony of mechanistic constructs need careful attention also. Explanations should have sufficient explanatory power to account for all major aspects of an issue, in a parsimonious way and with generality. One way to achieve such explanatory power is through theoretical reduction, as opposed to mere empirical verification/confirmation. Reduction means mapping conceptual entities in one theoretical framework to some other entities in another that is more fundamental, more tangible, or better understood, and thereby grounding the former framework in the latter. For example, identifying biological substrates of consciousness such as the two-pathway view is important, but it by itself does not explain the phenomenological distinction between the conscious and the unconscious. It is an empirical verification and confirmation of the distinction, but not a reduction and not a sufficient mechanistic explanation of the phenomenological distinction. Likewise, implementing consciously accessible and inaccessible knowledge in semantic networks and production rules, respectively, is useful in that it instantiates a cognitive model and enables the simulation of cognitive data, but it again does not constitute a reduction and does not explain the phenomenological distinction.

In contrast, the representational difference view (as reviewed earlier) does attempt to explain or reduce the phenomenological distinction in mechanistic terms (Sun, 1999). By the above considerations, only the representational difference view has some promise.

Let us look more closely into this view. One crucial issue fundamental to this view is the definition of explicitness (i.e., accessibility) of representation, which the whole representational difference explanation hinges on. We intuitively consider localist representation to be more explicit and therefore more accessible, compared with distributed representation. But how do we substantiate this claim? In other words, how do we define explicitness of a representation (and thus representational accessibility)? Notably, there have been a variety of conflicting theories concerning explicit vs. implicit representation, often debating among themselves (e.g., see Hadley, 1995; Sun & Bookman, 1994). One point these theories do agree on is the fact that localist representation is more explicit and thus more accessible. Instead of repeating some (or all) of these theories, I will propose another, more mathematically grounded way of defining explicitness. The definition is based on how easily a representational item can be transformed into a corresponding conceptual label, which is a prerequisite for verbal reporting and explicit thinking (Fodor, 1975) and thus a proper basis for defining explicitness. The transformation can be viewed as an algorithmic process — an algorithm transforms representations into corresponding labels. Thus, the explicitness of a representation can be defined based on the complexity of the simplest possible algorithm that can be used to

transform it. Explicitness of a representation is therefore defined as the inverse of Kolmogorov complexity of its transformation, a theoretical measure of the algorithmic complexity (Li & Vitanyi, 1997). Without getting into the technical details of this definition, it suffices to point out that the transformation of localist representation has lower Kolmogorov complexity compared with the transformation of distributed representation. This definition of explicitness gives us a solid foundation for claiming that the localist representation is more accessible than the distributed one, and thus for the hypothesis that the difference in representational accessibility can explain the difference in phenomenological accessibility.

In this regard, there have been various theoretical dichotomies of the mind: for example, William James's (1890) distinction of empirical thinking and true reasoning, Dreyfus and Dreyfus's (1987) distinction of intuitive thinking and analytic thinking, and Smolensky's (1988) distinction of subconceptual processes and conceptual processes. These dichotomies can be mapped onto the representational difference (between localist and distributed representation), as argued for and demonstrated by Sun (1994, 1995, 1999) as well as Sun, Merrill and Peterson (2001), through examining the underlying mechanistic difference between the two sides of each dichotomy.

There have also been various empirical dichotomies: for example, implicit vs. explicit learning (Reber, 1989; Berry & Broadbent, 1988), implicit vs. explicit memory (Jacoby, Toth & Yonelinas, 1993), automatic vs. controlled processes (Shiffrin & Schneider, 1977), and unconscious vs. conscious perception (Merikle, 1992). These dichotomies can be accounted for, qualitatively, by the representational difference view, as discussed in Sun (1999). Reber (1989), based on psychological data, hypothesized that the primary difference between the explicit and implicit learning processes lies in the forms of their representations. Lewicki, Hill and Czyzewska (1992) and Squire, Knowlton and Musen (1993) had similar views in interpreting their experimental data. My view is an extension of these previous conjectures. A model based on this view, CLARION, has been used to capture a variety of cognitive data in a wide range of domains (Sun, 1999; Sun et al., 2001). The model postulates the following: (1) representational difference: two subsystems employ two different types of representations and thus have different degrees of accessibility; (2) learning difference: different learning methods are used for the two subsystems and thus the two subsystems have different learning characteristics; (3) manipulability of interaction: the combination of the outcomes from the two subsystems can be altered based on task situations. Based on these postulates, the model accounted for the aforementioned cognitive data. The representational difference view is well supported both theoretically and empirically.

