Accounting for Similarity-Based Reasoning within a Cognitive Architecture

Ron Sun (rsun@rpi.edu)
Cognitive Sciences Department
Rensselaer Polytechnic Institute
Troy, NY 12180, USA

Xi Zhang (xzf73@mizzou.edu)
Department of CS
University of Missouri
Columbia, MO 65211, USA

Abstract

This work explores the importance of similarity-based processes in human everyday reasoning, beyond purely rule-based processes prevalent in AI and cognitive science. A unified framework encompassing both rule-based and similarity-based reasoning may provide explanations for a variety of human reasoning data.

The paper implements this analysis in a cognitive architecture Clarion, which has previously succeeded in capturing a variety of human learning data in simulations. The exploration of similarity-based reasoning in this architecture leads to a more complete and more comprehensive framework of human reasoning and learning. The simulation within this architecture accurately captures human reasoning data, including numerical measures and verbal protocols. This work demonstrates the significant role played by similarity-based reasoning. Furthermore, it demonstrates how such a reasoning process falls out of the existing structure in the cognitive architecture Clarion.

Introduction

What is human everyday reasoning like? Is it suitably captured by formal models developed by logicians and AI researchers? Or is it different? What are its similarities and differences to these models? After all, computationally speaking, what are the essential patterns in such reasoning?

In this paper, we will attempt to describe some data of human everyday (i.e., mundane or "commonsense") reasoning in computational terms. We will instantiate our analysis in the form of a computational model implemented in a generic cognitive architecture — Clarion (Sun 2002).

A little background is in order here. Sun (1991) proposed a theory of human everyday reasoning based on a combination of rule-based reasoning and similarity-based reasoning, implemented with a mixture of localist and distributed connectionist models. This theory was further developed and elaborated in Sun (1995). The basic tenet of this theory is that, to a significant extent, human everyday reasoning may be described by a combination of rule-based and similarity-based reasoning. Human everyday reasoning may be reduced to these two types of processes. The intermixing of rule-based and similarity-based reasoning can lead to complex patterns of inferences as commonly observed in human everyday reasoning. And these two types of processes may be captured within a unified connectionist model; that is, they fall out of the very same model (albeit with a combination of localist and distributed representations).

The theory was backed up by psychological evidence in the form of verbal protocols as in Collins (1978) and Collins and Michalski (1989). In Sun (1995), these protocols were analyzed based on two mechanisms: rules and similarity (Tversky 1977, Hahn and Chater 1998). The analysis showed that vast majority of the protocol data might be easily captured by the intermixing of these two mechanisms. This theory was crystallized into a two-component model whereby rule-based reasoning was carried out in one component with localist representation, and similarity-based reasoning in another with distributed representation (Sun 1995). Relevant to this approach, Sloman (1993) published a set of experiments, which provided support to the hypothesis of Sun (1991) (see also Sun 1995). He found that similarity played a significant role in determining outcomes of inductive reasoning and similarity might be characterized by feature overlapping (as in Sun 1991). Five years later, Sloman (1998) described further experiments that again supported the hypothesis that there were two parallel mechanisms at work in human everyday reasoning (Sun 1991).

In the remainder of this paper, we first describe the three pertinent experiments of Sloman (1998), which were consistent with the theory advanced in Sun (1991) and Sun (1995). We then describe the generic cognitive architecture, Clarion, used in capturing human everyday reasoning. Next, the particular setup of the architecture for capturing this set of human experiments is described. We then describe the results from simulating the experiments of Sloman (1998) using Clarion. Finally, some general discussion completes the paper.

The Categorical Inference Task

Let us examine some human reasoning data that illustrates combinations of similarity-based and rule-based reasoning (SBR and RBR, respectively). We will look into the data from experiments 1, 2, 4, and 5 of Sloman (1998), which are most relevant to this issue.

Among them, according to our interpretation, although experiment 1 used forced choice while experiment 2 used rating of argument strength, both involved SBR to a very significant extent. Experiment 4 involved explicit use of categorical relations, and thus mainly RBR. Experiment 5 involved more of SBR, as well as RBR.

