Title
Imagery and Categories: The Indeterminacy Problem

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 13(0)

Author
Ioerger, Thomas R.

Publication Date
1991

Peer reviewed

Imagery and Categories: The Indeterminacy Problem

Thomas R. Ioerger*
Beckman Institute
University of Illinois
405 N. Mathews Ave.
Urbana, IL 61801
ioerger@cs.uiuc.edu

Abstract

One of the classical problems faced by theories of mental imagery is the Indeterminacy Problem: a certain level of detail seems to be required to construct an image from a generating description, but such detail might not be available from abstract, categorical descriptions. If we commit to unjustified details and incorporate them into an image, subsequent queries of the image might indiscriminantly report not only information implied by the description but also information that was arbitrarily fixed. The Indeterminacy Problem is studied in a simplified domain, and a computational model is proposed in which images can be incrementally adjusted to satisfy a set of inter-constraining assertions as well as possible. In this model, queries can discriminate between those details in an image which are necessary (implied by the generating description) and those which are incidental (consistent but arbitrarily fixed). The computational model exploits the graded prototypicality of the categorical relations in the simplified domain, and suggests the importance of a grounded language for reasoning with categories.

Introduction

Although the psychological validity of mental imagery has long been debated, images have often served usefully for modeling knowledge in computational systems (Funt 1980; Waltz & Boggess 1979). One of the arguments against imaginal representations and for propositional ones has been that images seem to possess a certain level of detail in representation that cannot always be provided by an abstract description being used to generate the image (Pylyshyn 1973). For example, in imagining a tiger, people often report that the tiger has stripes but they cannot say how many (Kosslyn 1980). Yet if an image of a striped tiger had been generated, it is argued, surely its stripes could have be counted. If default information had been used to fill in the required details, the image would have been unjustly over-committed; the necessary information in the image would have been indistinguishable from those details which were consistent but arbitrary. Pinker (1984) claims that this problem "would speak against a totally factored structural description," suggesting that imaginal representations are inherently incapable of handling such indeterminacy. We call this the Indeterminacy Problem.

The Indeterminacy Problem has been acknowledged in at least one computational system that uses images for reasoning (Waltz & Boggess 1979). Boggess's computer program takes a sequence of (restricted) natural language sentences as input, constructs an internal image of the spatial relationships implied by the prepositional phrases, and answers questions about these spatial relationships by examining the image. Nouns like "table," "box," and "goldfish bowl" are drawn as right parallelpipeds with default height, width, and depth in a three-dimensional bitmap representations of space.[1] Prepositions like "on" and "in" are realized as procedures for placing and locating objects in the image relative to other objects.

In Boggess's program, the meaning of a preposition can be sensitive to the kinds of objects being related, so a shadow can be "on" a wall in a different way than a book can be "on" a table. Nevertheless, for any instantiated proposition, the program has default knowledge for drawing and evaluating images. For example, to draw a book on a table, a specific spot which "tends to a particular corner" of the table is used. This of course simplifies the task of finding the book on the table later, but the use of such default information can lead to some embarrasing conclusions. Consider Waltz and Boggess's own example: the program is given "The shelf is on the wall" and "The fly is on the wall" as input. Using default knowledge, the program constructs an acceptable model in which the fly is below the shelf. When the program is queried with "Is the fly below

*This material is based upon work supported under a National Science Foundation Graduate Fellowship and ONR grant N00014-88K0124.

[1] Bitmaps are quantized representations of space and its occupation by objects. In this paper, a Cartesian coordinate system will be assumed.

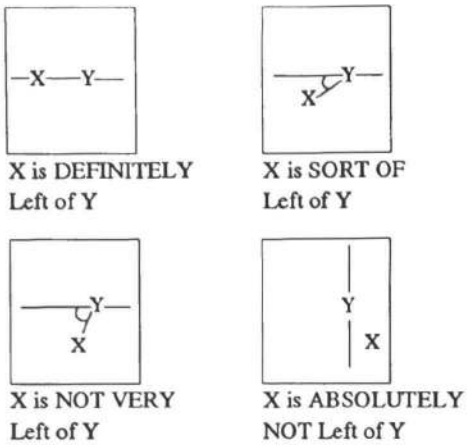

Figure 1: The categorical preposition Left has many meanings, which are distinguished by the use of various hedges in natural language.

the shelf?" it responds affirmatively. Yet the program was not told this directly, nor was it implied by the input. The information was derived as an artifact of using defaults to construct the image; this is a direct demonstration of the Indeterminacy Problem.

