

ISR Interdisciplinary Systems Research/Interdisziplinäre Systemforschung

88

The Mathematics of Fuzzy Systems

Antonio Di Nola
Aldo G. S. Ventre



Verlag TÜV Rheinland 

A MODEL FOR SCHEMA-GUIDED REASONING

Alf C. Zimmer

Universität Regensburg

Lehrstuhl für Psychologie

Abstract

It is assumed that world knowledge in general consists of representations analogous to quantified statements. Usually the truth value for these statements about real-world events or rules is neither 1 nor 0 but usually in between. Approaches to formalize plausible reasoning on the basis of world knowledge are successful in modelling single deductive steps or chains of independent inferences. However, a general consequence of these models is that the longer the chains of inferences are, the less plausible is the final conclusion. This is apparently not the way people usually evaluate their conclusions because they exhibit considerable (and quite often justified) confidence in the final result of their reasoning. The alternative way of evaluating world knowledge as proposed here starts from the analysis of arguments and the way they are evaluated. Fuzzy schemata are defined which describe the rules underlying the above described revision of world knowledge in the face of either new situations which necessitate decisions or new information which is corroborating, contradicting, or irrelevant. The application of this notion of a fuzzy schema is demonstrated in the analysis of visual scenes. It is argued that visual information processing can be used as a model for information processing in general which is dependent on world knowledge.

One important feature of normal discourse, that is, a discourse between people in a given situation sharing background knowledge is that the information given or requested is not completely specified. For instance, if during a conversation I am asked "Will you be at the meeting?" I will usually be able to come up with an answer immediately. But in order to give this answer I have to access my knowledge about the topic of the ongoing conversation that I assume to be the context of the question, (e.g. meetings which are due in the immediate future or meetings in general). In the case where deductions from my knowledge are not explicit enough to give an answer I can either probe deeper into my knowledge or I can request more information from the speaker.

Without going into details about how common knowledge is generated and known to be mutual (the most recent overview from a more psychological point of view is Clark & Carlson (1982) this example is applied in order to shed light on the following facts: (i) people quite often are extremely fast in making long chains of inference from their knowledge, (ii) they are able to take into account not only standard knowledge but also new information necessitating the revision of the existing knowledge, (iii) in general they come up with a conclusion in finite time, and (iv) they are usually convinced of the correctness or at least aptness of the conclusion underlying their answer.

The analysis of different ways to model this kind of efficient utilization of vague or "fuzzy" information is the aim of this paper. Starting from Toulmin's (1959) analysis of arguments and argumentation, different approaches are presented which can be used to model chainlike deductional structures. The results of these models are then compared to the facts I have pointed out about the informal way of reasoning in the initial example. One general result from these models is that the longer a deductive chain is, the less confident one should be about the final conclusion. While this is supported by experimental evidence from studies of the memory for unrelated items it is in conflict with the certituede expressed in the inferences underlying the interpretation of utterances in conversation. The reason for this difference is that in conversations the bits of informations are

usually related by a common context and structured by an underlying argumentative strategy. For the modelling of this kind of inferential chains I draw upon perceptual schemata as an analogy. I suggest a schema oriented way of evaluating conclusions drawn from world knowledge which takes into account the overall structure of this world knowledge.

The Structure of Arguments.

According to Toulmin (1959) an argument consists of a series of propositions which are evaluated one for one by comparing them with the evidence provided by the context of the whole argument. Both the piecewise and the global evaluation are done in order to convince the user of the argument as well as his or her addresses. An argument consists of at least one claim (usually a hierarchy of claims), based on *evidence* and a *warrant*. In a warrant generally known and accepted rules about the world are stated or referred to. The argumentative reasoning takes place in the context of the *backing*, that is, the world knowledge from where the warrant is derived, and the *rebuttals*, that is, alternative claims. The successful refutation of plausible rebuttals strengthens the convincingness of the argument. Warrant, evidence, and claim are analogous to the major premise, the minor premise, and the conclusion in the syllogistic schema. But, whereas the emphasis in syllogistic thinking lies in the pinpointing of inconsistencies, in argumentation the whole argument including the backing and the refuted rebuttals has to be evaluated in the face of inevitable inconsistencies. This evaluation is done by working through possible alternatives (rebuttals) and choosing the chain of claims which is least inconsistent and closest to the overall claim, that is, the claim highest in the hierarchy. This theory of arguments by Toulmin (1959) can serve as a framework and a normative standard for formalizations of the rules underlying the usage of world knowledge.

