The Emergence of Cognitive Phenomena from Sub-symbolic Processes

David E. Rumelhart
Institute for Cognitive Science
University of California, San Diego

Discussion of cognition, especially of language and thought often revolves arouind a discussion of the *rules of language* and the *rules of thought*. The former, namely the rules of language, we often call the grammar of the language – in part I suppose by analogy with the grammar rules we all learn in school. The latter, the rules of thought, we often call natural logic or just plain logic – again I suppose by analogy with the rules of logic we learn in school. These rules that people have in mind are normally expressed as relations among cognitive elements – relationships among NP's and VP's or premises and conclusions. Moreover, in many models of language and thought, these rules play an operative role, that is they are interpreted by some interpreter and decisions are made on the basis of the rules themselves. There are obvious reasons why it is tempting to think of language and thought in just this way – as a set of rules we follow while producing or interpreting what we see or hear in the case of language and as the set "of rules of inference" we employ in reasoning. Language is not haphazard. In English words *must* be used in certain orders, certain inflections *must* be used to signal certain meanings (the s for plural, the -ed for past etc.) and certain words *must* be used for certain meanings. Most importantly, most sentences are *entirely novel* – they are obviously *generated*, not stored and recalled again. On this analysis, explict rules would seem to be a very parsimonious explanation of the facts. This suggests that language learning involves the abstraction of a set of such rules and that language production and comprehension involves the application of these *general rules* to the situation at hand. Similarly, in the case of reasoning, it would appear that the most parsimonious explanation of our ability to produce and interpret arguments is that we have somehow abstracted from our experience a set of rules of inference which are then interpreted in any given situation.

As neat as these accounts have seemed, there are serious problems with them. There are characteristic flaws in our reasoning – sometimes we don't follow the rules. Similarly, language is *full* of exceptions to the rules – cases where the general rule doesn't seem to apply. This includes straight forward cases like the fact that rang is the past of

ring, not ringed as we might have thought to more interesting examples such as the fact that "cold" means something rather different in the phrase "cold person" than it does in the phrase "cold water." Within the framework of the *rule account* we must proliferate rules, differentiate between rule governed and non-rule governed cases or distinguish between competence conditions and performance conditions.

It has seemed to me for some years now that the "explicit rule" account of language and thought was *wrong*. It has seemed that there must be a *unified* account in which the so-called *rule-governed* and *exceptional* cases were dealt with by a unified underlying process -- a process which produces rule-like and rule exception behavior through the application of a single process. On this account, the rules that we analysts discover are more in our analysis than in the heads of our subjects. In short, I have come to believe that the explanation of human cognition by appealing to the interpretation of rules stated at the symbolic level is not, in general, going to work.

I have instead become very interested in a much different conception of human cognitive processing -- a system in which cognitive performance is not produced by the processing of symbolic rules, but one in which both the rule like and non-rule-like behavior is a product of the interaction of a very large number of "sub-symbolic" processes. In this sense the rule-like behavior is seen to "emerge" from these interactions rather than to have the processor in any sense "interpret" the rules at hand. This view has been motivated by two very different concerns -- on the one hand, I have been increasingly disillusioned with attempts to formulate an adequate set of explanations at the rule level. The more I learn about the way language and thought proceeds the less it seems like an application of general abstract rules. On the other hand, I have become increasingly impressed with the power of what we have come to call *Parallel Distributed Processing* (PDP) systems as an alternative to the more conventional accounts.

Parallel distributed processing is my short-hand for *brain-like* or *neurally inspired* processing systems. I am convinced that brains process information in ways fundamentally different from conventional digital computers. Whereas modern computers are capable of carrying out

serial operations in 10s of nanoseconds, brains carry out their operations in times measured in the milliseconds -- brain units seem to process information 100,000 times slower than computers!! Yet, even our best artificial intelligence systems cannot come close to matching brains on simple tasks like recognizing a spoken word or catching a ball.

What then is the brain's advantage? I suspect that this lies in the kind of computation the brain is able to carry out. Primarily, the brain succeeds because is has an enormous number of processing units all working in parallel and cooperatively *settling* into a solution -- rather than *calculating* a solution. Processing is done by cooperating coalitions of independent units each working on the information made available to it. It is as if computation were done by having each little processor carry out its small computation and then *vote* on the answer to the question. Solutions are reached by majority rule, or by reaching a compromise. There is no central processor, rather a highly distributed set of units whose combined activity pushes the whole system toward an action. This is a very different view that that implicit in the symbolic rule oriented processesing systems. There is no interpreter, there is no place in the system where the rules abid. This does not deny the existence and importantce of symbols. Symbols themselves are emergent properties of the interactions of such a set of processing units. Symbols are not, however, processed. The processing occurs at a sub-symbolic level. Regularities at the symbolic level occur and we can write descriptions of those regularities, but whenever we formulate a rule at that level we must recognize that the rule (or law) is not interpreted by the system any more that a ball flying through the air computes the differential equations which describe its behavior. It is simly a description of the system at the symbolic level.

The research program I have been carrying out in conjunction with James McClelland and several other colleagues has been to show how the cooperative interactions among many of these sub-symbolic processing units can account for the regularities which have led to the postulatation of specific rules and which, at the same time, can allow us to account for phenomena which are difficult for an explicitly rule based account. McClelland and I have produced models in two cases which I believe offer a general paradigm for this sort of information processing system. We have

shown how a simple activation model of word perception can behave as if it knows the rules of English orthography and we have shown how a simple associative memory system can mimic the aquisition of past tense verb morphology without explicitly distinguishing between regular and exception verbs. These are two examples of a large number of cases that we have been investigating. In both of these cases the system is generative, but in neither case is their an interpreter or anything that could be construed as an explicit representation of a rule.