Specifically, in experiment 1, subjects were given pairs of arguments, either in the form of *premise specificity*:

a. All flowers are susceptible to thrips. $\Longrightarrow$ All roses are susceptible to thrips.

b. All plants are susceptible to thrips. $\Longrightarrow$ All roses are susceptible to thrips.

or in the form of *inclusion similarity*:

a. All plants contain bryophytes. $\Longrightarrow$ All flowers contain bryophytes.

b. All plants contain bryophytes. $\Longrightarrow$ All mosses contain bryophytes.

Subjects were to pick the stronger of the two arguments from each pair. 73 subjects were tested and each was given 18 pairs of arguments (among other things not related to this task).

The results showed that the more similar argument from each pair of arguments was chosen 82% of times (for inclusion similarity) and 91% of times (for premise specificity). t tests showed that these percentages were significantly above chance, either by subjects ($t(72) = 18.64$ and $t(72) = 33.09$ for premise specificity and inclusion similarity, respectively; $p < 0.0001$) or by argument pairs ($t(8) = 6.97$ and $t(8) = 15.61$ respectively; $p < 0.0001$). We note that, if only RBR had been used, then similarity should not have made a difference, because the conclusion category was contained in the premise category and thus both arguments in each pair should have been equally, perfectly strong. Therefore, the data suggest that SBR was involved to a significant extent.

In experiment 2, subjects were instead asked to rate the likelihood ("conditional probability") of each argument. Ratings could range from 0 to 1. 18 subjects were tested.

The mean rating was 0.89 for inclusion similarity and 0.86 for premise specificity. Both were significantly below 1, both by subjects ($t(17) = 2.75$ and $t(17) = 3.23$ respectively; $p < 0.01$), and by arguments ($t(17) = 8.87$ and $t(17) = 6.14$ respectively; $p < 0.0001$). Again we note that it would have been the case that the outcome was 1 if only RBR had been used (because the conclusion category was contained in the premise category). Thus, SBR was significantly present here too. Indeed, ANOVA showed that across subjects, there was a significant main effect of similarity (low vs. high; $F(1, 17) = 18.90, p < 0.001$). So was the case across argument pairs ($F(1, 16) = 12.64, p < 0.001$).

In experiment 4, subjects were asked to rate the likelihood of each argument. Ratings could range from 0 to 1. However, in this case, each category inclusion relation was specifically presented as part of each argument. For example,

All plants contain bryophytes. All mosses are plants. $\Longrightarrow$ All mosses contain bryophytes.

The results showed that the mean judgment was 0.99. 23 out of 27 subjects gave all 1's. 32 out of 36 arguments received judgments of all 1's (excluding one individual who gave 0.99 throughout). In other words, the similarity-based phenomena almost disappeared. Instead, an explicit RBR mode based on category inclusion relations was used.

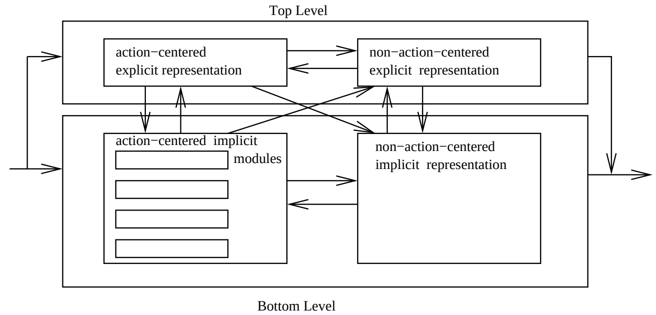

Figure 1: The CLARION architecture.

Experiment 5 was similar to experiment 2, in that ratings were obtained. However, before any ratings were done, subjects were asked to make category inclusion decisions. Thus, in this case, subjects were reminded of rule-based reasoning explicitly involving category inclusion relations. Therefore, they were more likely to use RBR, although probably not as much as in experiment 4, due to the separation of category inclusion judgment and argument likelihood rating in the experiment procedure (unlike that of experiment 4).