Models for Reasoning with Inconsistent or Vague Premises.

Rescher & Manor (1970) have proposed a formalization of plausible

reasoning, that is, a connected system for the formal interpretation of statements which are neither completely true nor completely false, and of procedures underlying judgments and conclusions in this framework. The proposed methods consist primarily in (i) a strategy to cope with assertions which are neither true nor false but more or less plausible and (ii) a strategy to draw conclusions from inconsistent premises:

in (i) the plausibility for all premises is evaluated and subsequently integrated into the overall plausibility of the argument, that is, in Toulmin's terms the *plausibility of the claim*, and

in (ii) one searches for the maximal consistent subset of premises in the argument.

Both strategies can be combined by: first, evaluating the plausibilities; second, determining plausibility thresholds which might be thought to depend on the overall importance of the argument; and, third, searching for the maximal subset consisting entirely of above-threshold plausible premises. Rescher & Manor propose the MIN-operator for the evaluation of the overall plausibility: $\text{plaus}\{\text{conclusion}\} = \text{MIN}(\text{plaus}\{\text{Minor premise}\}; \text{plaus}\{\text{major premise}\})$

It can easily be seen that the longer the chain of deductive steps is the smaller the overall plausibility will be.

This result is very similar to the results in fuzzy reasoning where three different approaches can be differentiated:

- (i) the evaluation of propositions by fuzzy truth values which is practically equivalent to Rescher's strategy (ii)
- (ii) the interpretation of propositions by means of operations on fuzzy sets (e.g. "all X are Y" is interpreted as "X is a fuzzy subset of Y")
- (iii) the interpretation of quantifiers as possibility functions over the range of relative frequencies of instances.

Approach (i) enables use to resolve the old sophistic paradox of the bald man who is still bald if one adds one hair and then another and so on until infinitum (Goguen, 1969). The assumption of an initial truth value insignificantly smaller than 1.00 and the evaluation of every single situation by multiplying the truth values of all steps between the initial state and the present situation leads to ever decreasing truth values. By virtue of this kind of evaluation, approach (i) accounts well for unrelated steps of inference but fails for integrated arguments as much as Rescher & Manor's approach does.

The situation for approach (ii) is similar to that for approach (i) as can be seen from the definition of a fuzzy subset:

If A is a fuzzy subset of B then the membership function for any element in A is equal or smaller than its membership function for B.

If therefore the warrant of an argument is expressed as a statement about a fuzzy subset and the evidence is given by the membership function for an element of this subset then again the claim (the conclusion) can be only equally or less valid than either the warrant or the evidence.

The interpretation of quantifiers as possibility functions over the range of the relative frequencies of the occurrence of instances, approach (iii), (see Fig. 1) models fairly well the meaning

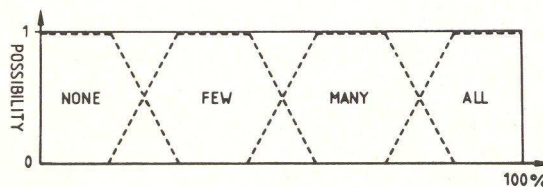


Figure 1 Possibility functions for natural-language quantifiers, the scope gives the possibilities of events

of quantifiers in natural language. Furthermore by means of scope-functions for the contexts in which quantified statements

occur it becomes possible to model context dependent meanings of quantifiers (Zadeh, 1982, Zimmer 1982a). But again, the algorithm for deduction with these quantifiers implies that in a chain of inferences the possibility value of the complete line of arguments can be maximally as high as the possibility value of its weakest link.

