

Dual-Process Theories, Cognitive Architectures, and Hybrid Neural-Symbolic Models

Ron Sun^{a*}

^a*Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

Editors: First Editor, University or Company name, Country; Second Editor, University or Company name, Country

Solicited reviews: First Solicited Reviewer, University or Company name, Country; Second Solicited Reviewer, University or Company name, Country

Open reviews: First Open Reviewer, University or Company name, Country; Second Open Reviewer, University or Company name, Country

Abstract. This paper discusses relationships among hybrid neural-symbolic models, dual-process theories, and cognitive architectures. It provides some historical backgrounds and argues that dual-process (implicit and explicit) theories have significant implications for developing neural-symbolic models. Furthermore, computational cognitive architectures can help to disentangle issues concerning dual-process theories and help the development of neural-symbolic models (in this way and in other ways).

Keywords: neural-symbolic, neurosymbolic, dual process, cognitive architecture, Clarion, implicit, explicit, psychology

1. Introduction

In this relatively short position paper, I will briefly discuss hybrid neural-symbolic models, dual-process theories, and cognitive architectures, as well as mutually beneficial relationships among them and their relevance to each other. I will provide some historical backgrounds. Based on that, I argue that dual-process (implicit and explicit) theories have significant implications for developing neural-symbolic models. Furthermore, through computational means, cognitive architectures can help to disentangle complex issues concerning dual-process theories and thus help with the further development of hybrid neural-symbolic models (in this as well as other ways). The present position paper inevitably reflects, of course, a personal perspective resulting from personal experiences and is not meant to be comprehensive (nor technically detailed).

Neural networks and deep learning are good at many things, but they have shortcomings and limitations, despite their popularity, as has long been known (e.g., difficulties with certain types of extrapolation, generalization, abstraction, explanation, and so on; see, e.g., very early arguments at the beginning

*Corresponding author. E-mail: dr.ron.sun@gmail.com

of “new connectionism”, such as Fodor and Pylyshyn, 1988; Marcus, 2001; Pinker and Prince, 1988).¹ But, as many have noticed, neural networks and symbolic models have some complementary strengths. Thus, combining them in some way seems a good idea naturally, and the resulting models can be advantageous — For instance, they can be more expressive, more versatile, or more capable otherwise. Hence the idea of hybrid neural-symbolic (or “neurosymbolic”) models was born, transcending many heated, conflicting claims of advantages and disadvantages of each paradigm (Sun, 2002a, 2015b).

This idea immediately harkens back to the 1990s when hybrid models *first* emerged (see, e.g., Sun and Bookman, 1994). At the present time (at the time when the present article is being written), this idea is clearly not something entirely new (or new at all). In other words, the old has become new again (in ways much like fashion, albeit with some necessary modifications and enhancements). Over time, such models have been known under various names. Besides neural-symbolic or “neurosymbolic” models, they have also been variously known as connectionist-symbolic model, hybrid symbolic neural networks, or simply hybrid system, during different periods in their history (see, e.g., Sun, 2002a, 2015b; Sun and Bookman, 1995; Wermter and Sun, 2000).

Despite the fact that this important (or even crucial) idea was very much ignored or downplayed early on by the mainstreams on both the neural and the symbolic side during the 1990s, by now, even some strong advocates for and major contributors to symbolic AI have suggested that hybrid neural-symbolic models should indeed be pursued (e.g., Kautz, 2020). At the same time, some strong advocates for and major contributors to neural networks and deep learning have also started to see that. When people from the two opposing camps converge onto the same idea (along with people who have always been in the middle already), the likelihood of this idea leading to major advances has just skyrocketed (some of them even invoked good old dual-process theories in support of their arguments; more on this later). There have also been quite a number of recent events promoting this theme (which I shall not enumerate, for this would take too long), which include the launch of the present journal.

However, there have been many possible ways of structuring such models; that is, there exist many different ways of combining symbolic processing and neural networks (Sun, 2002a). Many of them, however, appear ad hoc or task specific, or otherwise problematic. Therefore, a major question that immediately comes to mind in regard to hybrid models is: how *should* we structure them after all? How do we structure them in a way that avoids being ad hoc? How do we make them more principled, that is, justified or justifiable by some fundamental laws or principles? Obviously, it is preferable that they should have a more solid theoretical (as well as empirical) basis — that is, to be justifiable by some fundamental laws or principles.

