Nonstandard Set Theories and Information Management

VAROL AKMAN
akman@troy.cs.bilkent.edu.tr
Department of Computer Engineering and Information Science, Bilkent University, Bilkent, Ankara 06533, Turkey

MÜJDAT PAKKAN
pakkan@trboun.bitnet
Department of Computer Engineering, Bosphorus University, Bebek, Istanbul 80815, Turkey

Abstract. The merits of set theory as a foundational tool in mathematics stimulate its use in various areas of artificial intelligence, in particular intelligent information systems. In this paper, a study of various nonstandard treatments of set theory from this perspective is offered. Applications of these alternative set theories to information or knowledge management are surveyed.

Keywords: set theory, knowledge representation, information management, commonsense reasoning, nonwellfounded sets (hypersets)

1. Introduction

Set theory is a branch of modern mathematics with a unique place because various other branches can be formally defined within it. For example, *Book 1* of the influential works of N. Bourbaki is devoted to the theory of sets and provides the framework for the remaining volumes. Bourbaki said in 1949 (Goldblatt, 1984) [1]:

> "[A]ll mathematical theories may be regarded as extensions of the general theory of sets ...[O]n these foundations I can state that I can build up the whole of the mathematics of the present day."

This brings up the possibility of using set theory in foundational studies in AI. McCarthy (1983, 1984) has emphasized the need for fundamental research in AI and claimed that AI needs mathematical and logical theory involving conceptual innovations. In an opening address (McCarthy, 1985), he stressed the feasibility of using set theory in AI because there is considerable beauty, economy, and naturalness in using sets for information modeling and knowledge representation.

In this paper, we first give a brief review of "classical" set theory. We avoid the technical details—which the reader can find in texts like (Halmos, 1974), (Fraenkel et al., 1973), and (Suppes, 1972)—and instead focus on the underlying concepts. While we assume little or no technical background in set theory per se, we hope that the reader is interested in the applications of this formal theory to the problems of intelligent information management. We then consider the alternative set theories which have been proposed to overcome the limitations of the standard theory. Finally, we survey various "nonstandard" treatments of set theory, each innovating different aspects such as urelements, cumulative hierarchy, self-reference, cardinality, well-orderings, and so on. It is shown that such treatments—which are all very recent and sometimes esoteric—are quite useful to the IIS community, for there are assorted technical problems in information management (e.g., commonsense reasoning, terminological logics, etc.) that may profit from such nonstandard approaches.

2. Early developments in set theory

G. Cantor's work on the theory of infinite series should be considered as the foundation of the research in set theory. In Cantor's conception, a *set* is a collection into a whole of definite, distinct objects of our perception or our thought (the *elements* of the set). This property of *definiteness* implies that given a set and an object, it is possible to determine if the object is a member of that set. In other words, a set is completely determined by its members.

In the initial stages of his research, Cantor did not work from axioms. However, all of his theorems can be derived from three axioms: *Extensionality* which states that two sets are identical if they have the same members, *Abstraction* which states that for any given property there is a set whose members are just those entities having that property, and *Choice* which states that if b is a set, all of whose elements are nonempty sets no two of which have any elements in common, then there is a set c which has precisely one element in common with each element of b.

The theory was soon threatened by the introduction of some paradoxes which led to its evolution. In 1902, B. Russell found a contradiction in G. Frege's foundational system which was developed on Cantor's naive conception (van Heijenhoort, 1967). This contradiction could be derived from the Axiom of Abstraction—Axiom V in Frege's system—by considering "the set of all things which have the property of not being members of themselves." This property can be denoted as $\neg(x \in x)$ (or simply $x \notin x$) in the language of first order logic. The Axiom of Abstraction itself can be formulated as

$$\forall x \exists y [x \in y \leftrightarrow \varphi(x)],$$

where $\varphi(x)$ is a formula in which y is free. In the case of Russell's Paradox $\varphi(x) = x \notin x$ and we have: $\forall x \exists y [x \in y \leftrightarrow x \notin x]$. Substituting y for x, we reach the contradiction $y \in y \leftrightarrow y \notin y$.

Another antinomy occurred with the "set of all sets," $\mathrm{V} = \{x : x = x\}$. The well-known Cantor's Theorem states that the power set of V has a greater cardinality than V itself. This is paradoxical since V by definition is the most inclusive set. This is the so-called Cantor's Paradox and led to discussions on the size of comprehensible sets. Strictly speaking, it was Frege's foundational system that was overthrown by Russell's Paradox, not Cantor's naive set theory. The latter came to grief precisely because of the preceding "limitation of size" constraint. Later, J. von Neumann would clarify this problem as follows (Goldblatt, 1984): "Some predicates have extensions that are too large to be successfully encompassed as a whole and treated as a mathematical object."

Paradoxes of the preceding sort were instrumental in new axiomatizations of the set theory and in alternate approaches. However, it is believed that axiomatic set theory would still have evolved in the absence of paradoxes because of the continuous search for foundational principles.

2.1. Alternate axiomatizations

The new axiomatizations took a common step for overcoming the deficiencies of the naive approach by introducing *classes*. NBG, which was proposed by von Neumann (1925) and later revised and simplified by P. Bernays (1937) and Gödel (1940), was the most popular of these. In NBG, there are three primitive notions: set, class, and membership. Classes are considered as totalities corresponding to some, but not necessarily all, properties. The classical paradoxes are avoided by recognizing two types of classes: sets and proper classes. A class is a set if it is a member of some class. Otherwise, it is a proper class. Russell's Paradox is avoided by showing that the class Y $= \{x : x \notin x\}$ is a proper class, not a set. V is also considered as a proper class. The axioms of NBG are simply chosen with respect to the limitation of size constraint.

Other approaches against the deficiencies of the naive approach alternatively played with its language and are generally dubbed *type-theoretical*. Russell and Whitehead's *Theory of Types* is the earliest and most popular of these (Whitehead & Russell, 1910). In this theory, a hierarchy of types is established to forbid circularity and hence avoid paradoxes. For this purpose, the universe is divided into types, starting with a collection M of *individuals*. The elements of M are of type 0. Sets whose members are of type 0 are said to be of type 1, sets whose elements are of type 1 are said to be of type 2, and so on. The membership relation is defined between sets of different types, e.g., $x^n \in y^{n+1}$. Therefore, $x \notin x$ is not even a valid formula in this theory and Russell's Paradox is trivially avoided.

Similar to the Theory of Types is Quine's *New Foundations* (NF) which he invented to overcome some unpleasant aspects of the former (Quine, 1937). NF uses only one kind of variable and a binary predicate letter $\in$ for membership. A notion called *stratification* is introduced to maintain the hierarchy of types [2]. In NF, Russell's Paradox is avoided as in the Theory of Types, since the problematic wff is not stratified.

2.2. ZF set theory

Zermelo-Fraenkel (ZF) is the earliest axiomatic system in set theory. The first axiomatization was by E. Zermelo (1908). A. A. Fraenkel (1922) observed a weakness of Zermelo's system and proposed a way to overcome it. His proposal was reformulated by T. Skolem (1922) by introducing a new axiom. ZF is carried out in a language which includes sets as objects and $\in$ for membership. Equality is defined externally by the *Axiom of Extensionality* which states that two sets are equal if and only if (iff) they have the same elements.

ZF's essential feature is the *cumulative hierarchy* it proposes (Parsons, 1977). The intention is to build up mathematics by starting with $\emptyset$ and then construct further sets in a stepwise manner by various defined operators. Hence there are no individuals (urelements) in the universe of this theory. The cumulative hierarchy works as follows (Tiles, 1989).