It is possible to combine the three approaches by defining quantifiers as elastic constraints on statements:

If one defines the fuzzy sets \underline{X} and \underline{Y} by the membership functions μ_Y and μ_X then the following elastic constraints can be applied to model quantifiers:

$$\text{ALL } (\underline{X}; \underline{Y}) = \text{poss}\left[\frac{\mu_Y - \mu_X}{\mu_Y} \rightarrow 0\right] \ll \text{poss}\left[\frac{\mu_Y - \mu_X}{\mu_Y} \leq 0\right] \quad (1)$$

$$\text{MANY } (\underline{X}; \underline{Y}) = \text{poss}\left[\frac{\mu_Y - \mu_X}{\mu_Y} \rightarrow 0\right] < \text{poss}\left[\frac{\mu_Y - \mu_X}{\mu_Y} \leq 0\right] \quad (2)$$

$$\text{FEW } (\underline{X}; \underline{Y}) = \text{poss}\left[\frac{\mu_{\text{not } Y} - \mu_X}{\mu_{\text{not } Y}} \rightarrow 0\right] < \text{poss}\left[\frac{\mu_{\text{not } Y} - \mu_X}{\mu_{\text{not } Y}} \leq 0\right] \quad (3)$$

$$\text{NONE } (\underline{X}; \underline{Y}) = \text{poss}\left[\frac{\mu_{\text{not } Y} - \mu_X}{\mu_{\text{not } Y}} \rightarrow 0\right] \ll \text{poss}\left[\frac{\mu_{\text{not } Y} - \mu_X}{\mu_{\text{not } Y}} \leq 0\right] \quad (4)$$

\underline{X} and \underline{Y} are fuzzy sets in the universe of discourse standardized to the scope function which captures the possibility of events. The relations (\ll , $<$, $>$, and \gg) are assumed to be fuzzy (Kaufmann, 1975). Thereby comparisons of evidence are made possible.

In this approach deductive chains can be modelled by assigning

the following values to conclusions: if a conclusion falls in between the elastic constraints of a certain quantifier then the most typical value of this quantifier is taken to evaluate the conclusion (e.g. a conclusion A falling in between the elastic constraints of MANY would be evaluated by $\text{poss}(A) = .65$ according to Figure 1). This approach is hardly applicable to chains of arguments because if by chance information with $\mu(A) = 0$ is used as evidence in an inferential step then the overall evaluation might become NONE if no provisions are made to prevent this kind of instability in the deductive process. Furthermore this formalization has a bias towards negative conclusions.

The above sketched approaches of reasoning with vague concepts fail for lines of arguments or, at least, lead to conclusions which contradict known facts about natural inferences, because they imply independence of warrants and evidence. This kind of null assumption about the structure of knowledge is apparently wrong for human world knowledge. Most easily this can be demonstrated for the perception of visual scenes which form a major basis for the world knowledge internally represented. For the analysis of complex visual scenes a number of processing models have been suggested which assume that either perception is predominantly driven by stimuli provided in the environment or that it is mostly influenced by higher-order concepts, that is, the world knowledge. In the most extreme data-driven models the human visual system is equated to something like a camera. While such reductionist models fail even for simple phenomena like form, size, or color constancies, it is possible to develop robust data-driven models which are able to account for some of these effects. These models can be expanded too to chains of inferential reasoning.

Modelling Robust Inference for Independent Information.