The advantage of the representational difference view lies in the explanation of consciousness in terms of a mechanistic (computational) distinction, reducing a set of vague notions needing explanation to a set of notions that are much better understood, i.e. the reduction of the dichotomy of the conscious and the unconscious, and other similar or related dichotomies, to the more tangible dichotomy of the localist (symbolic) representation and the distributed representation.

2. Teleological views of consciousness

In terms of mechanistic explanations of consciousness (e.g., in the form of a cognitive model), I have shown that the explanations in terms of knowledge content, knowledge organization, or knowledge processing are untenable. Thus, it leaves us with one possibility — knowledge representation; that is, the difference in representation accounts for the phenomenological distinction between the conscious and the unconscious. However, a further issue concerning the functional role of consciousness needs to be addressed: why did evolution create such a distinction?

Let us explore the teleology of the conscious/unconscious distinction, in terms of both access and phenomenal consciousness. Access consciousness refers to the direct availability of the mental content for access (e.g., verbal report), while phenomenal consciousness refers to the phenomenal quality of mental content, that is, what something feels like and the immediacy and vividness of such feeling, as Thomas Nagel brought out in 'What is it like to be a bat?' (Nagel, 1974).

2.1. Access consciousness

With regard to the functional role of access consciousness, there have been various suggestions: for example, the veto view as suggested by Libet (1985), which states that the function of consciousness is to veto unconsciously initiated actions, or the counterbalance view as suggested by Kelley and Jacoby (1993), which is a generalization of the veto view. The question which these two views did not address is why one needs counterbalance, whether in the form of occasional veto or something else. On the other hand, Reber (1989), Stanley et al. (1989), and others claim that conscious and unconscious processes are each suitable for different situations, so that either a conscious or an unconscious process will be applied to a situation depending on which one is more suitable. Similarly, the language/planning view of Crick and Koch (1990) suggests that the function of consciousness is to enable the use of language and (explicit) planning. However, the question remains why one should use language and planning (in an explicit and conscious way) on top of unconscious processes.

An alternative explanation, based on the representational difference view of consciousness, is that the function of the conscious/unconscious distinction lies in the synergy that this distinction creates. As shown in Sun (1994, 1999), the interaction of the conscious and the unconscious (as two distinct processes, with different representations) can, in many common circumstances, lead to synergy in performance. This is the synergy view of consciousness proposed in Sun (1994, 1995, 1999).

Let us discuss this view in more detail. First of all, there is the question of the source of the synergy. As indicated by psychological data (e.g., on implicit learning and implicit memory), conscious processes tend to be more crisp and focused (selective), while unconscious processes tend to be more complex, broadly scoped (unselective), and context-sensitive (see Reber, 1989; Berry & Broadbent, 1988; Seger, 1994 regarding complexity; see Hayes & Broadbent, 1988 regarding selectivity). Due to their vastly different characteristics, it should not come as a surprise that the interaction of the conscious and the unconscious leads to synergistic results. In the statistical literature, it is well known that combining diversified processes can improve performance (e.g., Breiman, 1996; Efron & Morris, 1973). It is not farfetched to speculate that synergy was the reason evolution created consciousness.

There is some psychological evidence directly in support of the synergy view. Willingham et al. (1989) found that those subjects who acquired full explicit (conscious) knowledge (in a serial reaction time task) appeared to learn faster than those who did not have full explicit knowledge. Stanley et al. (1989) reported that subjects' learning improved (in a dynamic control task) if they were asked to generate verbal instructions for other subjects along the way during learning. That is, a subject was able to speed up his/her own learning through the emphasis on explicit processes. In addition, in terms of learned performance, Willingham et al. (1989) found that subjects who verbalized (while performing serial reaction time tasks) were able to attain a higher level of performance than those who did not verbalize. This phenomenon may also be related to the self-explanation effect reported in the cognitive skill acquisition literature (Chi, Bassok, Lewis, Reimann & Glaser, 1989). In all of these cases, it can be the emphasis on conscious processes that helped performance. Finally, synergy has also been demonstrated through computational simulations, in the domains of commonsense reasoning (Sun, 1994, 1995) and skill learning, including all of the aforementioned tasks (Sun, 1999; Sun et al., 2001).

