# Strategies for Hypothesis-Driven Recognition in Rule-Based Systems

**B. Rode**

Cycorp
3721 Executive Center Drive
Austin, Tx.
ph. 512-342-4022
fax: 512-342-4040
ben@cyc.com

**Abstract**

*Recognition* continues to be one of the most difficult problems for systems utilizing rule-based reasoning, primarily because of the inadequacy of classical deduction for dealing with the problem and the likelihood of typological mismatch between pattern and data. Following a brief overview of the issues, this paper outlines a hypothesis-driven model for the recognition process that addresses some of the technical challenges while identifying several strategies for implementation based on data characteristics. The conclusion situates the work within a broader biological context.

## I. Introduction

Of all of the problems confronting automated reasoning and artificial intelligence, *recognition* has proved to be one of the most challenging. Not only is it still an open question how to do it, but the term itself has proved curiously intractable of formal definition. Although this is due in part to the fact that the problem cannot be treated as one of deduction, and to the parallel difficulty of scaling probabilistic methods to useful domains, there also remains the problem of formally specifying what recognition really *means* in a way that takes account of behavioral response. The intuition is that it is not enough to define recognizing that an instance of type $\varphi$ exists as concluding that $\exists x \varphi x$, and that the question of how probabilistic or defeasible the inference was is somewhat beside the point. The question, so to speak, is what happens *after the conclusion is drawn*. In the realm of behavioral neuroscience, recognition is seen as being intimately linked to physiological and behavioral response, to a point where it is impossible to draw a single definitive division between afferent perceptual signaling and efferent physiological response (Guillery, 2003), and where more sophisticated signal processing and comprehension is widely acknowledged to derive from the evolutionary arms race proceeding from the signal-detection capabilities of lower organisms. The difficulty is compounded by the strong presentiment deriving from research in cognitive science that recognition and deliberative reasoning are intimately linked, an intuition summarized by Hofstadter's slogan that cognition is recognition (Hofstadter, 1995), The close affiliation between perception and cognition on the one and, and between high-level perception and low-level efferent response on the other, poses certain difficulties for automated recognition systems, where the historical tendency has been to put the focus on signal detection, and where the identifier or identifiers of the type to be detected are often taken as a given. Indeed, insofar as these systems are representational and rule-based – and regardless of whether the rules are logical or probabilistic – one may fairly ask the question of where the 'atomic' predicates used by the representation scheme come from, how and in virtue of what they acquire their extensions and semantics, and how the terms denoting them work as such in the context of the system.

The definitional problem is compounded by the fact that the cognitive mechanisms involved in recognition cannot be purely deductive. At some level this has been known ever since the beginning of the early modern period of philosophy, insofar as it is precisely the criticism that the early British empiricists directed against the Cartesian rationalists who sought some level of empirical experience that might serve as a basis for deductive certainty. Many facts *follow* deductively from the fact that a certain concept is instantiated, but it seems that no single fact is logically sufficient for concluding *that* the concept in question is instantiated, or at least, any fact that is thus sufficient would seem to be an infinitary conjunction. The lesson commonly drawn is that recognition cannot involve *just* quantified knowledge and logically sound inference from input data. What else may be involved is a matter of some debate, but the possibilities most often considered in the literature of the last three decades are defeasible or non-monotonic reasoning, abductive reasoning, and probabilistic reasoning.

All of these, insofar as they involve introduction into the reasoning process of elements that are not arrived at deductively, point to the singular importance of hypotheses for the recognition problem, where by *hypothesis* is meant any propositional element in the cognitive process that a reasoner introduces by other-than-deductive means from input: that is to say, by means of rules of reasoning that are not strictly guaranteed to preserve truth. So central to recognition are the processes of hypothesis generation and management that it is not exaggerating to say they *are* the recognition problem. The main objective of this paper will be in part to argue for the viability of a representational approach to recognition, based on a combination of deductive and abductive reasoning with an emphasis on hypothesis generation and confirmation, but it will also suggest that the solution to the riddle of 'natural' denotation is to be found in hypothesis-driven systems that close the loop between hypothesis confirmation and background knowledge by means of learning. The prospect will also be raised that a perspective which views recognition as hypothesis-driven allows us to deal with recognition as fundamentally ratiocinative, not only in the case of 'higher level' processes that readers may be accustomed to associate with reasoning, but also with respect to 'low-level' instances wherein data from multiple sensors is 'fused' to generate a coherent result.