The *Null Set Axiom* guarantees that there is a set with no elements, i.e., $\emptyset$. This is the only set whose existence is explicitly stated. The *Pair Set Axiom* states the existence of a set which has a member when the only existing set is $\emptyset$. So the set $\{\emptyset\}$ can now be formed now and we have two objects $\emptyset$ and $\{\emptyset\}$. The application of the axiom repetitively yields any finite number of sets, each with only one or two elements. It is the *Sum Set Axiom* which states the existence of sets containing any finite number of elements by defining the union of already existing sets. Thus $\bigcup\{\{\emptyset,\{\emptyset\}\},\{\{\emptyset,\{\emptyset\}\}\}\} = \{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\}$. However it should be noted that all these sets will be finite because only finitely many sets can be formed by applying Pair Set and Sum Set finitely many times. It is the *Axiom of Infinity* which states the existence of at least one infinite set, from which other infinite sets can be formed. The set which the axiom asserts to exist is $\{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\},\{\emptyset,\{\emptyset\},\{\emptyset,\{\emptyset\}\}\},\ldots\}$.

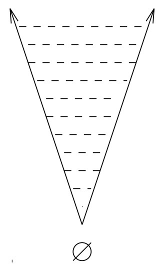

Figure 1: The cumulative hierarchy

The cumulative hierarchy is depicted in Figure 1. Thus, the ZF universe simply starts with the $\emptyset$ and extends to infinity. It can be noticed that cumulative hierarchy produces all finite sets and many infinite ones, but it does not produce all infinite sets (e.g., V).

While the first five axioms of ZF are quite obvious, the *Axiom of Foundation* cannot be considered so. The axiom states that every set has elements which are minimal with respect to membership, i.e., no infinite set can contain an infinite sequence of members $\ldots \in x_3 \in x_2 \in x_1 \in x_0$. Infinite sets can only contain sets which are formed by a finite number of iterations of set formation. Hence this axiom forbids the formation of sets which require an infinity of iterations of an operation to form sets. It also forbids sets which are members of themselves, i.e., circular sets. Russell's Paradox is avoided since the problematic set $x = \{x\}$ cannot be shown to exist [3]. The *Axiom of Separation* makes it possible to collect together all the sets belonging to a set whose existence has already been guaranteed by the previous axioms and which satisfy a property φ:

$$\forall x \exists u [x \in u \leftrightarrow x \in v \,\&\, \varphi(x)].$$

The axiom does not allow to simply collect all the things satisfying a given meaningful description together into a set, as assumed by Cantor. It only gives permission to form subsets of a set whose existence is already guaranteed. It also forbids the universe of sets to be considered as a set, hence avoiding the Cantor's Paradox of the set of all sets. The *Axiom of Replacement* is a stronger version of the Axiom of Separation. It allows the use of functions for the formation of sets but still has the restriction of the original Axiom of Separation. It should be noted that these two axioms are in fact not single axioms but *axiom schemes*. Therefore, ZF is not finitely axiomatizable [4].

The *Power Axiom* states the existence of the set of all subsets of a previously defined set. The formal definition of the power operation, P, is $P(x) = \{y : y \subseteq x\}$. The Power Axiom is an important axiom, because Cantor's notion of an infinite number was inspired by showing that for any set, the cardinality of its power set must be greater than its cardinality.

The *Axiom of Choice* is not considered as a basic axiom and is explicitly stated when used in a proof. ZF with the Axiom of Choice is known as ZFC.

It should be noted that the informal notion of cumulative hierarchy summarized above has a formal treatment. The class WF of *wellfounded* sets is defined recursively in ZF starting with $\emptyset$ and iterating the power set operation P where a *rank* function $R(\alpha)$ is defined for $\alpha \in Ord$, the class of all ordinals [5]:

- $R(0) = \emptyset$,

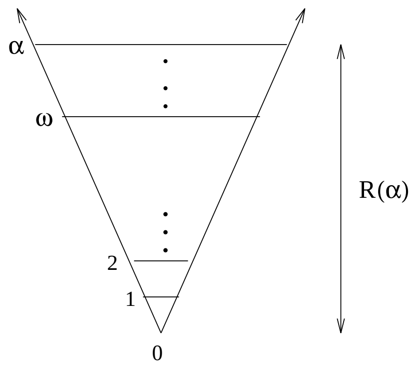

Figure 2: The WF universe in terms of ordinals

- $R(\alpha + 1) = P(R(\alpha))$,
- $R(\alpha) = \bigcup_{\beta<\alpha} R(\beta)$ when α is a limit ordinal,
- $\mathrm{WF} = \bigcup\{R(\alpha) \ : \ \alpha \in Ord\}$.

The WF universe is depicted in Figure 2 which bears a resemblance to Figure 1. This is justified by the common acceptance of the statement that the universe of ZF is equivalent to the universe of WF.

ZF and NBG produce essentially equivalent set theories, since it can be shown that NBG is a *conservative extension* of ZF, i.e., for any sentence φ, if ZF $\models \varphi$, then NBG $\models \varphi$. The main difference between the two is that NBG is finitely axiomatizable, whereas ZF is not. Still, most of the current research in set theory is being carried out in ZF. Nevertheless, ZF has its own drawbacks (Barwise, 1975). While the cumulative hierarchy provides a precise formulation of many mathematical concepts, it may be asked whether it is limiting, in the sense that it might be omitting some interesting sets one would like to have around, e.g., circular sets. Clearly, the theory is weak in applications involving self-reference because circular sets are prohibited by the Axiom of Foundation.

Strangely enough, ZF is too strong in some ways. Important differences on the nature of the sets defined in it are occasionally lost. For example, being a prime number between 6 and 12 is a different property than being a solution to $x^2 - 18x + 77 = 0$, but this distinction disappears in ZF. Besides, the *Principle of Parsimony*, which states that simple facts should have simple proofs, is quite often violated in ZF (Barwise, 1975). For example, verification of a trivial fact like the existence in ZF of $a \times b$, the set of all ordered pairs $\langle x, y\rangle$ such that $x \in a$ and $y \in b$, relies on the Power Set Axiom [6].

3. Alternate approaches

3.1. Admissible sets

Admissible sets are formalized in a first order set theory called *Kripke-Platek* (KP). Barwise weakened KP by readmitting the *urelements* and called the resulting system KPU (Barwise, 1975). Urelements are the objects (or individuals) with no elements, i.e., they can occur on the left of $\in$, but not on the right. They are not considered in ZF because ZF

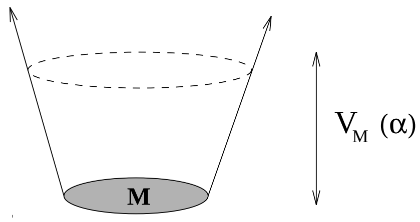

Figure 3: The universe of admissible sets

is strong enough to live without them. But since KPU is a weak version of KP, Barwise decided to include them.

KPU is formulated in a first order language L with equality and $\in$. It has six axioms. The axioms of *Extensionality* and *Foundation* are about the basic nature of sets. The axioms *Pair*, *Union*, and Δ_0-*Separation* [7] treat the principles of set construction. These five axioms can be taken as corresponding to ZF axioms of the same interpretation. The important axiom of Δ_0-*Collection* assures that there are enough stages in the (hierarchical) construction process.

The universe of admissible sets over an arbitrary collection M of urelements is defined recursively:

- $V_M(0) = \emptyset$,
- $V_M(\alpha + 1) = P(M \cup V_M(\alpha))$,
- $V_M(\lambda) = \bigcup_{\alpha<\lambda} V_M(\alpha)$, if λ is a limit ordinal,
- $\mathrm{V}_M = \bigcup_\alpha V_M(\alpha)$.