The localist representation that conscious processes use enables explicit control and manipulation, which constitute meta-level processes (Nelson, 1993). Such control and manipulation can include, for example, selecting a reasoning method, controlling the direction in which reasoning goes, enable/disable certain inferences, or evaluating the progress of reasoning. When meta-level processes become assimilated into regular processes, further meta-level processes can be developed on top of them. Thus, potentially, we may have many levels of self-control and self-manipulation of mental processes. (By no means am I claiming that unconscious processes cannot be controlled and/or manipulated at all, but that it is clearly more difficult to do so due to the less accessible representation.) Meta-level processes may be another reason for synergy.

Therefore, conscious processes are characterized by explicit (localist/symbolic) representations, as well as explicit meta-level regulation (i.e., control

and manipulation of processes operating on explicit representations). These two aspects together distinguish conscious processes from unconscious processes (as in the CLARION model; Sun et al., 2001). The teleological explanation of access consciousness follows directly from these two aspects. Synergy results from the co-existence of the two different types of representations and consequently the co-existence of the two different types of processes operating on the two types of representations, respectively.

It should be emphasized that this dichotomous structure is minimal and necessary for understanding cognition, in the sense of a minimal architecture. The previous arguments show that there are qualitative and fundamental differences between the two types of processes. It is difficult to believe that one side of the dichotomy can be derived from the other ontogenetically, without some minimal structure to begin with. Thus, the distinction should be somehow innate. Notice also that most of the high animal species are not capable of developing an elaborate and complete conceptual system (with symbolic processing abilities) while humans rarely fail to develop such a system. The constancy of this interspecies difference points to the innateness of such a difference and the innateness of the dichotomous structure. It is thus more convincing to hypothesize that the dichotomous structure is a given innate structure for humans, and should be incorporated into the architecture itself.

This explanation of consciousness also encompasses the other views concerning the functional roles of consciousness. According to the synergy view, consciousness can certainly veto or counterbalance unconsciousness, given the right circumstances when such veto or counterbalance improves performance (that is, if they lead to synergy). Likewise, the synergy view can explain the situational difference view, in that, while in general both types of processes are present due to their synergistic effect, in some extreme cases it may be advantageous to use only conscious or unconscious processes. For example, when a task is well practiced and therefore there is no longer a need for synergy, unconscious processes alone suffice, which leaves conscious processes for other tasks. This phenomenon is known as automatization. The synergy view also encompasses the language/planning view, because it explains why one should use (consciously) language/planning on top of unconscious processes: it is because of the possibility of improved performance through the interaction of both types of processes.

The question now is whether conscious awareness arises solely as a by-product of this distinction of two representations, or whether it is the necessary result of the distinction. Since localist representation is a prerequisite for synergy and meta-level control, and localist representation necessarily leads to accessibility, we believe that conscious awareness is a necessary result, not a mere coincidence. I will discuss this point further next, under the heading of 'phenomenal consciousness'.

2.2. Phenomenal consciousness

The functional role of phenomenal consciousness is a far more difficult question. Let us examine the notion of qualia, which refers to the 'phenomenal quality' of conscious experience (Nagel, 1974; Chalmers, 1993; Block, 1994). The major problem with the notion lies in the difficulty it poses for functionalism, the currently popular view in philosophy of mind that the defining feature of any mental state is the set of causal relations it bears to other mental states, environmental effect on the body, and behavior of the body. If cognitive functioning can occur without qualia, then qualia may not have a functional role and functionalism cannot explain them. I contend that it is possible that there are many functionally equivalent organizations of mental states, at least at a gross level. Many of these functional structures, though capable of generating certain behavioral patterns, do not lead to phenomenal experience (or at least not the right kind of phenomenal experience). A functional organization capable of generating behavior alone is not a necessary and sufficient condition for consciousness (Searle, 1980). However, there is nothing that prevents some functional organization (among those that are capable of the range of behavior) having phenomenal conscious experience.

We should distinguish different possible functional structures capable of the same behavior, especially in terms of separating those that are capable of the behavior but not the more intrinsic properties of the mind from those that are capable of both the

behavior and the more intrinsic properties of the mind (Sun, 1999). Thus, the search for phenomenal consciousness can, after all, be the search for a functional structure — the right functional structure.