A Simplified Domain

To study the Indeterminacy Problem, we define a simplified domain in which categorical relations are used to generate and query images. The domain consists of three structureless points, labelled X, Y, and Z, initially assumed to exist in some arbitrary spatial arrangement on a plane. We decompose the use of images into two processes that interact. There is a client process which can make a sequence of (possibly interleaved) assertions and queries to an image server in a propositional language. The language consists of three constants–X, Y, and Z–and four two-place predicates–Left, Right, Above, and Below. The image server must update the image when given assertions, and must derive the answers to queries by examining the image (as opposed to proving theorems with the assertions).

The categorical nature of the four relations can be explored by first making some intuitive observations on how people might hedge descriptions in this simplified domain (see Figure 1). There is a technical sense of the Left relation between two points, say X and Y: X is "definitely" left of Y if X is on the left half-plane with respect to Y and the line XY is horizontal. However, if the absolute value of the angle between the line XY and the horizontal is small but non-zero, we might say that X is "sort of" left of Y. As the line XY approaches verticality, we might say that X is "not very" left of Y. And finally, if X is anywhere on the right half-plane with respect to Y, then X is "absolutely not" left of Y. The other relations can be defined symmetrically. Note that these definitions do not depend on the distance between X and Y and are invariant under translation. Also notice the internal symmetry in each relation; for example, given any image, X is just as much to the left of Y as the image in which X is flipped across the horizontal passing through Y.

Lakoff (1973) has shown that the use of hedges with propositions constructed from a particular predicate strongly suggests that the propositions should be allowed to take on not only True and False as truth values, but also intermediate degrees of truth. If we follow the conventions of Fuzzy Logic (Zadeh 1965) and map truth values into the unit interval, [0,1], where 0 represents falsity and 1 represents truth in their classical senses, the intermediate values can be interpreted as prototyicality ratings–estimates of the closeness of some state of the world (in this case, modeled by an image) to the "central meaning" of the categorical predicate (Smith & Medin 1984). Given these two representation languages–images and fuzzy-truth-valued propositions–we must specify how they are related so that asserted propositions can be used to update images and so images can be used to answer queries. Specifically, for any instantiated proposition, such as '(Left X Y), we must describe how images are related to truth values.

The examples of hedging in Figure 1 suggest that the prototypicality rating of an image for the the proposition '(Left X Y) is a function of only the absolute value of the angle between the line XY and the horizontal (based on independence from distance and on internal symmetry).[2] Prototypicality is highest when the angle is zero, decreases monotonically as the absolute value of the angle increases to 90 degrees, and is essentially zero when X is in the right half-plane with respect to Y (although there are still slight differences in distances from the prototypical Left relationship). Figure 2 describes a method for computing the truth value of an image for the proposition '(Left X Y) using a function LEFT that has these properties. The other relations can be computed similarly to the Left relation by first rotating the image. The negation of a proposition with a fuzzy truth value q is defined to be the same proposition with the truth value 1-q (Zadeh 1965).

This method for computing prototypicality only models the "perception" of images in terms of fuzzy-truth-valued propositions. The inverse relation is one-to-many because, for any instantiated proposition and

[2] We assume that the hedges in Figure 1 roughly reflect prototypicality. That is, for any proposition P, an example of "definitely" P is more prototypical than an example of "sort of" P, which is more prototypical than an example of "not very" P, which is more prototypical than an example of "absolutely not" P. See Lakoff (1973) for a formal discussion of the relationships of hedges to prototypicality based on possibility distributions.