2. Some Early History (Or What You Might Call Pre-history)

Some early work on neural-symbolic models is worth mentioning here, due to their relevance to what has been discussed thus far. Even during the early days of such models (that is, in the 1990s), there was some important work done. One may choose to view such early work as the pre-history to today’s neural-symbolic models (especially when they involve complex deep neural networks that were not yet fully developed or popularized back then), or may just simply view it as the early history of this evolving field, thus perceiving a more continuous process of development and progression. In particular, it should be noted that in the 1990s, there were some edited volumes (books) with contributions from many

¹ To that list of shortcomings, one may also add difficulties with catastrophic forgetting, transfer of learning, distributional changes, amount of data required, computational cost, and so on, as researchers discovered over time.

authors, which meant that they were likely representative of the state of the art of the work done during that period and contained many different models and ideas, some of which may still be relevant and useful today. Therefore, we shall have a look.

One of them is Sun and Bookman (1994) (which, by the way, happened to be the first such volume on the topic of neural-symbolic models). One of the contributions of this book (and some other related ones) is dealing with *analysis, classification, categorization, and cataloguing* of neural-symbolic models (e.g., different ways of designing and constructing such models, different techniques used, and so on).

For example, some general categories discussed in Sun and Bookman (1994) included: (1) designing specialized, structured localist neural networks for symbolic processing (e.g., some forms of logical reasoning); (2) embedding symbolic processing within usual neural networks with distributed representations; (3) combining separate neural and symbolic processing modules; (4) using neural networks as elements (modules) in a structured, symbolic architecture (see Sun and Bookman, 1994 for details). On the other hand, from the perspective of a whole system, (1) a system could be a single-module system, in which (1.1) representations can be symbolic, localist, or distributed, and (1.2) mappings between symbols and representations could be direct, translational, or transformational (fully utilizing complex neural dynamics); or (2) it could be a multi-module system, which (2.1) could be either homogeneous or heterogeneous and (2.2) could vary in terms of granularity of modules, and (2.3) the relationship between modules could be either loosely or tightly coupled. There can also be finer distinctions and variations (Sun and Alexandre, 1997). Logics and symbolic or probabilistic reasoning could have various neural implementations or approximations. These ideas represent the state of the art in the 1990s, so, obviously, there may be many newer methods and ideas by now (e.g., back then, there was no complex, sophisticated deep learning models or Big Data). However, note that these models and methods were designed to address the shortcomings of purely neural or purely symbolic models; these shortcomings were known even then but remain a challenge today despite the progress that has been made. Many of the concepts described above (such as localist vs. distributed representations, single- vs. multi-module systems, loosely vs. tightly coupled modules) and many common structures of neural-symbolic systems identified then (such as direct mapping of symbolic models to neural networks, embedding symbolic processing within neural networks, multiple neural and/or symbolic modules with different control and coordination methods) remain viable in describing today's neural-symbolic models.

However, one question that one faced then (and is still facing today) is: Given these options, that is, given these many different ways of combining symbolic processing and neural networks, structurally, mechanistically, and process-wise, how can we produce better hybrid models going forward? That is, how should we best structure them, choosing among all these options? In particular, how do we structure them in a principled and justifiable way, rather than in an ad hoc or task-specific way?

I argued then and is still arguing today (e.g., in my books, such as Sun, 1994, 2002b, 2016b) that a better approach is structuring them in a cognitively (i.e., psychologically) motivated and justified way — that is, based on human mental architecture (which is, in turn, inferred based on experimental psychology, neuroscience, etc.; see Newell, 1990; Sun, 2008). In particular, within the human mental architecture, we need to take into account dual processes (e.g., as has been variously termed as implicit vs. explicit, unconscious vs. conscious, intuition vs. reason, System 1 vs. System 2, and so on, albeit sometimes with somewhat different connotations). Incidentally, dual-process (or two-system) theories have become quite popular lately (see, e.g., Kahneman, 2011), although they are not new either (see, e.g., Reber, 1989; Sun, 1994; etc.). Dual-process theories can provide both philosophical and psychological justifications for (at least some of) hybrid neural-symbolic models. In fact, they constitute both theoretical and empirical grounding for hybrid neural-symbolic models.