One interpretation of how complex visual scenes can be perceived as integrated wholes (Gestalten) is that "unconscious conclusions" (Helmholtz, 1896) are drawn to put together the available pieces of information. By virtue of their being unconscious these conclusions can be assumed not to be affected by the above shown impairments of chains of inferences. This can

be modelled by a theory of robust inferences about imprecise statements which consists of a renormalizing procedure after each deductive step and a memory span limited to the most recent state. In (Zimmer, 1983) I have shown that the "availability bias" (Tversky & Kahneman, 1973) can be modelled fairly well with these two elementary processes. One problem connected with this model is that, due to the limited memory capacity, it does not account for the tendency of knowledge to resist changes which are too abrupt. This tendency seems to be of high evolutionary importance because it prevents instability. This model can be elaborated further by taking into account the belief strength in the existing knowledge (the inertia of the system) and the impact of new information (its saliency for the existing system). Continuous processing of independent pieces of information as, for instance, in forecasting can be represented in this framework (Zimmer 1982 b). The belief strength (b_i) is assumed to be a monotonically decreasing function of the fuzzy distance between the quantified knowledge (Q) at time $t(i)$ and at time $t(i+1)$ normalized to the interval $[0,1]$.

$$b_i = f [d(Q_i; Q_{i-1})] \quad 0 \leq b_i \leq 1 \quad (5)$$

The saliency parameter for novel information, a , is restricted to the interval $[0,1]$. The resulting change of the knowledge in the light of new information with a given saliency can then be described by the following formula:

$$Q_{i+1}^{(x)} = \text{MAX}_x \left[b Q_i^{(x)} + (1-b) I_{i+1}^{(x)}; \frac{M_x [N_x [a \cdot I_{i+1}^{(x)}; Q_i^{(x)}]]}{\text{MAX}_x [I_{i+1}^{(x)}; Q_i^{(x)}]^r} \right] \quad (6)$$

In this formula $Q(x)$ represents the quantification of event x at time $i+1$, $I(x)$ is the novel information about x at time $i+1$. The parameter 'r' reflects the degree to which the chain of arguments is integrated: for $r = 0$ the conclusions are independent and for $r = 1$ the conclusions are maximally interdependent ($0 \leq r \leq 1$).

In Fig. 2 the quantified old knowledge, the quantified incoming information, and the predicted

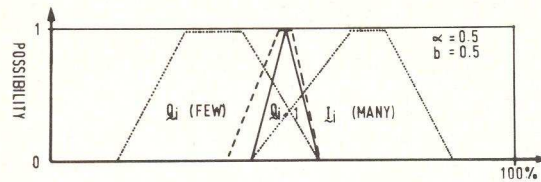


Figure 2 Comparison of the predicted (____) and the actual change (-----) of knowledge.

versus the actual change in quantification are shown for an r -value of 1. The good match between the predicted and the observed possibility function indicates that the subjects had implied that the information gathered by them is tightly interconnected as shown by the high r -value.

It has to be kept in mind that in these models no structural assumptions about the knowledge and the active search for information (the argumentative strategy) have been made, except for the saliency which is supposed to be given or otherwise set to .5, and except for the ' r ' parameter which reflects the empirically determined interdependency of the warrants, evidence, and claims. In most approaches to reasoning they are assumed to independent.

The research on judgmental processes as done by Kahneman and Tversky (see Kahneman, Slovic, and Tversky, 1982) reveals that in many situations the reasoning based on simplifying heuristics leads the decision maker astray because it either neglects necessary structural information or it overgeneralizes structural assumptions. In order to avoid the pitfalls of reasoning mentioned, it is necessary to integrate the structural knowledge into the reasoning process. It can be shown that for a self-organizing system - and world knowledge seems to be such a system - it is necessary to be able not only to return to a stable state after perturbation but also to be able to investigate new developmental pathways through successive states of instability. The performance of a system in coping with its environment

depends on how much its complexity is counterbalanced by its structural integration and its reliability (Sahal, 1979). While in the above sketched model the reliability has been taken care of by the "inertia" of the system as expressed in the belief function, ways to get hold of the structure of the system have still to be developed. As pointed out before, visual perception can serve as a good example for such a system insofar as it exhibits not only homeostatic features as, for instance, the constancy effects, but also it is highly integrated. This can be most strikingly demonstrated in the perception of complex visual scenes where the background knowledge, the intention in perceiving (see Barwise, 1982), and the physically describable stimuli give rise to an integrated perception about what is going on.

Schematic Perception as a Model for World Knowledge in General.