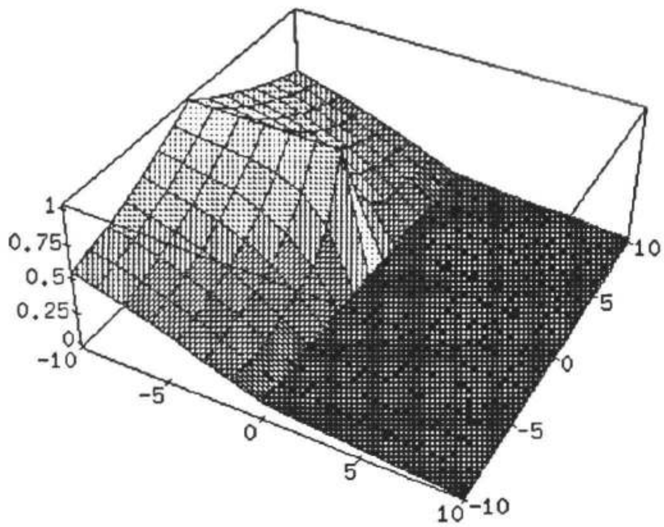

Figure 2: To compute the truth value of '(Left X Y) for a particular image, translate both X and Y so that Y is at the center of the image. Then apply a function LEFT, which might look like this plot but must be investigated empirically, to the translated location of X. The number that is returned represents the prototypicality of the image as an example of '(Left X Y). The function LEFT decreases monotonically as the absolute value of the angle between the line XY and the horizontal increases from 0 to 90 degrees. There is a slight gradient in LEFT when X is in the right half-plane with respect to Y because, eventhough X is "definitely not" left of Y for any of these positions, they are not equivalent; some relative positions of X are closer to the prototypical Left relationship with Y than others.

some truth value, there are many equivalent images which differ in distance between the arguments or which are symmetric. Also note that the "meaning" of a proposition becomes a possibility function among images such that the distribution reflects prototypicality. This property will later be exploited to generate default information when updating images with asserted propositions.

In this simplified domain, queries should return one of four values:

- YES means the queried proposition was perviously asserted verbatim, or the image is sufficiently constrianed by the previous assertions that the queried proposition must be true (have an acceptably high truth value). For example, after asserting '(Above X Y) and '(Above Y Z), the query '(Below Z X) should return YES.
- NO means the previous assertions sufficiently constrain the image so that the negation of the queried proposition must be true. For example, after asserting '(Above X Y), the query '(Above Y X) should return NO.
- UNSPECIFIED indicates that the assertions have not constrained the model enough to favor either the queried proposition or its negation; both are consistent with the previous assertions. For example, initially X, Y, and Z are in some arbitrary arrangement. Since no assertions have been made, any query should return UNSPECIFIED.
- CONTRADICTION should be returned if the previous assertions contained two contradictory propositions, like '(Above X Y) and '(Below X Y), or were similarly overconstrained to make any sense.

A Computational Model

We propose two procedures as a computational model of how the image server can process a sequence of assertions and queries and still avoid the Indeterminacy Problem. Consider the case when the image server is given an assertion that is categorical and has some choice as to how to update the image to satisfy the assertion. As a working example, suppose we first assert '(Above X Y). As long as X is on the upper half-plane with respect to Y, the image would satisfy the assertion somewhat. But, since there are no other constraints, the best image is one in which X is directly above Y. Even if we choose this spatial arrangement we must arbitrarily choose a distance between X and Y to construct the image. To see how these choices can be defeated with additional constraints, suppose we now assert '(Right X Y). As long as X is on the half-plane to the right of Y, the image would satisfy this new proposition somewhat, and the way to satisfy the proposition in the most prototypical sense would be to have X directly to the right of Y. But these constraints interact; intuitively, the most appropriate spatial arrangement in which X is both above and to the right of Y has X on a 45 degree angle above the horizontal from Y. Thus we have to modify some of the details that we arbitrarily chose for modeling the first assertion.

Figure 3 shows the Adjustment Procedure, which can be used by the image server to handle a sequence of assertions properly. The Adjustment Procedure assumes the image server is maintaining a record of the previous assertions and has a current image which models the previous assertions as well as possible. To determine how well an image models a set of propositions, we take the minimum[3] prototypicality rating of the image for any previously asserted proposition; this is called the evaluation of the image with respect to the set of propositions. When an assertion is made to the image server, it calls the Adjustment Procedure, passing to it the set of previous assertions, the new assertion, and the current image. The image server resets the current image to the one which is returned by

[3] This is a standard aggregator function for conjunctions in Fuzzy Logic (see Zadeh 1965).