3. Dual-Process Theories

So, at this juncture, we should examine (briefly) dual-process theories that have been proposed — There have been quite a few of them out there, and their historical roots go even deeper. Therefore, it is worthwhile to look into them.

Theoretical or philosophical proposals concerning two types of processes have long existed, dating back to even before the inception of cognitive science and artificial intelligence. For instance, Martin Heidegger's distinction of the pre-ontological versus the ontological is an early, very abstract version of such a duality (Heidegger, 1927). William James's (1890) distinction between "empirical thinking" and "true reasoning" is even more clearly relevant. Going back even earlier, Immanuel Kant's (1781) dictum was: "Thoughts without intuitions are empty, intuitions without concepts are blind". "Intuition and concepts constitute... the elements of all our knowledge, so that neither concepts without an intuition in some way corresponding to them, nor intuition without concepts, can yield knowledge" (Kant, 1781).

In contemporary experimental psychology and neuroscience, there have been empirically based proposals in this regard. The distinction of implicit and explicit processes has been empirically demonstrated in the implicit memory literature (e.g., Schacter, 1987). The distinction of implicit and explicit processes has also been empirically demonstrated in the implicit learning literature (e.g., Reber, 1989). In social psychology, there have been dual-process models that are roughly based on the co-existence of implicit and explicit processes (Chaiken and Trope, 1999). Similar distinctions have been proposed by other researchers, based on similar or different empirical or theoretical considerations (e.g., Erickson and Kruschke, 1998; McClelland, McNaughton, and O'Reilly, 1995; Milner and Goodale, 1995; Sun, 1994, 1995; etc.).

Here is my version of a dual-process theory from a long time ago. In my 1994 book, I made the following distinction:

".... cognitive processes are carried out in two distinct 'levels' with qualitatively different mechanisms. Each level encodes a set of knowledge for its processing, and the coverage of the two sets of knowledge encoded by the two levels overlaps substantially." (Sun, 1994)

That is, the two "levels" (the two systems or modules, or two types of processes) encode somewhat similar content, but they encode their content in different ways: Symbolic versus subsymbolic representations are used, respectively. Therefore, they utilize different mechanisms. They can thus have qualitatively different "flavors", including different degrees of accessibility (leading to different amounts of conscious awareness). One may be termed the explicit level, and the other the implicit level (or, alternatively, they may be termed System 1 and System 2, respectively, or whatever other terms one may use as mentioned earlier).

Furthermore, according to this account, one reason for having the two different levels is that these different levels (with different representations and different mechanisms) can potentially work together synergistically, complementing and supplementing each other (as argued in Sun, 1994, 1995, 2002b, and Sun et al., 2005). This is probably, at least in part, the reason why there are these two levels — That is why nature has chosen this design. I refer to the above two points as the dual representation hypothesis and the synergy hypothesis, respectively (see Sun, 2002b, 2016a).

Many other dual-process theories exist, although some of them can be problematic sometimes. For instance, a comparatively more recent view (compared with those precedent views mentioned earlier) was proposed by Kahneman (e.g., Kahneman, 2011). The main ideas of his were as follows: There are

two styles of processing: intuition and reasoning. Intuition (or System 1) is based on associative processes, fast and automatic, involving strong emotional bonds, based on formed habits, and difficult to change or manipulate. Reasoning (or System 2) is slower, more volatile, and subject to conscious judgments and attitudes.

At roughly around the same time, Evans (2003) espoused essentially the same view. According to Evans, System 1 is “rapid, parallel and automatic in nature: only their final product is posted in consciousness”. He also noted the “domain-specific nature” of System 1 learning. System 2 is, on the other hand, “slow and sequential in nature and makes use of the central working memory system”; it “permits abstract hypothetical thinking that cannot be achieved by System 1”. Furthermore, there is the “default-intervention” relationship between the two systems: System 1 forms the default response unless there is active intervention from System 2

However, it should be noted that some of these claims above, unfortunately, may be somewhat simplistic. For one thing, intuition can be slow sometimes (see, e.g., Helie and Sun, 2010). For another, intuition can be subject to conscious control and manipulation; that is, it may not be entirely “automatic” (e.g., Curran and Keele, 1993; Stadler, 1995). Furthermore, intuition can be subject to conscious “judgment” (e.g., Libet, 1987). Moreover, explicit thinking may be engaged right from the outset and can be constantly present, not just an occasional intervention, and there may be complex interactions between the two systems (Helie and Sun, 2010). The reader is referred to Sun (2015a, 2016a) for detailed discussions of these points. For various alternative views, see also Evans and Frankish (2009) and Macchi et al. (2016).