## II. Data and Knowledge

Apart from the aforementioned problem of defining recognition itself, understanding how to define the term *pattern* is possibly the most vexatious question confronting recognition-related research,, a problem which is belied by the term's ubiquity of use. Everyone, it seems, knows what a pattern is, but saying what one is, even colloquially, is another matter. It is not, for example, particularly helpful from the standpoint of analysis to say that a pattern facilitates recognition, or even that a pattern is necessary for recognition; although these things are true of patterns, they are true of many other things besides. Nor is it particularly helpful to employ terms like 'template', which mean different things to different people, and which may carry technical baggage, unless one is prepared to explain precisely what one means by them. And yet, to the extent that it *is* true that patterns are necessary to recognition, understanding them is indispensable from the standpoint of our investigation.

Let us begin with the familiar concepts of *instance* and *type*. Patterns, as the term is being defined here, are not types, although they are closely related to them. A pattern, rather, is a specification that allows, either for an instance of a given type to be *produced* (e.g., the pattern employed by a Jacquard Loom) or for an instance of a given type to be recognized as such, so that appropriate action may be taken. In some cases, the same pattern may do for both. The use of the term 'specification' is highly instructive in this respect, and the example of the loom even more so: in fine, there is no reason for distinguishing what is being described here from a performance specification for a machine with numerically encodable inputs and outputs. Patterns, in other words, are identical with a class of Turing Machine specification, where the machine in question is functioning in a context according to which its output is interpreted either as systematic manufacture of instances of a given type or (the case which concerns us) systematic recognition of instances of a given type. Indeed, because recognition is hypothesis driven, the use of patterns in *recognizing* instantiation of a given type is not wholly separable from using them for producing instances: to the extent that a hypothesis replicates relevant features of the instances it is used to recognize, hypothesis generation constitutes a form of production. This view, incidentally, is strongly corroborated by converging lines of evidence deriving from the past thirty years of clinical neuroscience, which strongly suggest that perception is fundamentally a proactive and hypothesis driven process with many mutual processing dependencies and 're-entrant' feedback loops obtaining between the efferent and afferent regimes (Creutzfeld, 1977), (Guillery, 2003).

Patterns may be encoded in any number of ways, and we may, in principle, distinguish components of patterns that are of particular interest to us – for example, in the case of a rule-based reasoner, we might be particularly concerned with those components of the specification that encode quantified, i.e., typological knowledge about the world. But in any case, it is knowledge that is being encoded, which brings us to the key distinction between instance-level data and quantified knowledge. *Data* represents propositional information about instances, like the fact that a car is in a garage at a certain time, and is what recognition systems operate on, in the form of primary or medial input. *Knowledge* represents propositional information quantifying over a domain of instances and stating logical dependencies between types. Relevance of the types referenced is determined by the purpose of the recognition system, but because encoding in syntactic structures can never fully account for the semantic richness of the world, encoding is limited in ways that insure that deductive inference alone will be insufficient to achieve recognition in most cases. There are two particularly pervasive reasons for this: first, because of complications owing to conditional representation, and second, because of what might be termed the *typological divide* between data and knowledge.