This universe is depicted in Figure 3, adapted from Barwise (1977). It should be noticed that the KPU universe is like the ZF universe (excluding the existence of urelements), since it supports the same idea of cumulative hierarchy.

If M is a structure [8] for L, then an *admissible set over* M is a model U_M of KPU of the form $U_M = (M; A, \in)$, where A is a nonempty set of nonurelements and $\in$ is defined in $M \times A$. A *pure admissible set* is an admissible set with no urelements, i.e., it is a model of KP.

KPU is an elegant theory which supports the concept of cumulative hierarchy and respects the principle of parsimony. (The latter claim will be proved in the sequel.) But it still cannot deal with self-reference because of its hierarchical nature.

3.2. Hypersets

It was D. Mirimanoff (1917) who first stated the fundamental difference between wellfounded and nonwellfounded sets. He called sets with no infinite descending membership sequence wellfounded and others nonwellfounded. Nonwellfounded sets have been extensively studied through decades, but did not show up in notable applications until Aczel (1988). This is probably due to the fact that the classical wellfounded universe was a rather satisfying domain for the practicing mathematician—"the mathematician in the street" (Barwise, 1985). Aczel's work on nonwellfounded sets evolved from his interest in

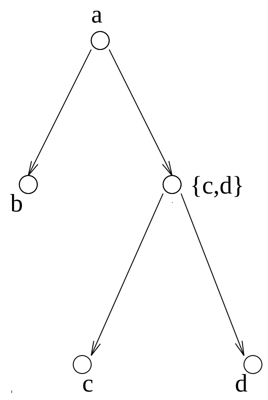

Figure 4: $a = \{b, \{c, d\}\}$ in hyperset notation

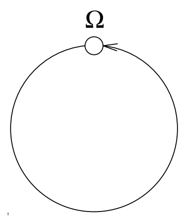

Figure 5: The Aczel picture of Ω

modeling concurrent processes. He adopted the graph representation for sets to use in his theory. A set $a = \{b, \{c, d\}\}$ can be unambiguously depicted as in Figure 4 in this representation (Aczel, 1988), where an arrow from a node x to a node y denotes the membership relation between x and y (i.e., $x \in y$).

A set (pictured by a graph) is called wellfounded if it has no infinite paths or cycles, and nonwellfounded otherwise. Aczel's *Anti-Foundation Axiom* (AFA) states that every graph, wellfounded or not, pictures a unique set. Removing the Axiom of Foundation (FA) from the ZFC and adding the AFA results in the *Hyperset Theory* or ZFC$^-$/AFA. (ZFC without the FA is denoted as ZFC$^-$.) What is advantageous with the new theory is that since graphs of arbitrary form are allowed, including the ones containing proper cycles, one can represent self-referring sets (Barwise, 1992). For example, the graph in Figure 5 is the picture of the unique set $\Omega = \{\Omega\}$.

The picture of a set can be *unfolded* into a tree picture of the same set. The tree whose nodes are the finite paths of the apg [9] which start from the point of the apg, whose edges are pairs of paths $\langle n_0 \longrightarrow \cdots \longrightarrow n\,,\, n_0 \longrightarrow \cdots \longrightarrow n \longrightarrow n' \rangle$, and whose root is the path n_0 of length one is called the *unfolding* of that apg. The unfolding of an apg always pictures any set pictured by that apg. Unfolding the apg in Figure 5 results in an infinite tree, analogous to $\Omega = \{\{\{\cdots\}\}\}$.

According to Aczel's conception, for two sets to be different, there should be a genuine structural difference between them. Therefore, all of the three graphs in Figure 6 depict the unique nonwellfounded set Ω.

Aczel develops his own Extensionality concept by introducing the notion of bisimula-

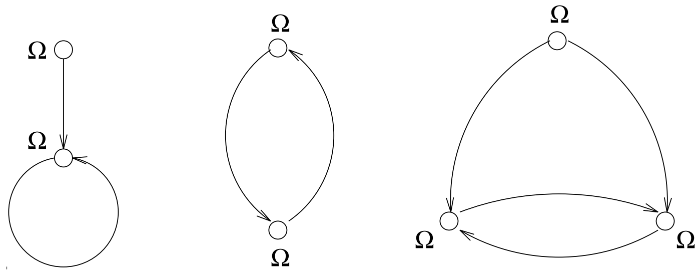

Figure 6: Other pictures of Ω

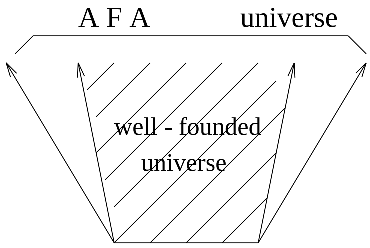

Figure 7: The AFA universe

tion. A *bisimulation* between two apg's, G_1 with point p_1 and G_2 with point p_2, is a relation $R \subseteq G_1 \times G_2$ satisfying the following conditions:

1. $p_1 R p_2$
2. if nRm then
   - for every edge $n \longrightarrow n'$ of G_1, there exists an edge $m \longrightarrow m'$ of G_2 such that $n'Rm'$
   - for every edge $m \longrightarrow m'$ of G_2, there exists an edge $n \longrightarrow n'$ of G_1 such that $n'Rm'$

Two apg's G_1 and G_2 are said to be *bisimilar*, if a bisimulation exists between them; this means that they picture the same sets. It can be concluded that a set is completely determined by any graph which pictures it [10].

The AFA universe can be depicted as in Figure 7, extending around the wellfounded universe, because it includes the nonwellfounded sets which are not covered by the latter.

3.2.1. Equations in the AFA universe. Aczel's theory includes another important useful feature: *equations* in the universe of Hypersets.

Let $\mathcal{V}_A$ be the universe of hypersets with atoms from a given set A and let $\mathcal{V}_{A'}$ be the universe of hypersets with atoms from another given set A' such that $A \subseteq A'$ and X is defined as $A' - A$. The elements of X can be considered as *indeterminates* ranging over the universe $\mathcal{V}_A$. The sets which can contain atoms from X in their construction are called *X-sets*. A *system of equations* is a set of equations

$$\{x = a_x \ : \ x \in X \ \wedge \ a_x \text{ is an } X\text{-}set\}$$

for each $x \in X$. For example, choosing $X = \{x, y, z\}$ and $A = \{\mathrm{C}, \mathrm{M}\}$ (thus $A' = \{x, y, z, \mathrm{C}, \mathrm{M}\}$), consider the system of equations

$$x = \{\mathrm{C}, y\},$$
$$y = \{\mathrm{C}, z\},$$
$$z = \{\mathrm{M}, x\}.$$

A *solution* to a system of equations is a family of pure sets b_x (sets which can have only sets but no atoms as elements), one for each $x \in X$, such that for each $x \in X$, $b_x = \pi a_x$. Here, π is a *substitution operation* (defined below) and πa is the pure set obtained from a by substituting b_x for each occurrence of an atom x in the construction of a.

The *Substitution Lemma* states that for each family of pure sets b_x, there exists a unique operation π which assigns a pure set πa to each X-set a, viz.

$$\pi a = \{\pi b \ : \ b \text{ is an } X\text{-}set \text{ such that } b \in a\} \cup \{\pi x \ : \ x \in a \cap X\}.$$

The *Solution Lemma* can now be stated (Barwise & Moss, 1991) [11]. If a_x is an X-set, then the system of equations $x = a_x$ has a unique solution, i.e., a unique family of pure sets b_x such that for each $x \in X$, $b_x = \pi a_x$.