Beyond functionalism, we also need to address the physical nature of phenomenal consciousness. A mechanistic explanation must be attempted, despite the failure that we have been seeing in such attempts so far. There are plenty of examples of dualistic 'explanations' of difficult phenomena having evaporated after further explorations that led to a better mechanistic account.

To see how the representational difference view of consciousness can account for phenomenal consciousness, let us examine some possibilities. Qualia might be accounted for by the totality of a multimodal multi-level organization and its collective states (i.e., the total-states). These total-states are of extremely high complexity involving a nexus of external perception (of many modalities), internal states, emotion, implicit and explicit memory, implicit and explicit representation, implicit and explicit decision making, and so forth. This was termed the 'manifold' by Van Gulick (1993), and the 'superposed complex' by Lloyd (1995). In this approach, a particular kind of phenomenal quality may be accounted for by a particular region of a total-state space (involving the totality of all the aforementioned aspects), which gives rise to the sense of what something is like (Nagel, 1974) without explicit awareness of all the details. The complexity of organization may explain the 'irreducibility' of phenomenal experience — the difficulty (or impossibility) of describing phenomenal qualities (qualia). Clearly, a region in a total-state space can only be formed on the basis of particular organizations of modules and levels that support a space of total-states (Chalmers, 1993). Qualia are thus partially the result of the organization of cognitive apparatuses.

Are qualia epiphenomal according to this account? The answer is no, if we contrast the condition in which we are aware (with qualia) with the one in which we are not (without qualia). Qualia evidently entail high-intensity, qualitatively distinguishing states, demanding the attention of the agent. Although it is possible to imagine that there are corresponding unconscious states (without qualia) that can lead to similar information processing and action on the part of the agent (Searle, 1980), it is possible that the highly alerting characteristics of qualia serve some useful functions, such as attention getting, attention focusing, cognitive resources management (Nelson, 1993) and, in some cases, bringing about conscious processing. It is only natural to speculate that evolution created phenomenal consciousness because of these useful functions, which also lead to synergy (in a sense).

3. Concluding remarks

In this paper, I focused on the issue of the physical (mechanistic or computational) basis of consciousness, proposing the framework of a mechanistic account of consciousness and, in turn, a teleological account of consciousness in this framework. Analyses and argumentation showed that the difference between localist (symbolic) representations and distributed representations (as employed in the connectionist theorizing and modeling) led to a plausible account of consciousness and its functional role.

Much more work can be conducted along this direction centered on representational difference. Such work may include further investigation of detailed cognitive models (along the line of CLARION as mentioned earlier): for instance, we would like to see more detailed specifications of conscious processes, which should lead to a more detailed and more precise account of phenomenal consciousness; similarly, we would like to see more detailed specifications of unconscious processes as well, which should explain, in-depth, human intuition and reflexive behavior.

References

Anderson, J. R. (1983). *The architecture of cognition*, Harvard University Press, Cambridge, MA.

Anderson, J. R. (1993). *Rules of the mind*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Baars, B. (1988). *A cognitive theory of consciousness*, Cambridge University Press.

Berry, D., & Broadbent, D. (1988). Interactive tasks and the implicit–explicit distinction. *British Journal of Psychology 79*, 251–272.

Block, N. (1994). On a confusion about a function of consciousness. *Brain and Behavioral Sciences*.

Bower, G. (1996). Reactivating a reactivation theory of implicit memory. *Consciousness and Cognition 5*(1/2), 27–72.

Breiman, L. (1996). Bagging predictors. *Machine Learning 24*, 123–140.

Chalmers, D. (1993). Towards a theory of consciousness. Ph.D Thesis, Indiana University.

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, P. (1989). Self-explanation: how students study and use examples in learning to solve problems. *Cognitive Science 13*, 145–182.

Clark, A. (1992). The presence of a symbol. *Connection Science 4*, 193–205.

Collins, A., & Loftus, J. (1975). Spreading activation theory of semantic processing. *Psychological Review 82*, 407–428.

Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in the Neuroscience 2*, 263–275.

Damasio, A. (1994). *Decartes' error*, Grosset/Putnam, New York.

Dennett, D. (1991). *Consciousness explained*, Little Brown.