To come up with a more nuanced and more precise characterization of the two systems, it is important that we ask some pertinent questions taking into account architectural and mechanistic nuances of the mind. For instance, for either type of process, there can be the following important questions (which are mechanistic or process-oriented):

- How deep is the processing (in terms of precision, certainty, and so on)?
- How broad is the processing (e.g., how much information is involved)?
- How degraded (incomplete, inconsistent, or uncertain) is the information available?
- How typical or atypical is the situation to be addressed?
- Is the process involved procedural or declarative?
- How many processing cycles are needed considering the factors above?

There are many other similar or related questions, including those related to further divisions of modules within each level and across levels, such as procedural vs. declarative modules (see, e.g., Anderson and Lebiere, 1998) and their relations to the implicit-explicit distinction (Reber, 1989). (Note that the relationship between System 1 vs. 2 and procedural vs. declarative processes is a topic that is worth an article-length or even a book-length treatment by itself. It is certainly not simply the case that System 2 involves only declarative processes with complex knowledge representation, while System 1 involves only procedural processes with simple learning, as some researchers in the field tend to believe. See Sun, 2012 for details.)

These questions need to be addressed, to ensure cognitive-psychological realism of theories or models. To put it in another way, the complexity of the matter should be adequately understood and should be properly taken into account, especially in any mechanistic models or theories that purport to value cognitive-psychological realism (Sun, 2016a). Hybrid neural-symbolic models belong (or at least should belong) to this category of models.

4. Computational Cognitive Architectures

To better understand relevant issues and to better answer these questions raised earlier, we need an overall theoretical framework, as well as detailed computational (mechanistic) models that can substantiate proposed theories. That is where *computational cognitive architectures* (Newell, 1990) come into play — because they are both at the same time and thus they can go a long way in helping to clarify these issues. Note that the notion of computational cognitive architecture, as developed within cognitive science, refers to a broadly scoped, domain-generic computational cognitive (i.e., psychological) model, capturing the essential structures, mechanisms, representations, and processes of the mind, to be used for broadly scoped, multi-domain, multi-level analysis of behavior (Sun, 2016b). Needless to say, they are meant to be cognitively-psychologically realistic.

Computational cognitive architectures provide a general framework, which facilitates more detailed explorations of the mind (through providing constraints and tools; Taatgen and Anderson, 2023). This general framework can then lead to specific theories about components or functionalities of the mind. We accomplish this through specifying computational details of cognitive (psychological) mechanisms and processes; in other words, we embody our theory of the mind and its various functionalities in computer programs. In so doing, we generate runnable computer programs, which enables the simulation of empirical human data; simulation results can then be compared to human data, in a qualitative or quantitative way, to validate the models or to develop better ones. Through this iterative and detailed process of comparing human data and simulations, we can achieve better understanding of the inner workings of the mind. This is the essence of computational psychology (Sun, 2008). With cognitive architectures, we focus first on a general theory (a general framework), then move onto more specific theories (Newell, 1990).

As an example, the Clarion cognitive architecture is a dual-process, dual-representation cognitive architecture, incorporating both implicit and explicit processes. It uses dual representations, that is, both symbolic and neural representations, for computationally capturing and differentiating explicit and implicit processes (see Sun, 2002b, 2016b regarding Clarion; see also Sun, 1994 for ideas leading up to it). Clarion has been aimed to be an *integrative* and *comprehensive* computational framework. For instance, Clarion encompasses many different mechanisms, covers many different psychological functionalities and domains, and at the same time develops a more nuanced view of the two systems (or the two processes). Understanding intelligence and cognition (both natural and artificial kinds), especially issues related to dual processes, requires no less than such a comprehensive model. But, in turn, dual processes are crucial for this and many other cognitively-psychologically realistic models, serving as their theoretical inspiration and both theoretical and empirical justifications.