This lemma can be stated somewhat differently (Pakkan, 1993). Letting X again be the set of indeterminates, g a function from X to $P(X)$, and h a function from X to A, there exists a unique function f for all $x \in X$ such that

$$f(x) = \{f(y) \ : \ y \in g(x)\} \cup h(x).$$

Obviously, $g(x)$ is the set of indeterminates and $h(x)$ is the set of atoms in each *X-set* a_x of an equation $x = a_x$. In the above example, $g(x) = \{y\}$, $g(y) = \{z\}$, $g(z) = \{x\}$, and $h(x) = \{\mathrm{C}\}$, $h(y) = \{\mathrm{C}\}$, $h(z) = \{\mathrm{M}\}$, and one can compute the solution

$$f(x) = \{\mathrm{C}, \{\mathrm{C}, \{\mathrm{M}, x\}\}\},$$
$$f(y) = \{\mathrm{C}, \{\mathrm{M}, \{\mathrm{C}, y\}\}\},$$
$$f(z) = \{\mathrm{M}, \{\mathrm{C}, \{\mathrm{C}, z\}\}\}.$$

As another example due to (Barwise & Etchemendy, 1987), it may be verified that the system of equations

$$x = \{C, M, y\},$$

$$y = \{M, x\},$$

$$z = \{x, y\}.$$

has a unique solution in the universe of Hypersets depicted in Figure 8 with $x = a$, $y = b$, and $z = c$.

This technique of solving equations in the universe of hypersets can be useful in modeling information which can be cast in the form of equations (Akman & Pakkan, 1993), e.g., situation theory (Barwise & Perry, 1983), databases, etc. since it allows us to assert the existence of some graphs (the solutions of the equations) without having to depict them with graphs. We now give an example from databases.

3.2.2. AFA and relational databases. Relational databases embody data in tabular forms and show how certain objects stand in certain relations to other objects. As an

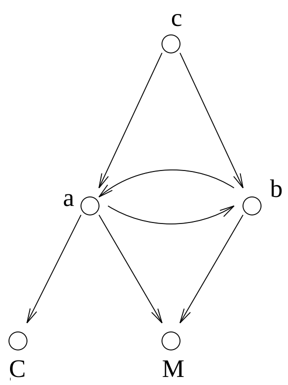

Figure 8: The solution to a system of equations

FatherOf

John	Bill
John	Kitty
Tom	Tim

MotherOf

Sally	Tim
Kathy	Bill
Kathy	Kitty

BrotherOf

Bill	Kitty

Figure 9: A relational database

SizeOf

FatherOf	3
MotherOf	3
BrotherOf	1
SizeOf	4

Figure 10: The SizeOf relation for the preceding database

example adapted from (Barwise, 1990), the database in Figure 9 includes three binary relations: FatherOf, MotherOf, and BrotherOf. A *database model* is a function M with domain some set Rel of binary relation symbols such that for each relation symbol $R \in Rel$, R^M is a finite binary relation that holds in model M.

If one wants to add a new relation symbol SizeOf to this database, then $Rel' = Rel \cup \{\text{SizeOf}\}$. A database model M for Rel' is *correct* if the relation SizeOf^M contains all pairs $\langle R, n\rangle$ where $R \in Rel$ and $n = |R|$, the cardinality of R. Such a relation can be seen in Figure 10. It may appear that every database for Rel can be extended in a unique way to a correct database for Rel'. Unfortunately, this is not so.

Assuming the FA, it can be shown that there are no correct database models. Because if M is correct, then the relation SizeOf stands in relation SizeOf to n, denoted by SizeOf SizeOf n. But this is not true in ZFC because otherwise $\langle \text{SizeOf}, n\rangle \in \text{SizeOf}$.

If Hyperset Theory is used as the meta-theory instead of ZFC in modeling such databases, then the solution of the equation

$$x = \{\langle R^M, |R^M|\rangle \ : \ R \in Rel\} \ \cup \ \{\langle x, |Rel| + 1\rangle\}$$

(which can be found by applying the Solution Lemma) is the desired SizeOf relation.

4. Commonsense set theory

If we want to design artificial systems which will work in the real world, they must have a good knowledge of that world and be able to make inferences out of their knowledge. The common knowledge which is possessed by any child and the methods of making inferences from this knowledge are known as *common sense*. Any intelligent task requires it to some degree and designing programs with common sense is one of the most important problems in AI. McCarthy (1959) claims that the first task in the construction of a general intelligent program is to define a naive commonsense view of the world precisely enough, but also adds that this is a very difficult thing. He states that "a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows."

It appears that in commonsense reasoning a concept can be considered as an indivisible unit, or as composed of other parts, as in mathematics. Relationships, again as in mathematics, can also be represented with sets. For example, the notion of "society" can be considered to be a relationship between a set of people, rules, customs, traditions, etc. What is problematic here is that commonsense ideas do not have very precise definitions since the real world is too imprecise. For example, consider the definition of "society" (adapted from *Webster's Ninth New Collegiate Dictionary*):

"Society gives people having common traditions, institutions, and collective activities and interests a choice to come together to give support to and be supported by each other and continue their existence."

It should be noted that the notions "tradition," "institution," and "existence" also appear to be as complex as the definition itself. So this definition should probably better be left to the experience of the reader with all these complex entities.

Nevertheless, whether or not a set theoretical definition is given, sets are useful for conceptualizing commonsense terms. For example, we may want to consider the set of "traditions" disjoint from the set of "laws" (one can quickly imagine two separate circles of a Venn diagram). We may not have a wellformed formula (wff) which defines either of these sets. Such a formation process of collecting entities for further thought is still important and simply corresponds to the set formation process of formal set theories, i.e., the comprehension principle.

Having decided to investigate the use of sets in commonsense reasoning, we have to concentrate on the properties of such a theory. Instead of directly checking if certain set-theoretic technicalities have a place in our theory, we first look from the commonsense reasoning point of view and examine the set-theoretic principles which cannot be excluded from such reasoning.

4.1. Desirable properties

We first begin with the general principles of set formation. The first choice is whether to allow urelements. This seems like the right thing to do because in a naive sense, a set is a collection of individuals satisfying a property. This is what exactly corresponds to the unrestricted Comprehension Axiom of Cantor. However, we have seen that this leads to Russell's Paradox in ZF. The problem arises when we use a set whose completion is not over yet in the formation of another set, or even in its own formation. Then we are led to the question when the collection of all individuals satisfying an expression can be considered an individual itself. Since we are talking about the individuals as entities formed out of previously formed entities, the notion of cumulative hierarchy immediately comes to mind.

The cumulative hierarchy is one of the most common construction mechanisms of our intuition and is supported by many existing theories, viz. ZF and KPU. It can be illustrated by the hierarchical construct in Figure 11, where we have bricks as individuals, and make towers out of bricks, and then make walls out of towers, and so on (Perlis, 1988). In the cumulative hierarchy, any set formed at some stage must be consisting of the urelements (if included in the theory) and the sets which have been formed at some previous stage (Shoenfield, 1977) (but not necessarily at the preceding stage, as in Russell's Type Theory).

At this point, the problem of sets which can be members of themselves arises, since such sets are used in their own formation. Circularity is obviously a common means of commonsense knowledge representation. For example, nonprofit organizations are sets of individuals and the set of all nonprofit organizations is also a set; all these are expressible in the cumulative hierarchy. But what if the set of all nonprofit organizations wants to be a member of itself, since it also is a nonprofit organization? This is not an unexpected event (Figure 12) because this umbrella organization may benefit from having the status of a nonprofit organization (e.g., tax exemption, etc.) (Perlis, 1988).