Following Cassirer (1944), a "schema" in perception can be defined as consisting of

- (i) a set of primitives which are not further analyzable in the given context,
- (ii) a set of organizational rules which can be paralleled to Helmholtz' logic of unconscious inferences
- (iii) a set of admissible transformations, that is, transformations which define the class of invariants of the objects in questions.

It has been shown that in general these schemata do not exist in separation but that they are structurally organized in hierarchies or in other structures. Hierarchical structures are of special importance because they exhibit the feature of near-decomposability (Simon, 1965) while at the same time higher level schemata impose constraints on the sets of admissible transformations in lower level schemata (see Fig. 3). The

application of fuzzy set theory to syntactic pattern recognition

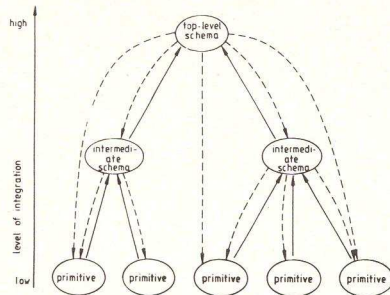
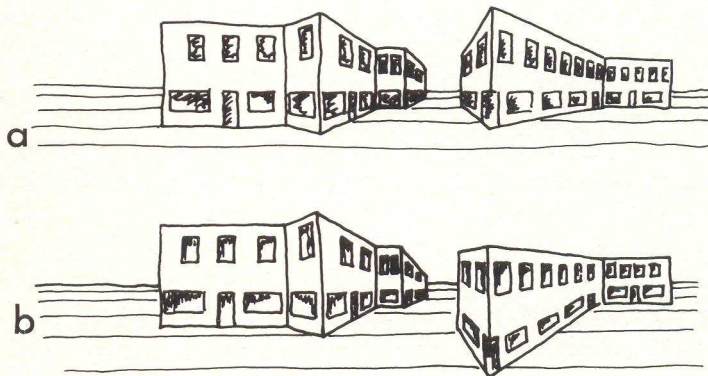


Figure 3 A hierarchy of schemata

(Freksa, 1981; Jain & Haynes, 1982) fits well into this theoretical framework because it enables people to discriminate between the different meanings a blurred part of a picture or a sloppily drawn line might carry. For instance Freyd (1983 a) has experimentally demonstrated that in the recognition of handwriting the very same line can have different meanings under the variation of handwriting methods: in one case it might provide structural information while in the other case it can be regarded as a dismissable error.

In Fig. 4 (b) through (d) it is shown how different elements of the original scene 4 (a) are picked up in the different drawings,



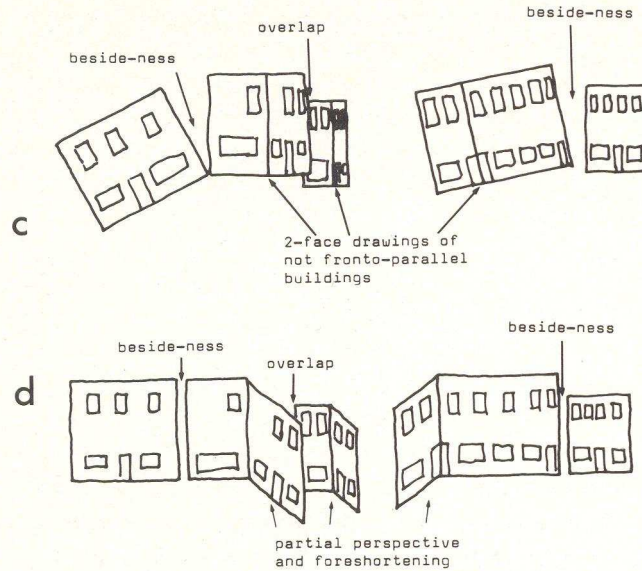


Figure 4 Line drawing of a photograph depicting a perspective scene (a), the same scene as drawn by a graphic designer (b), and by two students without training in drawing (c) and (d).

