I have a cautious, even grudging, appreciation for the ideas of Sigmund Freud. Though I am dispositionally inclined to dismiss Freud's ideas wholesale, I have come to credit him with some genuine insights, and to be curious about the automatic hostility that arose in me when first I encountered his ideas. In contrast, I was smitten from the start with Dan Sperber & Deirdre Wilson's (1995) Relevance theory, and especially the Cognitive principle of relevance (hence, CPR). The CPR states (I paraphrase) that ceteris paribus inferences are processed in an order determined by the size of their cognitive effects, from greatest to least. From the start I found this principle elegant and an enormously powerful analytic tool. But just as, with time, I came to some appreciation of Freud's ideas, so I came also to doubt the CPR.

My doubts focused on two issues in particular: the mechanism by which the relevance of an inference was assessed and the empirical tractability of Sperber & Wilson's theory. The problem of the mechanism could be summed up thus: how can the mind, when prioritizing two inferences, know which of them will lead to greater cognitive effects without actually processing them? The problem of empirical tractability was simply whether there could be any real empirical test of such a high-level theory. Was there any real reason to believe it?

Both questions were answered for me when I connected the CPR with John Holland's discussion of adaptation in classifier systems. Classifier systems are characterized by a set of inference rules (more strictly, IF-THEN statements defining state transitions) that define the behavior of the system in response to inputs from outside. In Holland's work (1992; but see also 1995), he envisions classifier systems embedded in adaptive agents that interact with their environment, which includes other similar agents. Classifier systems become adaptive when the rules receive feedback as to whether they contribute, however indirectly, to successful behavior by the agent. Holland assigns each rule a strength property to summarize its history of effectiveness, and competing rules are processed partly in order of decreasing strength. (The other consideration that determines their processing order is specificity of match between the rule's IF clause and environmental inputs.) Holland goes on to discuss other sorts of adaptive mechanisms, but strength is the critical one for the present discussion.

Any system characterized by competing rules faces the credit assignment problem, that is, the problem of passing credit for an outcome back to all the rules that contributed to the outcome in a way that is proportionate to that contribution. (The generalized delta rule handles this in feed-forward connectionist systems, but at the cost of requiring that credit assignment be handled by a distinct backward-looking process.) Holland solves this problem with what he calls the "bucket brigade algorithm," wherein each rule passes some of its strength back to its antecedents in a kind of bid for processing priority. In this way, credit assignment is handled as part of the regular sequential workings of the agent.

In this kind of system, a rule's processing priority grows in proportion to its history of contributions to successful behavioral outcomes. And this is, I think, precisely the sort of process required to implement the CPR: a rule's strength, in effect, sums its (history of) contributions to cognitive/behavioral effects. Thus Holland's classifier systems involve a mechanism that approximates the CPR in its operation. Not only is there a mechanism that approximates the CPR, but it does so simply as a consequence of its fundamental, adaptive architecture.

Two limitations to the approximation must be noted. First, the CPR requires that inferences be prioritized according to their actual relevance, not according to their history of relevance. Classifier systems use a rule's history of contributions to successful outcomes as a proxy for their future contribution to successful outcomes. This is a difference, but one that does not, so far as I can see, run afoul of anything important in the CPR. Second, a rule's strength (and thus its processing

priority) is determined partly by the specificity of its contribution to successful outcomes, and there is no analogy to this in the CPR, though I believe one could and probably should be developed, as otherwise a processor heavily favors very general inferences over specific ones, with the result that it incurs all the limitations of a general problem solving mechanism rather than a collection of specific problem-solving strategies.

If this approximation is admitted, then there still remains the question whether there is ground for believing the CPR. Although I am aware of some of the empirical work on it, I never feel quite sure that this work really shows the CPR, as opposed to something else, at work. I feel about this research like I feel about test of psychoanalytic theory: it doesn't disconfirm the hypothesis, but neither does it seem like strong support for such general claims. This is a general problem for all high level theories.

I think there are two arguments for accepting the CPR, one weak and one strong. The weak reason is the argument that human mental processes seem to lend themselves to description in terms of classifier systems, and that at least some cognitive dynamics could be understood as the result of a classifier-system architecture. We might call this the resemblance argument, and although I find it persuasive on the whole, certainly there is room for doubt. The stronger reason has to do with a property of classifier systems, computational completeness. Computational completeness means that classifier systems can simulate the behavior of any other computational system. If the human mind involves computational systems-which it certainly does, among other things-then the behavior of these systems can be mimicked by a classifier system. If its behavior can be mimicked by a system involving an approximation of the CPR, then it is as if human behavior were guided by the CPR. This latter argument may be rephrased that the CPR might as well be correct.

These considerations have persuaded me of the CPR enough to go ahead confidently using it as an analytic tool. Its application to particular situations is still fraught with difficulty, of course, but I feel that at least the tool is sound. This argument has been rambling around in my head since early 1997, and it seems sound to me, but I invite your comments and corrections.

References

Holland, John H. (1992). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence (1st MIT Press ed.). Cambridge, Mass.: MIT Press.

Holland, John H. (1995). Hidden order: How adaptation builds complexity. Reading, Mass.: Addison-Wesley.

Sperber, Dan, & Wilson, Deirdre (1995). Relevance: Communication and cognition. Oxford, UK ; Cambridge, Mass. :: Blackwell.