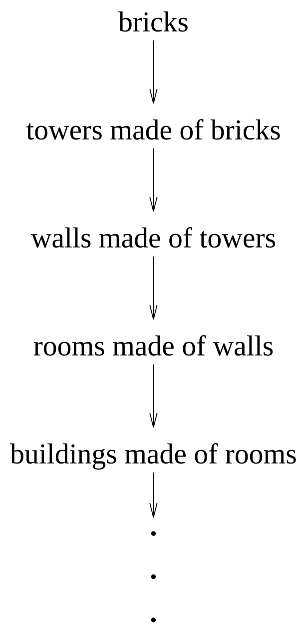

Figure 11: A construction à la cumulative hierarchy

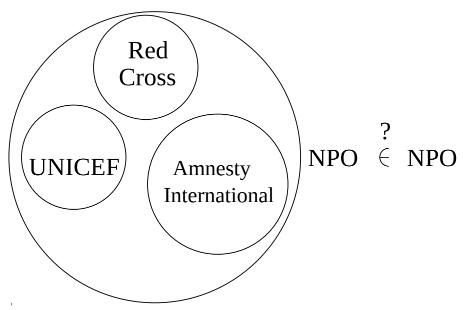

Figure 12: Nonprofit organizations

Thus we conclude that a possible commonsense set theory should also allow circular sets. This is an important issue in representation of meta-knowledge and is addressed in (Feferman, 1984) and (Perlis, 1985). In these references, a method which *reifies*—creates a syntactic term from a predicate expression—a wff into a name for the wff asserting that the name has strong relationship with the formula, is presented. In this way, any set of wff are matched with a set of names of wff, thereby allowing self-reference by the use of names. In (Feferman, 1984), the need for *type-free* (admitting instances of self-application) frameworks for semantics is especially emphasized. However, such formalizations which also capture the cumulative hierarchy principle are not very common. Among theories revised so far, Aczel's theory is the only one which allows circularity. By proposing his Anti-Foundation Axiom, Aczel overrode the FA of ZF which prohibits circular sets, but preserved the hierarchical nature of the original axiomatization.

4.2. Applications

4.2.1. Situations for knowledge representation. Situations are parts of the reality that can enter into relations with other parts (Barwise & Perry, 1983). Their internal structures are sets of facts and hence they can be modeled by sets. There has been a considerable deal of work on this especially by Barwise himself (Barwise, 1989a). He used his Admissible Set Theory (Barwise, 1975) as the principal mathematical tool in the beginning. However, in the handling of circular situations, he was confronted with problems and then discovered that Aczel's theory could be a solution (Barwise, 1989c). Circular situations are common in our daily life. For example consider the situation in which we utter the statement "This is a boring situation." While we are referring to a situation, say s, by saying "this situation," our utterance is also a part of that situation. As another example, one sometimes hears radio announcements concluding with "This announcement will not be repeated." If announcements are assumed to be situations, then this one surely contains itself.

Barwise defined the operation M (to model situations with sets) taking values in hypersets and satisfying [12]:

- if b is not a situation or state of affairs, then $M(b) = b$,
- if $\sigma = \langle R, a, i\rangle$, then $M(\sigma) = \langle R, b, i\rangle$ (which is called a *state model*), where b is a function on the domain of a satisfying $b(x) = M(a(x))$,
- if s is a situation, then $M(s) = \{M(\sigma) : s \models \sigma\}$.

Using this operation, Barwise proves that there is no largest situation (corresponding to the absence of a universal set in ZF).

We also see a treatment of self-reference in (Barwise & Etchemendy, 1987), where the authors concentrate on the concept of truth. In this study, two conceptions of truth are examined, primarily on the basis of the notorious Liar Paradox [13]. The authors make use of Aczel's theory for this purpose. A statement like "This sentence is not expressible in English in ten words" would be represented in Aczel's theory as in Figure 13 (adapted from (Barwise & Etchemendy, 1987)), where $\langle E, p, i\rangle$ denotes that the proposition p has the property E if $i = 1$, and it does not have it if $i = 0$ (which is the case in Figure 13 if we take E to be the property of "being expressible in English in ten words").

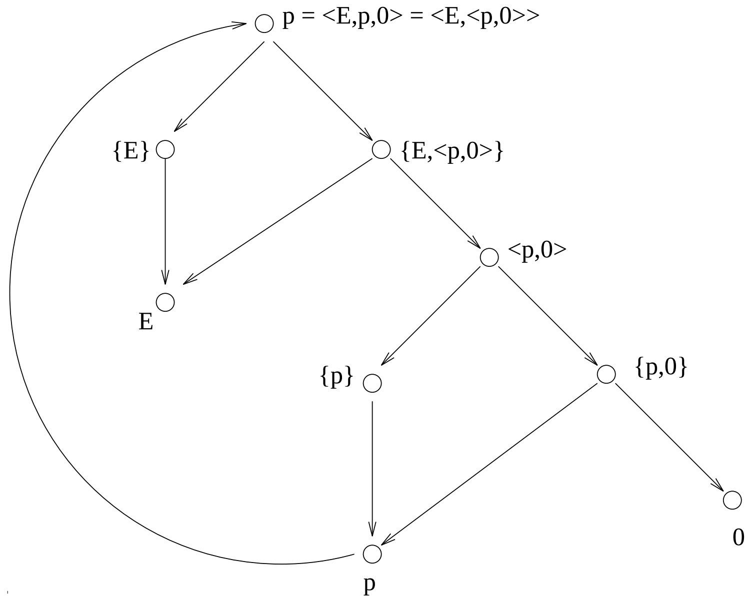

Figure 13: The Aczel picture of "This sentence is not expressible in English in ten words"

4.2.2. Common knowledge (mutual information). Two card players P_1 and P_2 are given some cards such that each gets an ace. Thus, both P_1 and P_2 know that the following is a fact:

$$\sigma\ :\ \text{Either } P_1 \text{ or } P_2 \text{ has an ace.}$$

When asked whether they knew if the other one had an ace or not, they both would answer "no." If they are told that at least one of them has an ace and asked the above question again, first they both would answer "no." But upon hearing P_1 answer "no," P_2 would know that P_1 has an ace. Because, if P_1 does not know P_2 has an ace, having heard that at least one of them does, it can only be because P_1 has an ace. Obviously, P_1 would reason the same way, too. So, they would conclude that each has an ace. Therefore, being told that at least one of them has an ace must have added some information to the situation. How can being told a fact that each of them already knew increase their information? This is known as *Conway's Paradox*. The solution relies on the fact that initially σ was known by each of them, but it was not *common knowledge*. Only after it became common knowledge, it gave more information (Barwise, 1989d).

Hence, common knowledge can be viewed as iterated knowledge of σ of the following form: P_1 knows σ, P_2 knows σ, P_1 knows P_2 knows σ, P_2 knows P_1 knows σ, and so on. This iteration can be represented by an infinite sequence of facts (where K is the relation "knows" and s is the situation in which the above game takes place, hence $\sigma \in s$): $\langle K, P_1, s\rangle$, $\langle K, P_2, s\rangle$, $\langle K, P_1, \langle K, P_2, s\rangle\rangle$, $\langle K, P_2, \langle K, P_1, s\rangle\rangle$, ... However, considering the system of equations

$$x = \{\langle K, P_1, y\rangle, \langle K, P_2, y\rangle\},$$
$$y = s \cup \{\langle K, P_1, y\rangle, \langle K, P_2, y\rangle\},$$

the Solution Lemma asserts the existence of the unique sets s' and$s \cup s'$ satisfying these

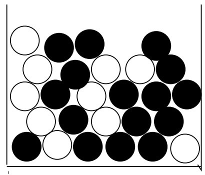

Figure 14: Finite cardinality without well-ordering

equations, respectively, where

$$s' = \{\langle K, P_1, s \cup s' \rangle, \langle K, P_2, s \cup s' \rangle\}.$$

Then, the fact that s is common knowledge can more effectively be represented by s' which contains just two infons and is circular.