The results showed that no one of the 18 subjects gave a likelihood judgment of 1 for every argument, indicating SBR was probably at work. Compared with experiment 2, having subjects make category inclusion judgments increased the likelihood ratings. The mean judgment for experiment 5 was 0.92 as opposed to 0.87 for experiment 2. [1] This increase might reflect the increased involvement of RBR. Nevertheless, ANOVA showed a significant effect of similarity (low vs. high), across subjects ($F(1, 17) = 9.33, p < 0.01$), and across argument pairs ($F(1, 16) = 11.42, p < 0.01$).

Below, we will utilize this task of categorical inference for the further testing of cognitive architecture CLARION. The simulation shows indications of the significance of similarity-based reasoning (as opposed to probabilistic or Bayesian reasoning; cf. Anderson and Lebiere 1998).

The CLARION Model

CLARION is an integrative model with a dual representational structure (Sun et al 2001, Sun 2002). It consists of two levels: the top level captures *explicit* processes and the bottom level captures *implicit* processes. See Figure 1.

First, the inaccessible nature of implicit knowledge is suitably captured by subsymbolic distributed representations provided by a backpropagation network. This is because representational units in a distributed representation are capable of accomplishing tasks but are

[1] However, the difference was not statistically significant by subjects, although significant by arguments ($t(35) = 3.81, p < 0.0001$).

subsymbolic and generally not individually meaningful (see Smolensky 1988, Sun 1995). This characteristic of distributed representation accords well with the (direct) inaccessibility of implicit knowledge.

In contrast, explicit knowledge may be captured in computational modeling by a symbolic or localist representation (Clark and Karmiloff-Smith 1993), in which each unit is more easily interpretable and has a clearer conceptual meaning. This characteristic captures the property of explicit knowledge being (directly) more accessible and more manipulable (Smolensky 1988, Sun 1995).

This radical difference in the representations of the two types of knowledge leads to a two-level model whereby each level using one kind of representation captures one corresponding type of process, either implicit or explicit. The model may select to use one level or the other, based on current circumstances (e.g., experimental conditions; see Sun 2002 for details). When both levels are used, the outcome from the two levels may be combined in some ways, which may be partially domain specific (Sun 2002).

At each level of the model, there may be multiple modules, both *action-centered* modules and *non-action-centered* modules (Schacter 1990, Moscovitch and Umilta 1991). The reason for having both action-centered and non-action-centered modules (at each level) is because, as it should be obvious, action-centered knowledge (roughly, procedural knowledge) is not necessarily inaccessible (directly), and non-action-centered knowledge (roughly, declarative knowledge) is not necessarily accessible (directly). Although it was argued by some that all procedural knowledge is inaccessible directly and all declarative knowledge is directly accessible, such a clean mapping of the two dichotomies is untenable in our view. We will refer to these two sets of modules as the *action-centered subsystem* (the ACS) and the *non-action-centered subsystem* (the NACS), respectively. There are also other components, such as working memory, episodic memory, etc., which are not important to this work.

In this work, we will focus on the NACS, due to the declarative nature of the task. This subsystem, as stated earlier, consists of (1) a top level (known as the GKS, or the general knowledge store), which is made up of a set of chunks and a set of explicit associative rules linking chunks, and (2) a bottom level (known as the AMNs, or the associative memory networks), which is made up of implicit associative memories (Sun 2002).

At the top level of the NACS, the essential elements are *chunks*, each of which is specified by a set of dimension-value pairs (i.e., attribute-value pairs) that describes an entity (or an object), along with a chunk label. Each chunk is represented by a chunk node, which is linked to the nodes at the bottom level (the AMNs) representing the individual dimension-value pairs involved.

The support for the conclusion of an associative rule, which is a chunk, is calculated as follows (Sun 1994):

$$S_j^a = \sum_i S_i^c * W_i^a \qquad (1)$$

where j indicates the jth rule at the top level, S_j^a is the support for associative rule j, S_i^c is the strength of the ith chunk in the condition of the rule, i ranges over all the chunks in the condition of rule j, W_i^a is the weight of the ith chunk in the condition of rule j (which, by default, is $W_i^a = 1/n$, where n is the number of chunks in the condition of the rule).