## III. Directionality and The Typological Divide

If first order predicate calculus is employed for representing the world knowledge needed for recognition in a particular domain, a number of sentences of the form $\forall\sigma\ (\varphi(\sigma)\rightarrow\psi(\sigma))$ will be used, where $\sigma$ is a sequence of variables, $\varphi(\sigma)$ and $\psi(\sigma)$ are open formulas featuring the same variables, and $\rightarrow$ is the material conditional. This last introduces a number of complications, the most familiar of which—the so-called paradoxes of the material conditional—to a certain extent mask the problems that are of greatest concern for automated recognition. The obvious problem and the most notorious challenge posed for common sense reasoning in general is that the truth conditions of $\rightarrow$ render the operator *monotonic*: that is, if it is the case that $\varphi\rightarrow\psi$, then $(\varphi\wedge\chi)\rightarrow\psi$ is also true, where $\chi$ could be any sentence whatever. This doesn't agree with colloquial or heuristic usage of the 'if…then' formulation, where, for example, if someone claims that drinking a cup of coffee before work will keep them alert through the morning, this doesn't mean that if they drink a cup of coffee before work and get hit by a truck while stepping off the curb, it will keep them alert all morning. Various well-known proposals for doing an end-run around monotonicity have been put forward over the years, most involving explicit circumscriptive assumptions, or else the introduction of 'non-monotonic' logical operators that engage implicit circumscriptive assumptions and allow for retraction of conclusions that have already been drawn as more information becomes available. This paper does not propose to come down in favor of any one of these approaches, but merely to point out that every one of them is at best a partial solution to the recognition problem, so long as it features a preferred *directionality* of inference. Thus, for example, if what I know is that every instance of making spaghetti pesto includes boiling spaghetti, then, it doesn't really matter whether the conditional used to represent this is material or defeasible: if what I *observe* is somebody boiling spagetti. Getting from that observation to the conclusion, however tentative, that somebody is making spaghetti pesto can't simply be a matter of inference, given the way the original rule was represented. What it is, strictly speaking, is the logical fallacy of affirmation of the consequent, but used virtuously (that is, probabilistically) as a form of abductive reasoning. This leads to the familiar problem that many hypotheses are sufficient (materially or probabilistically/defeasibly) for a particular conclusion, and even when one has eliminated the logically, physically, and pragmatically impossible, what remains may be so multifarious and mutually inclusive as to present a combinatoric challenge. But beyond the significant combinatoric issues attaching to abduction, the fact remains that *what* gets abduced as opposed to deductively concluded will depend on the *direction* of rules in the knowledge base. While this may be decided for the knowledge enterer in eight cases out of ten, there still remains a potential grey area of potentially useful features $\varphi$ and $\psi$, where an instance's having $\varphi$ will be a strong indicator of its having $\psi$, without being perfectly sufficient for it, and vice versa: for example, the property of having sickle cell trait and the property of having ancestors from a certain region of the African subcontinent. The fact it is possible to find a specialization $\varphi'$ of $\varphi$, or a generalization $\psi''$ of $\psi$, such that $\varphi' \rightarrow \psi$ or $\varphi \rightarrow \psi''$ may justifiably and feasibly be asserted; or a generalization $\varphi''$ of $\varphi$, or a specialization $\psi'$ of $\psi$, such that $\psi' \rightarrow \varphi$ or $\psi \rightarrow \varphi''$ may justifiably and feasibly be asserted, may or may not be a comfort, depending upon what needs to be hypothesized when.