One can view modeling and simulation based on cognitive architectures as a way of theoretical interpretation (albeit in a detailed and precise, namely mechanistic, fashion). In this way, Clarion has indeed helped to clarify many issues related to dual processes mentioned earlier. For instance, some claimed that implicit processes (“intuition”) are always faster than explicit processes (“reasoning”). While in terms of procedural processes, this is generally true, using theoretical interpretations through simulation (e.g., Sun et al., 2005). In terms of declarative processes, however, this is generally not true (e.g., Helie and Sun, 2010). See Sun (2015a, 2016a) for analysis of these and other issues related to dual-process theories and beyond.

The established methodology of developing computational cognitive architectures, on the basis of empirical data and findings from psychology, neuroscience, and so on, constitutes at the same time a principled (and thus arguably preferred) way of developing neural-symbolic systems in general, regardless of whether these systems are for theoretical purposes or for practical applications (cf. Pew and

Mavor, 1998; Sun, 2008). In relation to practical applications, the advantages of this approach include, for example, that a system resulting from this methodology will likely be more similar to humans in some fundamental way, that it is possible that such a system is more easily understood by humans (and vice versa), and that it may be easier for such a system to communicate with humans, not just through language, but also through other explicit or implicit means of communication (Sun, 2020).

5. Some Further Remarks

Without going further into details of how dual processes can be, and have been, further developed within Clarion and in turn helped the development of neural-symbolic models (due to the space limit, for one thing, but also due to the nature of a position paper), I would like to simply highlight a few key points:

- Dual-process theories have significant implications for developing computational cognitive architectures and cognitively-psychologically realistic neural-symbolic models (Sun, 1994, 2002).
- If cognitive-psychological realism is what one wants to achieve in developing theories or models, dual-process theories must be taken into consideration, not just as an option, not merely for expediency, but as one fundamental requirement (Sun, 2016a). (Of course, it is obvious that cognitive-psychological realism may not be for everyone, especially those who are only concerned with AI applications in the short run. But, even in those cases, there have been calls for dual-process theories to be applied, for the sake of practical expediency.)
- Dual-process theories can serve as the inspiration (and both theoretical and empirical justifications) for computational cognitive architectures and neural-symbolic models in general (especially cognitively-psychologically motivated ones).

Recent developments of deep learning (but see, e.g., Schmidhuber, 2015 for its long history) have been exciting, especially the successes that convolutional neural networks, Transformers, large language models based on Transformers, and other related approaches enjoyed. Many ideas that have been developing over the past three-four decades came together in a metaphoric big bang. Such developments have also introduced many new issues and many new possibilities in relation to neural-symbolic models (Garcez et al., 2015; Garcez and Jiménez-Ruiz, 2022; Hitzler et al., 2022), including, for example, probabilistic approaches towards integrating neural and symbolic processes, reasoning on the basis of large language models, and so on. However, there are reasons to believe that these fundamental points that were made earlier still stand (although adjustments and adaptations might be needed along the way, of course). It should be pointed out that the various recent calls for attention to the distinction between System 1 and System 2 in developing artificial intelligence systems involving deep learning, although timely and important, are nothing new either — Systems and models with the same or similar distinctions have been developed at one point or another, or (in some cases) have been in continuous development, non-stop, since the 1990s. Such ideas have stood the test of time.

Going back to the points mentioned earlier in our brief review of the early history of neural-symbolic models, dual-process theories point to multi-module, multi-representational, heterogeneous neural-symbolic systems as the basis of the human mental architecture (see Sun and Bookman, 1994; Sun and Alexandre, 2007; Wermter and Sun, 2001). Further empirical and/or theoretical support pointing in that direction has been mounting over the past decades.