The conclusion chunk has a strength level that is determined by the maximum of all the support from all the relevant rules:

$$S_{c_k}^c = \max_{j\text{:all associative rules leading to } c_k} S_j^a \qquad (2)$$

where $S_{c_k}^c$ is the strength of chunk c_k (resulting from associative rules), and j ranges over all the associative rules pointing to c_k.

In addition, similarity-based reasoning falls out of knowledge encoding with chunks (i.e., with sets of dimension-value pairs). A known (given or inferred) chunk is automatically compared with another chunk. If their similarity is high enough, then the other chunk is inferred. The strength of a chunk c_i as the result of similarity-based reasoning is:

$$S_{c_i}^c = \max_j (S_{c_j \sim c_i} \times S_{c_j}^c)$$

where $S_{c_j \sim c_i}$ measures the similarity from c_j to c_i (Tversky 1977), $S_{c_j \sim c_i} \times S_{c_j}^c$ measures the support to c_i from the similarity, and j ranges over all the chunks.

The default similarity measure (Sun 1995, Tversky 1977) is:

$$S_{c_1 \sim c_2} = \frac{N_{c_1 \cap c_2}}{f(N_{c_2})}$$

where $S_{c_1 \sim c_2}$ denotes the similarity from c_1 to c_2. $N_{c_1 \cap c_2}$ is the weighted sum of the identically valued dimensions in c_1 and c_2 (among all the specified dimensions of c_2 — the dimensions that have specified values). That is, $N_{c_1 \cap c_2} = \sum_{i \in c_2 \cap c_1} W_i^{c_2} \times A_i$, where A_i is the strength of the value of dimension i in chunk c_1, which is normally 1 (representing full strengths). The weights ($W_i^{c_2}$) in the weighted sum are specified with respect to c_2 (the target of similarity, not the source of it). Normally, these weights are the same and equal to 1. N_{c_2} is the weighted sum of the specified dimensions (the dimensions that have specified values) of c_2. That is, $N_{c_2} = \sum_{i \in c_2} W_i^{c_2} \times A_i$, where normally $A_i = 1$ and $W_i^{c_2} = 1$. f is a super-linear, but close to linear, function (such as $f(x) = x^{1.0001}$ as in our simulation of this task). [2] For further details, see Sun (1995).

Similarity is automatically computed whenever reasoning involves multiple chunks that are similar to one another. Therefore, there is no dedicated representation of similarity between any two chunks.

Similarity-based and rule-based reasoning can be inter-mixed. When both SBR and RBR are employed, we have:

$$S_{c_i}^c = \max(c_{14} \times \max_{j\text{:all rules leading to } c_i} S_j^a,$$

[2] Similarity is thus limited to $[0, 1)$.

$$c_{15} \times \max_{j\text{:all chunks similar to } c_i} (S_{c_j \sim c_i} \times S^c_{c_j}))$$

where c_{14} and c_{15} are two constants that balance the two measures (rule versus similarity), and $S_{c_j \sim c_i}$ is the similarity measure.

As a result of mixing SBR and RBR, complex patterns of reasoning can emerge. As explicated in Sun (1995), the conclusion from one step of reasoning can be used as the starting point of the next step. The iterative process of combined rule-based and similarity-based reasoning allows all possible conclusions to be reached (including "inheritance" reasoning; Sun 1995). These different sequences together capture essential patterns of human everyday reasoning (see Sun 1995 for details).

Note that all of the operations of the non-action-centered subsystem are under the control of the action-centered subsystem, which makes action decisions each step of the way. To do so, the top level of the ACS consists of a set of explicit action rules, either externally given or extracted from the bottom level (from implicit knowledge), while the bottom level consists of implicit decision networks (trained with reinforcement learning algorithms, negligible in this task). For details regarding the ACS and its parameters, see Sun et al (2001) and Sun (2002). We will not get into these details here, as they are not directly relevant to this work.

It is worth noting that Clarion has been successful in simulating a variety of cognitive tasks. These tasks include serial reaction time tasks, artificial grammar learning tasks, process control tasks, alphabetical arithmetic tasks, and the Tower of Hanoi task (Sun 2002, Sun and Zhang 2004). In addition, we have done extensive work on a complex minefield navigation task (Sun et al 2001, Sun and Peterson 1998). We are now in a good position to extend the effort to the capturing of a wide range of human reasoning and memory processes, through simulating reasoning and memory task data. This paper is but one aspect of this effort.