The problems of abductive hypothesis generation are compounded by the near-inevitability of a typological mismatch between data and pattern. By this, I mean that, barring deliberate design or incredible luck, the types likely to be referenced in the data schema are not likely to overlap with, or generalize, or specialize, the types referenced in the pattern. Historically, there has been less awareness of this problem because many of the use cases in current circulation, such as Henry Kautz's otherwise-excellent examples (Kautz 1991), have been constructed in ways that minimize or obviate it. In real life, the problem is likely to be chronic. So, for example, suppose we are utilizing a contract murder pattern whose explicit 'scenes' include hiring a hitman, payment, stalking, and murder, while attempting to recognize contract murders in data which includes things like banking records and maybe also the contents of police procedurals. Because the persons preparing the database have not anticipated the use to which the data is being put, and because their knowledge is limited, it is really hoping for too much to expect the relevant schemas to include types like 'payment', let alone types like 'payment to hitman' or 'stalking', which, if they *were* in the database, would constitute evidence, if not conclusive proof, that the prerequisite task of recognizing contract kills had already been performed. Thus, compounding the logical problem of hypothesis generation and management, we have the pragmatic problem that much of what is hypothesized may only be an indirect match for what is actually in the data, requiring further probabilistic and causal reasoning to bridge the gap. Moreover the extent and nature of the mismatch will vary dramatically with the data topic and the ontological choices made in setting up the schema. As we shall see, this can have a significant impact on the question of what strategy is appropriate for hypothesis generation and management.

## IV. Methods of Hypothesis and the Importance of Data Characterization

So far as a recognition process model is concerned, the foregoing considerations regarding the relative expense of hypothesis management and the inevitability of typological mismatch point to a pro-active inference process which begins with hypothesization of an instance of some category of interest, followed by a judicious admixture of reasoning methods designed to elaborate this hypothesis and match this elaboration to known elements in a body of suitably structured data. Regarding the available methodologies we must certainly include the standard methods of forward-chaining (propogate the deductive closures of a selected rule or set or rules prior to querying) and backward chaining (derive and store results on an as-needed basis in responding to a query), and in addition to these, some form of abductive inference featuring constraint-based pruning of results. In addition to these, to cope with the typological mismatch problem, it's reasonable to assume that something more free-wheeling than any of them will be required: something which on the one hand is deductively tolerant, casting a very broad net, but which at the same time features a quantitative metric for tracking the complexity of the method by which its results were arrived at, the assumption being that uncertainty scales in proportion with complexity. We might, for example, attempt to model various causal or situational dependencies at the type level in our ontology—the fact that wire transfers are often a means of implementing payments, for instance—and then try chaining known data elements to hypothesized elements based on conformity with these dependencies, with uncertainty increasing based on the length of the chain and the complexity of the generalization-based reasoning needed to arrive at it. We might also imagine attempting to incorporate some level of machine learning into this, so that past successes and/or failures could be leveraged in evaluating candidate linkages.

These inference methodologies may plausibly be combined in various recognition strategies wherein the choice of method is determined in large measure by the nature or the typological divide between knowledge and data. Where we expect this to be large, it is plausible to imagine starting with an initial seed hypothesis and using forward inference to elaborate this as much as possible, thereby subsuming the expense associated with forward reasoning and controlled hypothesizing in a precompiled step. Redaction algorithms then attempt to identify data elements whose definition exactly matches or entails the definitions of hypothesized elements, merging elements as appropriate. Once this is done, probabilistic reasoning attempts to match unmerged elements to footprints or indicators in the available data (fig. 1). The more data elements are merged or matched to the hypothesis, the more strongly validated it is considered to be.

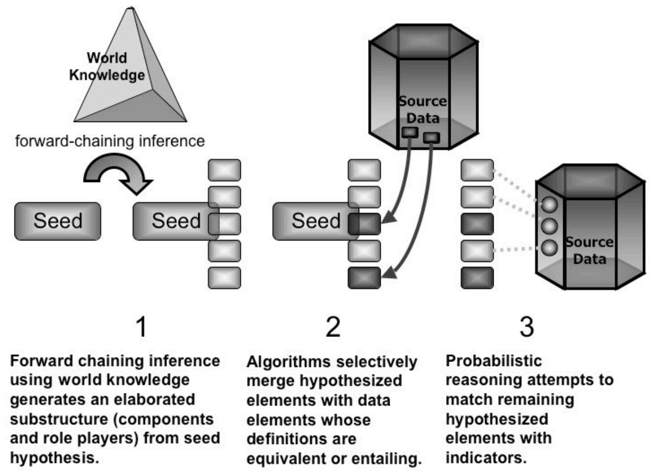


*fig. 1. Seed hypothesis is elaborated through forward propagation, with validation by data merge and subsequent indicator search.*