1. Generate a set Φ of adjusted images.
2. If no image in Φ has a better evaluation with respect to $\Gamma \cup \{P\}$ than I, then return I.
3. Else reset I to some image J in Φ such that no other image in Φ has a better evaluation with respect to $\Gamma \cup \{P\}$, and goto step 1.

Figure 3: The Adjustment Procedure takes as input a set of propositions Γ, a new proposition P, and a starting image I, which is expected to be the best model of Γ that could be found. The procedure returns an updated image that models the propositions $\Gamma \cup \{P\}$ as well as possible.

the Adjustment Procedure, and it also adds the new assertion to the set of previous assertions.

Rather than search all possible spatial arrangements for an image for which no other image has a better evaluation, it is a central proposal of the Adjustment Procedure that the search proceed in a directed and incremental way. The only images that are considered are the ones constructed by adjusting the position of the first argument of any least-satisfied assertion (with respect to the current image) a small distance in any direction,[4] and the ones formed by similarly adjusting the position of the second argument of such an assertion. If none of these evaluates better with respect to the set of propositions (including the most recent assertion), then we are done so we return the current image. Otherwise we reset the current image to one of the adjusted images such that none of the other adjusted images has a better evaluation, and then we iterate. This search procedure is called hill-climbing (Winston 1984). It is guaranteed to halt, possibly at a locally but not globally best image, if the representation of images admits only a finite number of states, which is true for bitmaps.

It should be clear that the Adjustment Procedure will find an appropriate spatial arrangement for our working example. Given an image that satifies the first assertion '(Above X Y) prototypically, the prototypicality of the second assertion '(Right X Y) will be nearly 0. According the the Adjustment Procedure, we construct and evaluate the images in which X or Y are adjusted independently, and we discover that moving X to the right or Y to the left increases the evaluation of the image. This process iterates until X is on a 45 degree angle above the horizontal with respect to Y; any adjustments of this image would have equivalent evaluations (if X were moved directly toward or away from Y) or the prototypicality of one of the asserted propositions would drop, decreasing the overall evaluation.

[4]With bitmap representations, space is already quantized (at a hopefully adequate resolution) so the finite number of possible adjustments to the position of an object is explicit.

1. Call the Adjustment Procedure with Γ, P, and I. Let A be the evaluation of the returned image with respect to $\Gamma \cup \{P\}$.
2. Call the Adjustment Procedure with Γ, the negation of P, and I. Let B be the evaluation of the returned image with respect to $\Gamma \cup$ {the negation of P}.
3. If $A \geq C$ and $B < C$, then return YES.
4. If $A < C$ and $B \geq C$, then return NO.
5. If $A \geq C$ and $B \geq C$, then return UNSPECIFIED.
6. IF $A < C$ and $B < C$, then return CONTRADICTION.

Figure 4: The Decision Procedure takes as input a set of propositions Γ, a queried proposition P, and an image I, which is expected to be the best model of Γ that could be found. The Decision Procedure returns one of four responses. A represents how consistent P is with Γ since it is the evaluation of the best model of $\Gamma \cup \{P\}$ that could be found by the Adjustment Procedure. Similarly, B represents how consistent the negation of P is with Γ. Since these consistency values come from evaluations of images, which are minima of prototypicality ratings, they fall into [0,1]. We differentiate high consistency from low consistency in this range with a cutoff C, which should be determined empirically.

Figure 4 shows the Decision Procedure, which uses the Adjustment Procedure to avoid the Indeterminacy Problem when answering queries about an image. To answer a query, we must first check to see how the queried proposition interacts with the previous assertions. So we call the Adjustment Procedure to find a best adjusted image if we were to assert the proposition, and we evaluate that image. If it has a high evaluation, then the Adjustment Procedure was able to find a spatial arrangement in which all of the asserted propositions plus the queried one could be satisfied fairly well. This alone does not imply that the previous assertions necessitate the queried one, however; it might spuriously be consistent or just not interact with the previous assertions. Thus we have to call the Adjustment Procedure to see what the effect would be if we had asserted the negation of the queried proposition, and we evaluate the image that is returned this time as well.