Simulation Setup

At the top level of the NACS (i.e., the GKS), all relevant category inclusion relations, such as "flowers are plants" or "mosses are plants", were encoded as associative rules. Chunk nodes in the GKS were used to represent the concepts involved, such as "flowers" and "plants". The dimensional values of these chunks were represented as separate nodes in the AMNs, and thus the chunk nodes were linked to the AMNs.

For simulating various experimental settings, the following manipulations were used: For simulating settings where SBR was dominant, RBR was de-emphasized. For simulating settings where RBR was dominant, RBR was emphasized. The relative emphasis of the two methods was accomplished through the *balancing* parameters. We set $c_{14} = 0.5$ and $c_{15} = 1.0$ for experiments 1 and 2, because of the heavy reliance on SBR as opposed to RBR as suggested by the analysis of the human data (see the earlier discussion of the human data). For simulating experiment 4, they were set at $c_{14} = 1.0, c_{15} = 1.0$, because this setting prompted more reliance on RBR as indicated by the human data. For simulating experiment 5, they were set at $c_{14} = 0.88, c_{15} = 1.0$, because the experiment involved an intermediate level of reliance on RBR as suggested by the human data. In all, these values were set in accordance with our interpretations of what happened under these different experimental conditions respectively.

At the bottom level of the NACS (the AMNs), although the associative memories were present, they were not very relevant for the performance of this task, because there was no sufficient prior training of the network with any data directly relevant to this task. [3]

Training of the model, before the simulation of the experimental test, consisted of presenting categorical features (dimension-value pairs) along with the category labels, to both levels of the NACS. The features (dimension-value pairs) captured similarities between entities. That is, if A was more similar to C than B was, then A would have more features in common with C than B would. And so on. Note that repeated presentations were not required. The one-pass presentation enabled the formation of chunks and associative rules in the GKS, but not much implicit knowledge in the AMNs. With a proper process of chunk encoding and associative rule encoding as in Clarion, one-pass presentation was sufficient for the GKS.

During test, when a category name was given, the category name was matched with a corresponding chunk label. The matching chunk was activated to the full extent (i.e., 1). Then, through associative rules as well as through similarity-based processes, conclusion chunks were also activated (to varying extents). Conclusion chunks were retrieved along with their strengths, combining SBR and RBR according to the balancing parameters.

For simulating ratings of conclusions (as in experiments 2, 4, and 5), the strengths of chunks derived from a proper combination of the results of SBR and RBR (as determined by the balancing parameters) were directly used. However, for simulating forced choices (as in experiment 1), a stochastic decision process based on the Boltzmann distribution was used to select between two possible outcomes.

Simulation Results

We simulated the data from experiments 1, 2, 4, and 5 of Sloman (1998) as described earlier. For each experiment, a set of simulation runs (i.e., simulated "subjects") equal to the number of the human subjects involved were used. The results and the statistical analysis of the results were as follows.

As described before, in experiment 1, subjects were to pick the stronger of the two arguments from each pair. The simulation of experiment 1 showed, the same as the human data, that the more similar argument from each

[3]For the associative memory network, the number of input units was 1800 (for representing all chunks specifiable with 60 dimensions of 30 possible values each), the number of hidden units was 500, and the number of output units was 1800. The learning rate was 0.2 and the momentum was 0.1.

pair of arguments was chosen more often: 82% of times (for inclusion similarity) and 83% of times (for premise specificity). t tests showed that these percentages were significantly above chance, either by subjects ($p < 0.001$) or by argument pairs ($p < 0.001$), the same as in the human data. In our simulation setup, there was a significant involvement of SBR (with $c_{14} = 0.5, c_{15} = 1.0$). If only RBR had been used, then similarity could not have made a difference, and thus both arguments in a pair should have been equally strong. This simulation demonstrated that the conjecture of the involvement of SBR in producing the human data in this experiment was a reasonable interpretation (see the earlier exposition of the human experiments), given the close match with the human data.