Conversely, in cases where there is an expectation of a tight typological match between pattern and data, a more conservative strategy suggests itself. Again, we begin with a seed hypothesis, but here, instead of elaborating this by forward means, we first use backward chaining queries and pattern-encoded world knowledge to try to identify data elements that are plausible components of the hypothesized entity, effectively sidestepping the expense associated with forward elaboration. Assuming the system is able to arrive at what is deemed a reasonable set of components, further querying against world knowledge can proceed to identify missing elements, which in turn are hypothesized, with probabilistic reasoning done in followup to link these hypothesized elements to their own data footprints, assuming these exist (fig. 2). This strategy depends on knowing in advance that the typological divide between pattern and data is small enough that active elaboration of the seed hypothesis is not a warranted expense, although it should be added that the option of recursively applying elaboration to secondary hypotheses remains, here, together with the challenges that implies.

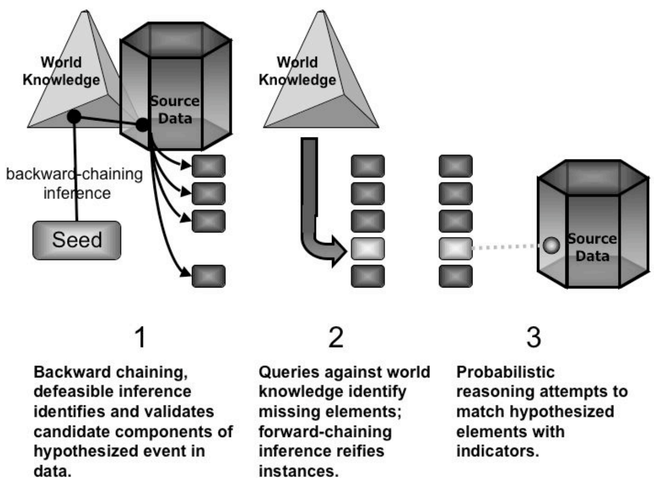


*fig. 2. Candidate components of the seed hypothesis are gathered through backward inference, with subsequent hypotheisizing of missing elements through forward reasoning, with subsequent validation by indicator search.*

Finally, we can imagine a somewhat more open-ended (and therefore riskier) version of this, in which the backward inferences used to collect candidate data elements utilize opportunistic abduction to generate 'missing' components which the system then attempts to link to data by indicator search (fig. 3).

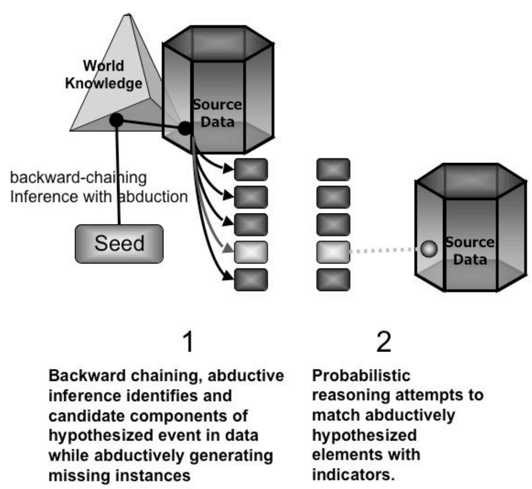


*fig. 3. Candidate components of the seed hypothesis are gathered through backward abductive inference with opportunistic hypothesizing of missing elements.*