The two evaluations are enough to make a decision. To respond affirmatively to the query, it must be the case that the queried proposition is consistent with the previous assertions, but its negation over-constrains the image, producing a low evaluation for the best model that could be found if the negated proposition had been asserted. The analysis is symmetrical for the negative response. If both the queried proposition and its negation are consistent with the previous assertions, then the previous assertions do not constrain the answer, and, regardless of the actual details of the image maintained by the server, the response is UNSPECIFIED. Finally, if neither the queried proposition nor its negation are consistent with the previous assertions, the previous assertions must be contradictory (over-constrained) since for any not-over-constrained model we expect at least one of the propostions to be true.

This computational model has been implemented for the simplified domain in ISR, a computer program written in Common LISP. An image is represented with a 10×10 bitmap, which is a list of ten lists of ten symbols. ISR avoids the Indeterminacy Problem by using the Adjustment Procedure to update images when assertions are made, and by using the Decision Procedure to answer queries about the current image. As an example, ISR begins with an arbitrary spatial arrangement of the points X, Y, and Z; all queries respond with UNSPECIFIED. If we assert '(Left X Y) and then '(Left Y Z), ISR can tell that us '(Left X Z) and '(Right Z X) are true but '(Right X Y) is false. If we query '(Above X Y), ISR responds UNSPECIFIED because X could be either above or below Y and still satisfy the constraints of the previous assertions. ISR can also handle interacting assertions and can find an appropriate image (with X on a 45 degree angle above and to the right of Y) in our working example, after '(Above X Y) and '(Right X Y) have been asserted.

Discussion

Kosslyn (1980) has developed a detailed computational model of imagery based on an imaginal representation called a visual buffer, which is essentially a two-dimensional bitmap. However, the Indeterminacy Problem is not a difficulty in this model because of the restricted range of tasks it was designed to explain. The phenomenon Kosslyn was trying to model is how, given a question about the appearance of some object, people seem to retrieve an image and examine it to find the answer. For example, it was hypothesized that, to compute how far apart the front and rear wheels of a car are, one simply scans the distance in the visual buffer between the tires in a default image of a car. Thus, default information was expected to be used to answer queries, and no subsequent assertions were made which could interact with and defeat such details in an image. When addressing the common introspection that images can somehow be sketchy and abstract, Kosslyn claims that this effect could be achieved in visual buffers by simply leaving out details, but it remains unclear how this would work.

From a general computational perspective, we might want to use images as mental models (Johnson-Laird 1983) to reason about knowledge in other formats. For example, images can form an oracle for inferring things from an alternate description because their fixed level of detail in representation makes implications explicit. The source of the Indeterminacy Problem lies in translating between the language in which the knowledge is described and the "language" of images. For propositional representations, descriptions might be indeterminate because they simply do not mention some aspect of the world, or because they are categorical in the sense that they can only give a possibility distribution over some states rather than implying which are necessarily true and which are necessarily false. This abstraction away from insignificant details is the power of propositional representations, but modeling knowledge in this format with images is not straightforward.

Our solution to using detailed images to model indeterminate propositional descriptions is to incrementally adjust an image to satisfy as many of the previously asserted constraints as possible, making and modifying consistent but not necessary choices as needed. To distinguish the necessitated from the just-consistent information in an image when we query it, we consider both the consistency of the queried proposition and of its negation with the previous assertions, again relying on adjustment to find the most satisfactory models. This computational model discredits a classical argument against imagery by showing how detailed models of indeterminate descriptions can be maintained and used without committing to consistent but unnecessary details required by the representation. It has been acknowledged that statements about the limitations of imaginal representations often fail to address the variety of possible processes that could operate on them (Anderson 1978; Johnson-Laird 1983). Our computational model suggests an interesting process that surprisingly would provide imaginal representations with the potential to handle indeterminacy.

If we view the image server from the perspective of the propositional client, it appears to be doing default reasoning with categories. This capability can be explained by the grounded nature (Lakoff 1987) of the "language" of images; this language is so close to perception that it can distinctly represent all the relevant states of the world, and the interactions among categories can be explicity described at this level. Our computational model demonstrates how a grounded language can be used to model categorical knowledge in such a way that 1) defaults can be assumed at one time and retracted or modified at a later time, and 2) implications of the generating description found by examining a detailed model can be distinguished from information in the model that is consistent but not necessary. Thus, if knowledge can be translated into a grounded language, default reasoning with categories reduces to a Constraint Satisfaction Problem (Mackworth 1990).