It is worth repeating that, despite the fact that dual-process theories are quite popular right now (as clearly indicated by the recent interests with regard to System 1 and System 2), some issues involved in

dual-process theories (e.g., relative speed, intention, control, and interaction) are more complex than often assumed. In this regard, it is important to gain a more fine-grained understanding, and further work is needed to disentangle these issues (Sun, 2016a). These issues, however, are important for developing cognitive architectures, and in turn, cognitive architectures can help in disentangling these and other theoretically important issues, by providing an overarching framework as well as by enabling computational simulation. Together, they may lead to better neural-symbolic models in the future; in other works, they may guide the development of future neural-symbol systems that are principled and cognitively-psychologically realistic. Nevertheless, there is, of course, still a long way to go, and many new developments, breakthroughs, enhancements, and refinements are very much needed (e.g., for some of the more recent developments, see Chaudhuri et al., 2021; Dong, 2021; Dong et al., 2022; Lamb et al. 2020; Mekik, 2022; amongst many others).

Acknowledgement

The present article was written while the author was supported (in part) by ARI grant W911NF-17-1-0236 and IARPA HIATUS contract 2022-22072200002. The views and conclusions contained herein are those of the author's and should not be interpreted as necessarily representing the official policies and positions, either expressed or implied, of those agencies.

References

- [1] Anderson, J. R. and C. Lebiere, (1998). *The Atomic Components of Thought*. Lawrence Erlbaum Associates, Mahwah, NJ.
- [2] Chaiken, S. and Y. Trope (eds.), (1999). *Dual Process Theories in Social Psychology*. Guilford Press, New York.
- [3] Chaudhuri, Swarat, Kevin Ellis, Oleksandr Polozov, Rishabh Singh, Armando Solar-Lezama, and Yisong Yue (2021), Neurosymbolic Programming, *Foundations and Trends in Programming Languages*. 7 (3), 158–243. DOI:10.1561/2500000049.
- [4] Curran, T. and S. Keele, (1993). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 19, 189-202.
- [5] Dong, T. (2021). *A Geometric Approach to the Unification of Symbolic Structures and Neural Networks*. Springer, Berlin.
- [6] Dong, T., Achim Rettinger, Jie Tang, Barbara Tversky, and Frank van Harmelen (eds.). (2022). Structure and Learning (Dagstuhl Seminar 21362), *Dagstuhl Reports*, 11 (8). 11-34. Published by Schloss Dagstuhl, Leibniz-Zentrum für Informatik, Dagstuhl, Germany. DOI:10.4230/DagRep.11.8.11
- [7] Erickson, M. and J. Kruschke, (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*. 127, 107-140.
- [8] Evans, J. (2003). In two minds: dual-process accounts of reasoning, *Trends in Cognitive Sciences*, 7 (10), 454-459.
- [9] Evans, J. and K. Frankish (eds.), (2009). *In Two Minds: Dual Processes and Beyond*. Oxford University Press, Oxford, UK.
- [10] Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2), 3–71. [https://doi.org/10.1016/0010-0277\(88\)90031-5](https://doi.org/10.1016/0010-0277(88)90031-5)
- [11] Garcez, Artur; Besold, Tarek; De Raedt, Luc; Földiák, Peter; Hitzler, Pascal; Icard, Thomas; Kühnberger, Kai-Uwe; Lamb, Luís; Mikkilainen, Risto; Silver, Daniel (2015). Neural-Symbolic Learning and Reasoning: Contributions and Challenges. AAAI Spring Symposium on Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches. Technical Report SS-15-03, AAAI Press, Palo Alto, CA.
- [12] Garcez, Artur d'Avila, and Ernesto Jiménez-Ruiz, (2022). *Neural-Symbolic Learning and Reasoning: Proceedings of the 16th International Workshop on Neural-Symbolic Learning and Reasoning*, Windsor Great Park, United Kingdom, September 28-30, 2022.
- [13] Heidegger, M. (1927). *Being and Time*. (English translation published by Harper and Row, New York. 1962.)