Of course, the difficulties attaching to all of these approaches are significant, and must not be understated. To the extent that a certain measure of abductive hypothesis generation is implicated in all of them, the most obvious problem concerns combinatorics. Realistically, we must assume that in any given case of recognition, multiple hypotheses are entertained in parallel, each of which turn can be elaborated in various ways depending on the wealth of world knowledge and available data: a situation which will only be exacerbated by the presence of noise and clutter at the data level. At a minimum, this indicates a need for aggressive constraint-based hypothesis pruning. Assuming that the patterns which are employed in recognition are subsumed within a coherent ontology of world knowledge, the issue of relevance reasoning also comes into play, as a means of up-front control of hypothesis generation through pre-selection and ranking of the types on which to abduce, as well as in the form of focus for forward inference (knowing where to cut off consideration of knowledge for forward propagation), and in the form of maximizing efficiency in backward chaining – and it is also advisable that he range of initial 'seed' hypotheses be constrained in some way by the nature and topic of the use case. Such considerations clearly bespeak the need for a quantitative means of dealing with relevance. Provided a sufficiently rich ontology is presupposed, a promising standard which presents itself is an ontological distance metric, or better still, a suite of metrics which, starting from the type of the seed hypothesis, would prioritize types which were candidates for instantiation and elaboration according to their connectivity with the starting type, measured via linkage through the shared rules of the knowledge base. In addition, given that the selection of recognition strategy is to be determined by the severity and character of the typological divide between data and knowledge, it follows that there must be some way of automatically assessing and declaratively summarizing this feature in a way that can make a difference at the strategic level.

A final point that should be made in connection with the matter of strategy concerns the importance of feedback and learning. The process of corroboration that proceeds under the guidance of one of the strategies described here may succeed or fail, and it is entirely reasonable to suppose that the record of the systems past failures and successes should be used in tuning behavior at multiple levels. To begin with, all of the strategies mentioned indicate adjustable parameters: most notably the extent to which the elaboration process and hypothesis generation are allowed to go forward. But besides this, another factor which can and should be influenced by experience is choice of strategy based on data characteristics and use case. And finally, there is the question of modification of the knowledge base that serves as the foundation for the corroborative process, through the editing of rules, through inductive learning of new rules, and through creation of new constants, which alone can address the issue of rule directionality that was touched on in section III. Although much of this is uncharted territory, its importance can scarcely be understated, inasmuch as many of the issues concerning optimization and adjustment of the recognition process are both empirical in nature, and liable to be highly contextualized. To put the point another way, engineering on the basis of *a priori* reasoning can only take one so far, and sooner than put the entire burden of empirical discovery on the engineer, it is better, and perhaps mandatory, that some of it be assumed within the functionality of the system.

## V. The Lessons of Biology

Although this paper will not explore the ramifications of biology for the theory of recognition presented here in any great detail, there are certain fairly obvious points which it may be as well to touch on by way of conclusion. To begin with, there is the extraordinary facility we humans seem to have for recognition—at least for objects falling within a certain range, and comprising a wide variety of different levels of abstraction, from familiar faces to complex social situations. We seem to do it with great efficiency, using extraordinarily plastic representations, in despite of the relevance issues involved and the computationally intensive nature of the reasoning process. This ability extends in varying degrees to the rest of the biological world—indeed, the instances of *animalia* at least may be thought of profitably *as* recognition machines: to a certain extent, active and massive hypothesis generation and testing seems to be what organized nervous systems have evolved to do. Whatever mechanism organic nervous systems have for generating and pruning hypotheses, it must be very efficient indeed. It is probably fair to say that, despite about a century-and-a-half of careful observation, clinical neuroscience is still only at the threshold of elucidating this mechanism, although promising hypotheses are beginning to be put forward, e.g., Hawkins, 2004.