4.2.3. Possible membership. One further aspect to be considered is "possible" membership which might have many applications, mainly in language oriented problems. This concept can be handled by introducing partial functions—functions which might not have corresponding values for some of their arguments. A commonsense set theory may be helpful in providing representations for dynamic aspects of language by making use of partiality. For example, partiality has applications in modality (the part of linguistics which deals with modal sentences, i.e., sentences of necessity and possibility), dynamic processing of syntactic information, and situation semantics (Mislove et al., 1990).

We had mentioned above that situations can be modeled by sets. Consider a situation s in which you have to guess the name of a boy, viz.

$s \models$ *The boy's name is Jon or the boy's name is John.*

This situation can be modeled by a set of two states of affairs. The problem here is that neither assertion about the name of the boy can be assured on the basis of s (because of the disjunction). A solution to this problem is to represent this situation as a *partial set*, one with two "possible" members. In this case s still supports the disjunction above but does not have to support either specific assertion. Another notion called *clarification*, which is a kind of general information-theoretic ordering, helps determine the real members among possible ones. If there exists another situation s', where $s' \models$ *The boy's name is Jon*, then s' is called a clarification of s.

4.2.4. Cardinality and well-ordering. Imagine a box of 16 black and 10 white balls (Figure 14). We know that there are 26 balls in the box, or formally, the cardinality of the set of balls in the box is 26. After shaking the box, we would say that that the balls in the box are not ordered any more, or again formally, the set of the balls does not have a well-ordering. This, however, is not true in classical set theory, because a set with finite cardinality must have a well-ordering (Zadrozny, 1989).

While the formal principles of counting are precise enough for mathematics, we can observe that people also use other quantifiers like "many" or "more than half" for counting purposes in daily speech. For example, if asked about the number of balls in the box in Figure 14, one might have simply answered "Many balls!" So, at least in principle, different counting methods can be developed for commonsense reasoning. It is natural to expect, for example, that a system which can represent a statement like "A group of kids are shouting" should probably decline to answer questions such as "Who is the first one?" (Zadrozny, 1989).

We also expect our theory to obey the parsimony principle. This is a very natural expectation from a commonsense set theory. We have observed that the proof of the existence of a simple fact like the Cartesian product of two sets $a \times b$, required the use of the Power Axiom in ZF [6]. The set obtained in this manner just consists of pairs formed of one element of the set a and one element of the set b. To prove this, the strong Power Axiom should not be necessary. We observe this in KPU set theory where the proof is obtained via definitions and simple axioms [14] (Barwise, 1975).

5. Applications of nonstandard theories in IIS

Cowen (1993) showed, in a landmark paper, that researchers in both set theory and computer science are studying similar objects: graphs, conjunctive normal forms, set systems, etc. and their interrelations. The moral is, he believes, advances in one area can then often be of use in the other. Gilmore (1993) provides evidence, within the framework of his NaDSet theory, that the absolute character of ZF or NBG set theories is not easy to support, cf. his remark: "The ad hoc character of the axioms of Zermelo-Fraenkel set theory and the equivalent class-set theory of Gödel-Bernays naturally leads to skepticism about their fundamental role in mathematics." Set theory has also been the subject of research in automated theorem proving. Brown (1978) gave a deductive system for elementary set theory which is based on truth-value preserving transformations. Quaife (1992) presented a new clausal version of NBG, comparing it with the one given in (Boyer et al., 1986), and claimed that automated development of set theory could be improved. We will now review some essential research efforts, by Perlis, Zadrozny, Mislove et al., Dionne et al., Barwise, and Yasukawa and Yokota.

5.1. Perlis's work

Perlis's approach was to develop a series of theories towards a complete commonsense set theory (Perlis, 1988). He first proposed an axiom scheme of set formation for a naive set theory which he named CST_0:

$$\exists y \forall x [x \in y \leftrightarrow \phi(x) \,\&\, Ind(x)].$$

Here ϕ is any formula and Ind is a predicate symbol with the intended extension "individuals." However, Ind can sometimes be critically rich, i.e., if ϕ is the same with Ind itself, then y may be too large to be an individual. (This is the case of Cantor's Axiom of Abstraction.) Therefore, a theory for a hierarchical extension for Ind is required. To support the cumulative hierarchy, Perlis extended this theory to a new one called CST_1 using the so-called Ackermann's Scheme:

$$HC(y_1) \& \ldots \& HC(y_n) \,\&\, \forall x[\phi(x) \rightarrow HC(x)] \rightarrow \exists z[HC(z) \,\&\, \forall x[x \in z \leftrightarrow \phi(x)]]$$

Here $HC(x)$ can be interpreted as "x can be built up as a collection from previously obtained entities." CST_1 is consistent with respect to ZF. Unfortunately, it cannot deal with self-referring sets.

Perlis finally proposed CST_2 which is a synthesis of the universal reflection theory of Gilmore-Kripke (Gilmore, 1974), which forms entities regardless of their origins and self-referential aspects. GK set theory has the following axiom scheme where each wff $\alpha(x)$ has a *reification* $[\alpha(x)]$ with variables free as in α and distinguished variable x

$$y \in [\alpha(x)] \leftrightarrow \alpha^*(y)$$

where y does not appear in α [15]. There is also a definitional equivalence (denoted by $\approx$) axiom:

$$w \approx z \leftrightarrow \forall x(x \in w \leftrightarrow x \in z).$$

GK is consistent with respect to ZF (Perlis, 1985). Perlis then proposed the following axioms to augment GK:

$$(\text{Ext1}) \quad x \approx y \leftrightarrow ext\ x = ext\ y$$

$$(\text{Ext2}) \quad x \approx ext\ x$$

$$(\text{Ext3}) \quad x \in HC \rightarrow \exists y(x = ext\ y)$$

$$(\text{A}_{ext}) \quad y_1, \ldots, y_n \in HC \,\&\, \forall x(\phi x \rightarrow x \in HC) \rightarrow ext[\phi] \in HC \,\&\, \forall x(x \in [\phi] \leftrightarrow \phi(x))$$

These axioms provide extensional constructions, i.e., collections determined only by their members. Thus, while GK provides the representation of circularity, these axioms support the cumulative construction mechanism. This theory can deal with problems like nonprofit organization membership described earlier (Perlis, 1988).

5.2. Zadrozny's work

Zadrozny does not believe in a "super theory" of commonsense reasoning about sets, but rather in commonsense theories involving different aspects of sets. He thinks that these can be separately modeled in an existing set theory. In particular, he proposed a representation scheme based on Barwise's KPU for cardinality functions, hence distinguishing reasoning about well-orderings from reasoning about cardinalities and avoiding the box problem mentioned earlier (Zadrozny, 1989).