In experiment 2, subjects were instead asked to rate the likelihood of each argument. In this simulation, the mean rating was 0.86 for inclusion similarity and 0.87 for premise specificity. Both were significantly below 1, different from what would have been predicted if only RBR had been used, both by subjects ($p < 0.001$) and by arguments ($p < 0.001$), the same as in the human data. ANOVA also showed that across subjects and across argument pairs, there was a significant main effect of similarity (low vs. high; $p < 0.001$). With the same setup as the previous simulation, this simulation again demonstrated the same pattern of significant involvement of SBR in the human performance.

In experiment 4, subjects were asked to rate the likelihood of each argument, right after being presented relevant category inclusion relations. The simulation produced the mean judgment 0.99, exactly the same as the human data. Compared with experiment 2, explicit RBR based on category inclusion was much more prominent in this case, as specified in our simulation setup ($c_{14} = 1.0, c_{15} = 1.0$), which captured the human data accurately.

In experiment 5, ratings were obtained after subjects were asked to make category inclusion decisions. In this case, subjects were reminded of RBR involving category inclusion relations and therefore they were more likely to use RBR (compared with experiment 2), although not exclusively (unlike experiment 4). In the simulation, the mean judgment for experiment 5 was 0.91 for both inclusion similarity and premise specificity, as opposed to 0.86 and 0.87 for the two cases in experiment 2. ANOVA also showed a significant main effect of similarity (low vs. high), across subjects ($p < 0.001$), and across argument pairs ($p < 0.001$). This simulation replicated the human data well, which showed that our interpretation as embodied in the simulation setup ($c_{14} = 0.88, c_{15} = 1.0$), that is, less involvement of RBR compared with experiment 4 but more compared with experiment 2, was a reasonable one.

In all, simulation of this task successfully validated the interpretation and the analysis of human performance in this task and, to some extent, our framework in general.

Concluding Remarks

Overall, the simulation accurately captured the human reasoning data from Sloman (1998). The simulation was conducted based on our framework of mixed rule-based reasoning and similarity-based reasoning, which, along with other simulations published elsewhere (e.g., Sun 1995, 2002, Sun et al 2001, Sun and Zhang 2004), showed the cognitive plausibility of the CLARION architecture to some extent.

This simulation demonstrates the importance of similarity-based reasoning in human everyday reasoning. This similarity-based process is quite distinct from probabilistic reasoning as implemented in other existing cognitive architectures, such as ACT-R (see Anderson 1993 or Anderson and Lebiere 1998). Let us compare the two different approaches. ACT-R as described in Anderson and Lebiere (1998) tries to capture all inferences in a probabilistic framework. In so doing, it lumps together all forms of weak inferential connections in a unified way. Although this approach leads to uniformity, it has shortcomings as well. All similarity relations between any pair of any two objects must be explicitly represented with all the associated parameters, which specify probabilistic computation used to capture similarity-based reasoning (along with other inexact inferences). The problem is the complexity of representing all similarity pairings. This complexity is very high in ACT-R but in contrast is avoided in CLARION.

The limitations of probabilistic reasoning (Pearl 1988) in general include its neglect of many heuristics, simplifications, and rules of thumb (Tversky and Kahneman 1983, Sun 1995, Yang and Johnson-Laird 2001) useful in reducing the computational complexity of formal mathematical models. As a result, it suffers from higher computational complexity (Sun 1995).

We should also look into the framework of Collins and Michalski (1989), which apparently incorporated "similarity-based" reasoning through explicitly representing similarity in a complex logical formalism. Similarity was explicitly represented as a logical operator: That is, for almost any pair of any two objects, there would be a logical relation explicitly represented, denoting their similarity. Inferences could be performed on the basis of similarity operators, using a search process. The complexity of this representational framework was extremely high.

In general, logic-based models suffer from a number of well known shortcomings, including their restrictiveness concerning pre-conditions, consistency, and correctness, and their inability in dealing with inexactness (see, e.g., Israel 1987, Sun 1995). Their restrictiveness renders such models costly, difficult to specify, and difficult to use.