which were intended to enable an onlooker to find out from which point of view they had been taken. In drawing 4 (b) the fuzziness consists entirely in a sort of wriggling while drawing straight lines. By the application of low-pass filtering the original drawing could be restored easily because the artist has applied the same kind of transformational constraints - that is, those of projective geometry - which are captured by a camera. The situation in drawings 4 (c) and (d) is different: here subjects have decomposed the integrated perspective scenes into only loosely connected subparts. However, these subparts exhibit the kind of downward constraints on transformations typical for integrated wholes. It turned out that for human observers the representations (a) to (d) were equally informative for the task at hand, that is, determining from which point of view these pictures had been taken (Zimmer, in preparation). This result indicates that the invariants which made possible the performance

of the observers are not necessarily those of projective geometry as in (a) and (b). The topological constraints which are preserved in drawings (c) and (d) resemble more the features of visual scenes selected by observers for verbal descriptions. This result is in line with Shareability theory (Freyd, 1983 b) which identifies the ways mental representations might be molded by the necessity to share knowledge.

Klix (1971) among others describes problem solving as transformations in the problem space linking the initial state with the state of the intended goal, the solution. What is lost in this model in comparison to Köhler's is the importance of autochthonous processes which structure automatically the applicability of transformations; these processes are reminiscent of Gibson's (1979) concept of affordances. What is common to these approaches is the distinction between initial states transformations (or means) and one or several final states. As mentioned above current classifications of problems according to these concepts bear strong similarities especially in the point that they do not include problems like the conversationally appropriate interpretation of an utterance as described in the introduction. Another structurally related problem is that of the typical detective story in which an active search is performed for mostly circumstantial evidence. The accumulation of circumstantial evidence is governed by the goal to exclude the innocents and to find the most probable culprit. The case against a possible culprit is strengthened by every refutation of arguments which speak against his or her guilt (see Toulmin, Rieke, Janik, 1979).

The contextual constraints work not only on the invariants but also on the definition what is to be regarded as fuzziness. Sperber and Wilson (1982) have analyzed discourse understanding from the point of view of contextual constraints. They claim that by merely applying the principle of maximal relevance it is possible to decide which utterance is informative and which is not. One problem connected with Sperber & Wilson's approach is that the determination of maximal relevance presupposes that for any utterance all possible non-trivial inferences have to be made (Moore, 1982). In the case of integrated arguments, however, the

exerted downward constraints prevent the possibility of an infinite regress.

Consequences for the Evaluation of Arguments.

In developing convincing strings of arguments people seem to follow a strategy quite analogous to the top-down analysis in visual perception of complex scenes: (i) the general claim, or the ultimate goal of the argument, is stated; (ii) in a top-down search a string of subclaims is generated that connects the available evidence and relevant world knowledge to the ultimate goal of the argument; (iii) new evidence is searched for if necessary; and finally, the available knowledge is checked for alternative strings leading to the same general claim or the string of claims is scrutinized to see if it supports alternative claims too. The organizing structure underlying this strategy can be assumed to be provided by the inferential schemata available. At least in our culture, causal schemata seem to make up the most convincing strings of arguments. Why this is the case can be seen from what people take as evidence for causality. For instance, Schustack and Sternberg (1980) report that their subjects' judgments of the strength of a causal relationship could be accounted for by the following variables: (a) joint presence of the cause and the effect, (b) violation of sufficiency, (c) violation of necessity, (d) joint absence of the cause and the effect, and (e) the strength of alternatives. In modelling the revision of world knowledge these variables can be used in order to determine the parameter of saliency, a , and the parameter of interdependence, r , in Equation (6) which capture structural information about the world knowledge. With the saliency of information as depending on the underlying causal structure the proposed formalization becomes a model for integrated argumentative reasoning.

Admittedly, this suggested approach for modelling the revision of world knowledge in an argumentative discourse or in an inferential search falls short of providing a fullfledged formalization or a general theory of plausible reasoning. I hope to have shown, however, that schematic perception can serve as an fruitful analogy for the research on argumentative reasoning in relation to the revision of world knowledge.