In this context, it is probably worth emphasizing that the close interlinkage of perception, hypothesis, and action is not an accidental feature in living systems, insofar as the behavior in question almost always relates to the survival of the organism and its line of inheritance. This bears closely and inescapably upon an issue that has long shadowed work on symbol-based recognition systems: namely, the problem of what determines the extensions of the so-called natural kinds that are recognized and operated on in rule-based systems. To put the problem another way: given that any number of kinds might be picked out of the general confusion of the world, and recognized and reasoned with on the basis of invariant features exhibited by their instances, is there any non-question-begging way of identifying an optimal set of kinds for use in recognition? The answer, one suspects, is to be found in selection and survival. It seems as certain as any result in science today that relevance and natural kinds are grounded in the exigencies of biology as determined by evolution: that is, if we think of an organism in terms of a set of self-sustaining boundary conditions that are maintained over time, what is relevant to the organism is, first and foremost, these conditions, and secondarily, any other environmental conditions that reliably bear upon their maintenance. It is to the organism's advantage to learn as many of these relationships as it can, which suggests that recognition in the animal world is very tightly coupled to inductive learning, concept acquisition, and adaptive behavior at the level of the individual organism, and to processes of natural selection operating at the level of the species.

The lesson for AI may be that the enterprise of implementing recognition computationally could benefit from closer study of how it fits within the larger evolutionary scheme, and particularly, of how the ability to induce rules and identify concepts within the constraints imposed by relevance arises in that context. More specifically, we may also hazard that the evidence of biology argues in favor of a general model of recognition that applies across the board, from simple feature detection to complex classification, and that pro-active hypothesis and behavioral output figure significantly in the picture, at every level and every step of the way. To say this is not by any means to subscribe to a naïve physicalism or behaviorism—both pitfalls that are properly abhorred—but only to suggest that the biological and evolutionary *milieu* is likely to be the only environment in which recognition, considered as an abstract information *process,* can rightly be understood. A sizable body of work that touches on this already exists and has been under development for the past thirty years, as exemplified by Calvin (Calvin, 1990) and Arbib (Arbib, 2000). More, however, could be done, and should be done, to center future investigation at the intersection of neurology, evolutionary biology, and information science.

## References


Arbib, M. A. and Erdi, P. 2000. Précis of *Neural Organization: Structure, Function, and Dynamics. Behavioral and Brain Science* 23. Cambridge University Press.

Calvin, W. 1987. The Brain as Darwin Machine. *Nature* 330.

Calvin, W. 1990. *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness.* Bantam Books.

Calvin, W. 1998. Competing for Consciousness: A Darwinian Mechanism at an Appropriate Level of Explanation. *Journal of Consciousness Studies* 5 (4).

Creutzfeld, O. 1977. Generality of the Functional Structure of the Neocortex. *Naturwissenschaften* 64. Springer-Verlag.

Goldman, R., Gelb, C., and Miller, C. 1999. A New Model of Plan Recognition. *Proceedings of the 1999 Conference on Uncertainty in Artificial Intelligence.*

Guillery, R.W. 2003. Branching Thalamic Afferents Link Action and Perception. *Journal of Neurophysiology* 90. American Physiological Society.

Hawkins, J. 2004. *On Intelligence.* Times Books

Hofstadter, D. 1995. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books

Kautz, H. A. 1991 A Formal Theory of Plan Recognition and its Implementation. *Reasoning About Plans*, Allen, J.F., Kautz, H. A., Pelavin, R. N., and Tenenberg, J.D., eds. San Mateo, CA. Morgan Kaufmann

Lenat, D. and Shepard, B. Cycorp Scenario Generation Final Report. Government Contract Number F306-00-0160 DARPA

Ramsey, F. P. 1990. *Philosophical Papers,* Mellor, D.H., ed. Cambridge University Press

Rode, B. 2005. Towards a Model of Pattern Recovery in Relational Data. *Proceedings of the 2005 International Conference on Intelligence Analysis.* https://analysis.mitre.org//proceedings/toc.html MITRE Corporation

Senator, T. 2000. Ongoing Management and Application of Discovery Knowledge in a Large Regulatory Organization: A Study of the Use and Impact of NASD Regulation's Advanced Detection System (RADS) . *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining.* pp. 44-53. ACM Special Interest Group on Knowledge Discovery in Data