Zadrozny interprets sets as directed graphs and does not assume the FA. A graph in his conception is a triple $\langle V, SE, E\rangle$ where V is a set of vertices, SE is a set of edges, and E is a function from a subset of SE into $V \times V$. It is assumed that $x \in y$ iff there exists an edge between x and y. He defines the edges corresponding to the members of a set as

$$EM(s) = \{e \in ES : \exists v[E(e) = (v, s)]\}.$$

In classical set theory, the cardinality of a finite set s is a one-to-one function from a natural number n onto a set, i.e., a function from a number onto the nodes of the graph of the set. However, Zadrozny defines the cardinality function as a one-to-one order preserving mapping from the edges $EM(s)$ of a set s into the numerals $Nums$ (an entity of numerals which is linked with sets by existence of a counting routine denoted by #, and which can take values like $1, 2, 3, 4$, or $1, 2, 3,$ *about-five*, or $1, 2, 3, many$). The last element of the range of the function is the cardinality. The representation of the four element set $k = \{a, b, \{x, y\}, d\}$ with three atoms and one two-atom set is shown in Figure 15, adapted from (Zadrozny, 1989). The cardinality of the set is *about–five*, i.e., the last element of $Nums$ which is the range of the mapping function from the edges of the set. (The cardinality might well be 4 if $Nums$ was defined as $1, 2, 3, 4$.) Zadrozny then proves two

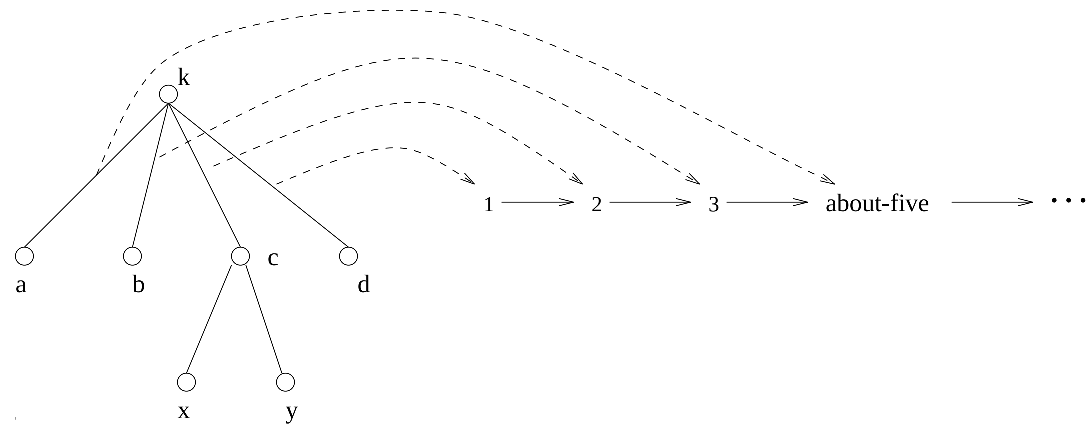

Figure 15: One-to-one order preservation

important theorems in which he shows that there exists a set x with n elements which does not have a well-ordering and there exists a well-ordering of type n, i.e., with n elements, the elements of which do not form a set.

More recent work of Zadrozny treating different aspects of computational mereology vis-*à*-vis set theory can be found in (Zadrozny & Kim, 1993). In this work, it is shown that it is possible to write formal specifications for the process of indexing or retrieving multimedia objects using set theory and mereology. Mereology is understood as a theory of parts and wholes, viz. it tries to axiomatize the properties of the relation PartOf. The paper argues that both set theory and mereology have specific roles to play in the theory of multimedia, for they both provide us with a language to address multimedia objects, albeit with complementary roles: the former is used to index, the latter to describe knowledge about indexes. This observation leads to a simple formalism regarding the storage and retrieval of multimedia material.

5.3. Mislove et al.'s work

Mislove, Moss, and Oles (1990, 1991) developed a partial set theory, ZFAP, based on *protosets*, which is a generalization of HF—the set of wellfounded hereditarily finite sets [16]. A protoset is like a wellfounded set except that it has some kind of *packaging* which can hide some of its elements. There exists a protoset $\perp$ which is empty except for packaging. From a finite collection $x_1, \ldots, x_n$, one can construct the *clear* protoset $\{x_1, \ldots, x_n\}$ which has no packaging, and the *murky* protoset $[x_1, \ldots, x_n]$ which has some elements, but also packaging. For example, a murky set like $[2, 3]$ contains 2 and 3 as elements, but it might contain other elements, too. We say that x is *clarified* by y, $x \sqsubseteq y$, if one can obtain y from x by taking some packaging inside x and replacing this by other protosets.

Partial set theory has a first order language L with three relation symbols, $\in$ (for actual membership), $\in_\diamond$ (for possible membership), and *set* (for set existence). The theory consists of two axioms and ZFA^{set}, the relativization of all axioms of ZFA (ZF + Aczel's AFA) to the relation *set*. The two axioms are (i) Pict, which states that every partial set has a *picture*, a set G which is a partial set graph (corresponding to the accessible pointed graph of Aczel) and such that there is a decoration d of G with the root decorated as x, and (ii) PSA, which states that every such G has a unique decoration. Partial set theory ZFAP is the set of all these axioms. ZFAP is a conservative extension of ZFA.

5.4. Dionne et al.'s work

Dionne, Mays, and Oles (1992) propose a new approach to intensional semantics of term subsumption languages. Their work is inspired by the research of Woods (1991) who suggested that a more intensional view of concept descriptions should be taken. In general, most of the work in semantics of term subsumption languages adopts an extensional view. Thus, concepts are interpreted as sets of objects from some universe. Roles of concepts are interpreted as binary relations over the universe. Concept descriptions are complex predicates and one inquires whether a given instance satisfies a complex predicate.

Dionne et al. consider a small subset of K-REP, a KL-ONE style of language (Brachman & Schmolze, 1985). They offer a general discussion of cycles, viz. how they arise, and how they are handled in K-REP. The restricted language they use allows simple concepts, concepts formed by conjunctions, and roles of concepts whose value restrictions are other concepts. In their vision, a knowledge base is a set of possibly mutually recursive equations, involving terms of this concept language. Essential use is made of Aczel's theory of nonwellfounded sets, especially the bisimulation relation. A so-called concept algebra is developed as a new approach to semantics of term subsumption languages. It is shown that for the above subset of K-REP, these algebras accurately model the process of subsumption testing—even in the presence of cycles. Concept algebras also allow for multiple definitions, i.e., concepts with different names that semantically stand for the same concept.

5.5. Barwise's work

Barwise (1989b) attempted to propose a set theory, *Situated Set Theory*, not just for use in AI, but for general use. He mentioned the problems caused by the common view of set theory with a universal set V, but at the same time trying to treat this universe as an extensional whole, looking from outside (which he names "unsituated set theory"). His proposal is a hierarchy of universes $\mathrm{V}_0 \subset \mathrm{V}_1 \subset \mathrm{V}_2 \subset \ldots$ which allows for a universe of a lower level to be considered as an object of a universe of a higher level. He leaves the axioms which these universes have to satisfy to one's conception of set, be it cumulative or circular. There are no paradoxes in this view since there is always a larger universe one can step back and work in. Therefore, the notions of "set," "proper class," and the set-theoretic notions "ordinal," "cardinal" are all context sensitive, depending on the universe one is currently working in. This proposal supports the *Reflection Principle* which states that for any given description of the sets of all sets V, there will always be a partial universe satisfying that description.

Barwise (1989c) also studied the modeling of partial information and again exploited Hyperset Theory for this purpose. For this purpose, he used the objects of the universe $\mathcal{V}_A$ of hypersets over a set A of atoms to model nonparametric objects, i.e., objects with *complete* information and the set X of indeterminates to represent parametric objects, i.e., objects with *partial* information. (The universe of hypersets on $A \cup X$ is denoted as $\mathcal{V}_A[X]$, analogous to the adjunction of indeterminates in algebra.)

```
object[ref='Akman et al. (1993)']/
      [
      date=1993/4/24,
      kind=paper,
      authors={'V. Akman', 'M. Pakkan', 'M. Surav'},
      title='Lectures in Commonsense Set Theory',
      journal='Proc. Natl. Acad. Sci. Utopia',
      volume=78,
      pages=405-409,
      year=1993
      ].
```

Figure 16: An object in $\mathcal{QUIXOTE}$

For any object $a \in \mathcal{V}_A[X]$, Barwise calls the set

$$par(a) = \{x \in X \; : \; x \in TC(a)\},$$

where $TC(a)$ denotes the *transitive closure* of a, the *set of parameters* of a. If $a \in \mathcal{V}_A$, then $par(a) = \emptyset$ since a does not have any parameters. Barwise then defines an *anchor* as a function f with $domain(f) \subseteq X$ and $range(f) \subseteq \mathcal{V}_A - A$ which assigns sets to indeterminates. For any $a \in \mathcal{V}_A[X]$ and anchor f, $a(f)$ is the object obtained by replacing each indeterminate $x \in par(a) \cap domain(f)$ by the set $f(x)$ in a. This is accomplished by solving the resulting equations by the Solution Lemma.