The domain in which we developed our computational model was greatly simplified; we should consider the possibility of extending the model to handle more interesting categories used in imagery (Herskowitz 1986). Even for Left, Right, Above, and Below, there are complications we have ignored, such as the fact that the sense of these predicates can be influenced by their arguments or by context. One subtle aspect of determining the meaning of spatial prepositional phrases, for example, is choosing a frame of reference (Clark 1973). As another example, the spatial preposition "over" has been shown to have nearly one hundred qualitatively different senses, as opposed to quantitative differences in distance or angle (Brugman 1983). Hopefully, these can be disambiguated prior to assertion to keep the prototypicality functions amenable to hill-climbing. Even still, it is possible that hill-climbing cannot find the globally best adjusted image. Maybe we can call on prototypes to help us "jump" out of local maxima. Or perhaps we can detect the need to change representations, like switching to higher resolution, three dimensions, or polar coordinates.

In the simplified domain, objects are structureless points, which we can easily locate. Imaging extended objects introduces a number of interesting complications. First, the objects themselves will be described by possibly categorical predicates. For example, our default image for a tiger could have exactly nineteen stripes, four legs, and a tail three feet long. Similarly, size, posture, and perspective (for projecting onto two dimensions) would have to be determined. Simply finding objects and features, which might have been adjusted, could be an extremely difficult recognition task. However, Kosslyn's computational model gives a simplified method for scanning for objects and features, and perhaps we could exploit the fact that, in images we construct, we have knowledge of what things are where. Additionally, we need to know a very limited number of relevant adjustments for each categorical predicate so the Adjustment Procedure will not degenerate into a full search.

Conclusion

We have presented a computational model of how detailed images can be used to model spatially indeterminate information. The central thesis of the model is that images should be incrementally adjusted to satisfy interacting constraints in the form of previously asserted categorical propositions. By considering the consistency not only of a queried proposition but also its negation with the set of previous assertions, the necessary information in the image can be distinguished from the details which were required for representation but not implied by the description. This computational model specifically relies on graded prototypicality functions which guide the search for the best model of a set of categorical propositions. The use of a grounded language like images for modeling knowledge reduces reasoning with categories to a Constraint Satisfaction Problem.

Acknowledgements

The author thanks Sheldon Nicholl, David C. Wilkins, and the Knowledge-Based Systems Group at the University of Illinois for helpful comments. Additional comments were kindly provided by Gordan Logan, Lois Boggess, and Anthony Maida.

References

Anderson, J. 1978. Arguments Concerning Representations for Mental Imagery. *Psychological Review* 85(4):249-277.

Brugman, C. M. 1983. *The Story of Over*. Bloomington, IN: Indiana University Linguistics Club.

Clark, H. H. 1973. Time, Space, Semantics, and the Child. In Moore, T. E. ed. *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

Funt, B. V. 1980. Problem-Solving with Diagrammatic Representations. *Artificial Intelligence* 13:203-230.

Herskowitz, A. 1986. *Language and Spatial Cognition*. New York: Cambridge University Press.

Johnson-Laird, P. N. 1983. *Mental Models*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M. 1980. *Image and Mind*. Cambridge, MA: Harvard University Press.

Lakoff, G. 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2:458-508.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

Mackworth, A. K. 1990. Constraint Satisfaction. In Shapiro, S. C. ed. *Encyclopedia of Artificial Intelligence*. 2nd edition. New York: John Wiley & Sons.

Pinker, S. 1984. Visual Cognition: An Introduction. *Cognition* 18:1-63.

Pylyshyn, Z. W. 1973. What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery. *Psychological Bulletin* 80:1-24.

Smith, E. and Medin, D. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Waltz, D. L. and Boggess, L. 1979. Visual Analog Representations for Natural Language Understanding. In *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* 926-934.

Winston, P. H. 1984. *Artificial Intelligence*. Reading, MA: Addison-Wesley.

Zadeh, L. 1965. Fuzzy Sets. *Information and Control* 8:338-353.