References

- Barwise, J. (1981) Scenes and other situations. *The Journal of Philosophy*, 78, 369-397.
- Cassirer, E. (1944) The concept of group and the theory of perception. *Philosophy and Phenomenological Research*, 5, 1-36.
- Clark, H.H. & Carlson, T.B. (1982) Speech acts and hearers' beliefs. In: Smith (1982).
- Freksa, C. (1981) *Linguistic Pattern Recognition*. Ph. D. thesis University of California, Berkeley.
- Freyd, J.J. (1983 a) Representing the dynamics of a static form. *Memory and cognition* 11(4), 342-346.
- Freyd, J.J. (1983 b) Shareability: the social psychology of epistemology. *Cognitive Science*, 7, 191-210.
- Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Goguen, J.A. (1969) The logic of inexact concepts. *Synthese*, 19, 325-373.
- Helmholtz, H. von (1896) *Handbuch der Physiologischen Optik*, Volume III, 2nd edition, Leipzig: Voss 1896.
- Jain, R. & Haynes, S. (1982) Imprecision in computer Vision. *IEEE Transactions on Computers*, August, 18, 39-48.
- Kahneman, D., Slovic, P., Tversky, A. (1982) *Judgments under Uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kaufmann, A., (1975) *Theory of Fuzzy Subsets*. Vol. 1, Academic Press.

- Klix, F. (1971) *Information und Verhalten*. Bern: Huber.
- Moore, T. (1982) Comments on Sperber and Wilson's paper. In: N.V. Smith (1982).
- Rescher, N. & Manor, R. (1970) On inference from inconsistent premises. *Theory and Decision*, 1, 179-217.
- Sahal, D. (1979) A unified theory of self-organization. *Journal of Cybernetics*, 9, 127-142.
- Simon, H.A. The architecture of complexity. *Proceedings of the American Philosophical Society*, 1965, 106, 467-482.
- Schustack, M.W. & Sternberg, R.J. Evaluation of evidence in causal inference. *Journal of Experimental Psychology - General*, 1980, 110, 101-120.
- Smith, N.V. (1982) *Mutual Knowledge*. New York: Academic Press.
- Sperber, D. & Wilson, D. (1982) Mutual knowledge and relevance in theories of comprehension. In: Smith (1982).
- Toulmin, S. (1959) *The Uses of Argument*. Cambridge: Cambridge University Press.
- Toulmin, S., Rieke, R. and Janik A. (eds.) 1979. *An introduction to reasoning*. New York: Macmillan Publ. Co., Inc.
- Tversky, A. & Kahneman, D. (1973) Availability: A heuristic for judging frequency and probability. *Cognition*, 5, 207-232.
- Zadeh, L. (1982) *A Computational Approach to Fuzzy Quantifiers in Natural Languages*. Berkeley: University of California, Memorandum No. UCB/ERL M82/36.
- Zimmer, A.C. (1982a) Some experiments on the meaning of quantifiers in normal language. In: L. Troncale (ed.) *General Systems Science* Louisville, Ky.: Society for General Systems Research, 1982.

Zimmer, A.C. (1982b) A model for the interpretation of verbal predictions. paper presented at the 2nd International Symposium on Forecasting, Istanbul.

Zimmer, A.C. (1983) Verbal versus numerical processing of subjective probabilities. In: R. Scholz (ed.) *Decision Making under Uncertainty*. Amsterdam: North-Holland.

Zimmer, A.C. (in preparation) What makes the eye intelligent?

Footnotes

Most of the theoretical work presented in this paper has been done while the author was on leave at the University of California, Berkeley, and at Stanford University. I want to thank Lotfi Zadeh especially for the many fruitful discussions we have had and for the opportunity to present my ideas in his Expert Systems Seminar. Furthermore I want to express my appreciation for the many helpful comments on earlier drafts of the paper I got from Jennifer Freyd. Her suggestions have improved the clarity of the argument and the quality of the style significantly. Furthermore, the suggestions of an unknown reviewer have helped me to express my intentions more clearly.