Dreyfus, H., & Dreyfus, S. (1987). *Mind over machine: the power of human intuition*, The Free Press, New York.

Edelman, G. (1989). *The remembered present: a biological theory of consciousness*, Basic Books, New York.

Efron, B., & Morris, C. (1973). Combining possibly related estimation problems. *Journal of the Royal Statistical Society 35*, 379–421.

Freeman, W. (1995). *Societies of brains*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Fodor, J. (1975). *The language of thought*, Crowell.

Hayes, N., & Broadbent, D. (1988). Two modes of learning for interactive tasks. *Cognition 28*, 249–276.

Hadley, R. (1995). The explicit–implicit distinction. *Minds and Machines 5*, 219–242.

Heidegger, M. (1927). *Being and time*, Harper and Row, New York, English translation, 1962.

Hunt, E., & Lansman, M. (1986). Unified model of attention and problem solving. *Psychological Review 93*(4), 446–461.

Jackendoff, R. (1987). *Consciousness and the computational mind*, MIT Press, Cambridge, MA.

Jacoby, L., Toth, J., & Yonelinas, A. (1993). Separating conscious and unconscious influences of memory: measuring recollection. *Journal of Experimental Psychology: General 122*, 139–154.

James, W. (1890). *The principles of psychology*, Dover, New York.

Kelley, C., & Jacoby, L. (1993). The construction of subjective experience: memory attribution. In: Davies, M., & Humphreys, G. (Eds.), Consciousness, Blackwell, Oxford.

LeDoux, J. (1992). Brain mechanisms of emotion and emotional learning. *Current Opinion in Neurobiology 2*(2), 191–197.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. *American Psychologist 47*, 796–801.

Li, M., & Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications*, Springer, Heidelberg.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences 8*, 529–566.

Lloyd, D. (1995). Consciousness: a connectionist manifesto. *Minds and Machines 5*, 161–185.

Logan, G. (1988). Toward a theory of automatization. *Psychological Review 95*(4), 492–527.

Mathis, D., & Mozer, M. (1996). Conscious and unconscious perception: a computational theory. In: Proceedings of the 18th Annual Conference of the Cognitive Science Society, Erlbaum, Mahwah, NJ, pp. 324–328.

Merikle, P. (1992). Perception without awareness: critical issues. *American Psychologists 47*, 792–795.

Milner, D., & Goodale, N. (1995). *The visual brain in action*, Oxford University Press, New York.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review 4*, 435–450.

Nelson, T. (Ed.), (1993). Metacognition: core readings, Allyn and Bacon.

Penrose, R. (1994). *Shadows of the mind*, Oxford University Press, Oxford.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General 118*(3), 219–235.

Revonsuo, A. (1993). Cognitive models of consciousness. In: Kamppinen, M. (Ed.), Consciousness, cognitive schemata and relativism, Kluwer, Dordrecht, pp. 27–130.

Schacter, D. (1990). Toward a cognitive neuropsychology of awareness: implicit knowledge and anosagnosia. *Journal of Clinical and Experimental Neuropsychology 12*(1), 155–178.

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

Seger, C. (1994). Implicit learning. *Psychological Bulletin 115*(2), 163–196.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing II. *Psychological Review 84*, 127–190.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences 11*(1), 1–74.

Squire, L., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology 44*, 453–495.

Stanley, W., Mathews, R., Buss, R., & Kotler-Cope, S. (1989). Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology 41A*(3), 553–577.

Sun, R. (1994). *Integrating rules and connectionism for robust commonsense reasoning*, Wiley, New York.

Sun, R. (1995). Robust reasoning: integrating rule-based and similarity-based reasoning. *Artificial Intelligence 75*(2), 241–296.

Sun, R. (1999). Accounting for the computational basis of consciousness: a connectionist approach. *Consciousness and Cognition 8*, 529–565.

Sun, R., & Bookman, L. (Eds.), (1994). Computational architectures integrating neural and symbolic processes, Kluwer Academic, Norwell, MA.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: a bottom-up model of skill learning. *Cognitive Science*, in press.

Van Gulick, R. (1993). Understanding the phenomenal mind. In: Davies, M., & Humphreys, G. (Eds.), Consciousness, Blackwell, Oxford.

Willingham, D., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition 15*, 1047–1060.