- [14] Helie, S., and R. Sun, (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, Vol.117, No.3, pp.994-1024. 2010.
- [15] Hitzler, P, et al. (2022). Neuro-symbolic approaches in artificial intelligence. *National Science Review*.
- [16] James, W. (1890). *The Principles of Psychology*. New York: Dover.
- [17] Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- [18] Kant, I. (1781). *Kritik der reinen Vernunft* (Critique of Pure Reason).
- [19] Kautz, H. (2020). The Third AI Summer, Robert S. Engelmore Memorial Lecture. *Thirty-fourth AAAI Conference on Artificial Intelligence*, New York, NY. <https://www.cs.rochester.edu/u/kautz/talks/Kautz%20Engelmore%20Lecture.pdf>.
- [20] Lamb, Luis, Garcez, Artur, Gori, Marco, Prates, Marcelo, Avelar, Pedro, and Vardi, Moshe. (2020). Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. 4810-4817. 10.24963/ijcai.2020/670.
- [21] Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*. 8, 529-566.
- [22] Macchi, L., M. Bagassi, and R. Viale, (eds.), (2016). *Cognitive Unconscious and Human Rationality*. MIT Press, Cambridge, MA.
- [23] Marcus, G. F. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science*. MIT Press., Cambridge, MA
- [24] McClelland, J., B. McNaughton and R. O'Reilly, (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102 (3), 419-457.
- [25] Mekik, C. (2022). *Raven's Matrices in The Clarion Cognitive Architecture*. Ph.D. Dissertation, Cognitive Science Department, Rensselaer Polytechnic Institute.
- [26] Milner, D., and M. Goodale. (1995). *The Visual Brain in Action*. Oxford University Press, Oxford, UK.
- [27] Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, MA.
- [28] Pew, R. W. and A. S. Mavor (eds) (1998). *Modeling Human and Organizational Behavior: Application to Military Simulations*. National Academy Press, Washington, DC.
- [29] Pinker, Steven and Prince, Alan (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28 (1-2), 73-193.
- [30] Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*. 118 (3), 219-235.
- [31] Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 501-518.
- [32] Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85-117. 10.1016/j.neunet.2014.09.003
- [33] Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 674-685.
- [34] Sun, R. (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. John Wiley and Sons, New York.
- [35] Sun, R. (1995). Robust reasoning: Integrating rule-based and similarity-based reasoning. *Artificial Intelligence (AIJ)*. Vol.75, No.2, pp.241-296. 1995.
- [36] Sun, R. (2002a). Hybrid connectionist symbolic systems. In: M. Arbib (ed.), *Handbook of Brain Theories and Neural Networks* (2nd Edition), pp.543-547. MIT Press, Cambridge, MA. 2002.
- [37] Sun, R. (2002b). *Duality of the Mind*. Lawrence Erlbaum Associates, Mahwah, NJ.
- [38] Sun, R. (ed.), (2008). *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, New York.
- [39] Sun, R. (2012). Memory systems within a cognitive architecture. *New Ideas in Psychology*, Vol.30, pp.227-240. 2012.
- [40] Sun, R. (2015a). Interpreting psychological notions: A dual-process computational theory. *Journal of Applied Research in Memory and Cognition*, Vol.4, No.3, pp.191-196. 2015.
- [41] Sun, R. (2015b). Artificial Intelligence: Connectionist and symbolic approaches. In: J. D. Wright (ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd edition), Vol.2. pp.35-40. Elsevier, Oxford, UK. 2015.
- [42] Sun, R. (2016a). Implicit and explicit processes: Their relation, interaction, and competition. In: L. Macchi, M. Bagassi, and R. Viale (eds.), *Cognitive Unconscious and Human Rationality*. MIT Press, Cambridge, MA.
- [43] Sun, R. (2016b). *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture*. Oxford University Press, New York.
- [44] Sun, R. (2020). Full human-machine symbiosis and truly intelligent cognitive systems. *AI and Society: Journal of Knowledge, Culture and Communication*, Vol.35, No.1, pp.17-28. 2020.
- [45] Sun, R. and Alexandre, F. (eds.), (1997). *Connectionist Symbolic Integration*. Lawrence Erlbaum Associates, Hillsdale, NJ.

- [46] Sun, R. and Bookman, L. (1994). *Computational Architectures Integrating Neural and Symbolic Processes*. Kluwer Academic Publishers, Amsterdam, The Netherland.
- [47] Sun, R., P. Slusarz, and C. Terry, (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review*, 112 (1), 159-192.
- [48] Taatgen, N. and J. Anderson, (2023). Chapter 8: Constraints in Cognitive Architectures. In: *The Cambridge Handbook of Computational Cognitive Sciences*. Cambridge University Press, Cambridge, UK.
- [49] Wermter, S. and R. Sun (eds.), (2000). *Hybrid Neural Systems*. Springer-Verlag, Berlin. 2000.