In a different vein, psychological work on reasoning is relevant also. Such work mostly centers around either mental logic (Rips 1994, Braine and O'Brien 1998) or mental models (Yang and Johnson-Laird 2001). Neither approach deals with similarity-based reasoning as captured in CLARION. Their focuses are elsewhere.

In sum, this line of work, combining similarity-based reasoning and rule-based reasoning (Sun 1995, Sloman

1998, Hahn and Chater 1998), offers a new approach for capturing some essential patterns of human everyday reasoning (albeit not all patterns of human reasoning). It complements logic-based "commonsense" reasoning models prevalent in AI, which is very much centered on logic and thus limited by logic. This work also points to new avenues of cognitive modeling, beyond the current psychology of reasoning (which largely focuses on various logics and mental models) and beyond existing cognitive architectures (Anderson and Lebiere 1998). In addition, this approach may well be extended to case-based and/or analogical reasoning (e.g., Sun 1995a).

Acknowledgment

This work has been supported in part by Army Research Institute contract DASW01-00-K-0012 to Ron Sun and Bob Mathews.

References

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

J. Anderson and C. Lebiere, (1998). *The Atomic Components of Thought*, Lawrence Erlbaum Associates, Mahwah, NJ.

M. Braine and D. O'Brien (eds.), (1998). *Mental Logic.* Lawrence Erlbaum Associates, Mahwah, NJ.

A. Clark and A. Karmiloff-Smith, (1993). The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind and Language.* 8 (4), 487-519.

A. Collins, (1978). Fragments of a theory of human plausible reasoning. In: D.Waltz (ed.), *Theoretical Issues in Natural Language Processing II*, 194-201. Ablex, Norword, NJ.

A. Collins and R. Michalski, (1989). The logic of plausible reasoning. *Cognitive Science*, 13(1), 1-49.

E. Davis, (1990). *Representations of Commonsense Knowledge.* Morgan Kaufman, San Mateo, CA.

U. Hahn and N. Chater, (1998). Similarity and rules: distinct? exhaustive? empirically distinguishable? *Cognition*, 65, 197-230.

D. Israel, (1987). What's wrong with non-monotonic logic? In: Ginsberg (ed.), *Readings in Non-monotonic Reasoning*, pp.53-55, Morgan Kaufman, San Mateo, CA.

M. Moscovitch and C. Umilta, (1991). Conscious and unconscious aspects of memory. In: *Perspectives on Cognitive Neuroscience.* Oxford University Press, New York.

J. Pearl, (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, San Mateo, CA.

L. Rips, (1994). *The Psychology of Proof.* MIT Press, Cambridge, MA.

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

S. Sloman, (1993). Feature based induction. *Cognitive Psychology*, 25, 231-280.

S. Sloman, (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, 35, 1-33

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

R. Sun, (1991). Connectionist models of rule-based reasoning. *Proceedings of the 13th Cognitive Science Conference*, pp.437-442. Lawrence Erlbaum Associates, Hillsdale, NJ.

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

R. Sun, (1995a). A microfeature based approach toward metaphor interpretation. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95).* Montreal, Canada. pp.424-430, Morgan Kaufmann, San Francisco, CA.

R. Sun, (2002). *Duality of the Mind.* Lawrence Erlbaum Associates, Mahwah, NJ.

R. Sun, E. Merrill, and T. Peterson, (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science.* Vol.25, No.2, 203-244.

R. Sun and T. Peterson, (1998). Autonomous learning of sequential tasks: experiments and analyses. *IEEE Transactions on Neural Networks*, Vol.9, No.6, pp.1217-1234.

R. Sun and X. Zhang, (2004). Top-down versus bottom-up learning in cognitive skill acquisition. *Cognitive Systems Research*, Vol.5, No.1, pp.63-89.

Y. Yang and P. Johnson-Laird, (2001). Mental models and logical reasoning problems in the GRE. *Journal of Experimental Psychology: Applied*, 7 (4), 308-316.

A. Tversky, (1977). Features of similarity. *Psychological Review*, 84(4), 327-352.

A. Tversky and D. Kahneman, (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 439-450.