Parametric anchors can also be defined as functions from a subset of X into $\mathcal{V}_A[X]$ to assign parametric objects, not just sets, to indeterminates. For example, if $a(x)$ is a parametric object representing partial information about some nonparametric object $a \in \mathcal{V}_A$ and if one does not know the value to which x is to be anchored, but knows that it is of the form $b(y)$ (another parametric object), then anchoring x to $b(y)$ results in the object $a(b(y))$ which does not give the ultimate object perhaps, but is at least more informative about its structure.

5.6. Yasukawa and Yokota's work

Yokota and Yasukawa (1992) proposed a knowledge representation language, called $\mathcal{QUIXOTE}$ for deductive and object-oriented databases. In $\mathcal{QUIXOTE}$, an object consists of an identifier (oid) and properties, each attribute of which is a triple $\langle label, operator, value \rangle$.

In (Yasukawa & Yokota, 1991), a partial semantics for the semantics of objects in $\mathcal{QUIXOTE}$ is given. This work is along the lines of (Dionne et al., 1992) and will now be summarized. Consider Figure 16 where the left hand side of "/" is an oid and the right hand side is the related properties. An object consists of an oid and its properties, and can be written as a set of attribute terms with the same oid as follows:

$$o/[l_1 = a, l_2 = b] \Longleftrightarrow o/[l_1 = a], o/[l_2 = b]$$

An oid can be defined intensionally by a set of rules as follows (note the Prolog-like style):

$$path[from = X, to = Y] \Leftarrow arc[from = X, to = Y]$$

$$path[from = X, to = Y] \Leftarrow arc[from = X, to = Z], path[from = Z, to = Y]$$

Here, $path[from = X, to = Y]$ is transitively defined from a set of facts such as $arc[from = a, to = b]$, etc., and the oid is generated by instantiating X and Y as a result of the execution of the program. This guarantees that an object has a unique oid even if it is generated in different environments. Furthermore, circular paths can be defined: $X@o[l = X] \Longleftrightarrow X|\{X = o[l = X]\}$. This denotes that a variable X is an oid with the constraint $X = o[[l = X]$. The semantics of oids is defined on a set of labeled graphs as a subclass of hypersets. Thus, an oid is mapped to a labeled graph and an attribute is related to a function on a set of labeled graphs. Since the metatheory is ZFC$^-$/AFA, all of the familiar set-theoretic techniques to deal with circular phenomena are brought to bear.

6. Conclusion

Set theory can be useful in intelligent information management. The methodology may change, of course. A universal set theory, answering many technical questions in information management, can be developed by means of proposing new axioms or modifying existing ones. Alternatively and more conservatively, different set-theoretic concepts may be examined and modified based on existing set theories. No matter what proposal is followed, we believe that further research in this field should be to the advantage of the IIS community.

Acknowledgments

We would like to thank an anonymous referee of the *Journal of Intelligent Information Systems* for critical comments which were crucial in revising the original manuscript. Larry Kerschberg and Maria Zemankova, two of the Editors-in-Chief of *JIIS*, were extremely kind and accommodating when we first submitted the paper. We are also indebted to Wlodek Zadrozny (IBM T. J. Watson Research Center) for moral support and technical advice. As usual, we are solely responsible for this final version.

The first author's research is supported in part by the Scientific and Technical Research Council of Turkey under grant no. TBAG–992.

Notes

1. On the other hand, another great mathematician of this century, R. Thom, said in 1971: "The old hope of Bourbaki, to see mathematical structures arise naturally from a hierarchy of sets, from their subsets, and from their combination, is doubtless, only an illusion" (Goldblatt, 1984).

2. A wff w is said to be *stratified* if integers are assigned to the variables of w such that all occurrences of the same free variable are assigned the same integer, all bound occurrences of a variable that are bound by the same quantifier are assigned the same integer, and for every subformula $x \in y$, the integer assigned to y is equal to the integer assigned to $x + 1$. For example, $(x \in y) \,\&\, (z \in x)$ is stratified as $(x^1 \in y^2) \,\&\, (z^0 \in x^1)$.

3. Let us recall Russell's Paradox. We let r be the set whose members are all sets x such that x is not a member of x. Then for every set x, $x \in r$ iff $x \notin x$. Substituting r for x, we obtained the contradiction.

   The explanation is not difficult. When we are forming a set z by choosing its members, we do not yet have the object z, and hence cannot use it as a member of z. The same reasoning shows that certain other sets cannot be members of z. For example, suppose that $z \in y$. Then we cannot form y until we have formed z. Hence y is not available and therefore cannot be a member of z. Carrying this analysis a bit further, we arrive at the following. Sets are formed in "stages." For each stage S, there are certain stages which are before S.

   Stages are important because they enable us to form sets. Suppose that x is a collection of sets and Σ is a collection of stages such that each member of x is formed at a stage which is a member of Σ. If there is a stage after all of the members of Σ, then we can form x at this stage. Now the question becomes: Given a collection Σ of stages, is there a stage after all of the members of Σ? We would like to have an affirmative answer to this question. Still, the answer cannot always be "yes"; if Σ is the collection of all stages, then there is no stage after every stage in Σ.

4. A theory T (any set of formulas closed under implication, i.e., for any φ, if $T \models \varphi$, then $\varphi \in T$) is *finitely axiomatizable* iff there is a finite $T' \subseteq T$ such that for every ψ in T, $T' \models \psi$.

5. A linear ordering $<$ of a set a is a *well-ordering* if every nonempty subset of a has a least element. Informally, an ordered set is said to be *well-ordered* if the set itself, and all its nonempty subsets have a first element under the order prescribed for its elements by that set. An ordinal number stands for an order type which is represented by well-ordered sets. First, call a set a *transitive* if $\forall x(x \in a \rightarrow x \subseteq a)$. Then, a set is an ordinal number (or ordinal) if it is transitive and well-ordered by $\in$. It should be noted that in case of finite sets, the notions of cardinal number and ordinal number are the same. The class of all ordinal numbers is denoted by Ord. The relationship $<$ between two ordinals α and β, $\alpha < \beta$, is defined iff $\alpha \in \beta$. If $\alpha = \beta + 1$, then α is called a *successor ordinal*; else it is a *limit ordinal*. The Axiom of Infinity guarantees the existence of limit ordinals other than 0. In fact, ω, the set of natural numbers, is the next limit ordinal (Figure 2).

6. We have to prove the theorem

   $$\exists c \forall x[x \in c \leftrightarrow \exists y \exists z(y \in a \,\&\, z \in b \,\&\, x = \langle y, z \rangle)]$$

   to show that the Cartesian product set exists. The main point of the proof is that if $x = \langle y, z \rangle$, and if $y \in a$ and $z \in b$, then $x \in P(P(a \cup b))$. Then, by the Axiom of Separation,

   $$\exists c \forall x[x \in c \leftrightarrow x \in P(P(a \cup b)) \,\&\, \exists y \exists z(y \in a \,\&\, z \in b \,\&\, x = \langle y, z \rangle)]. \tag{1}$$

   The theorem to be proved is equivalent to (1) without the statement $x \in P(P(a \cup b))$. Then we must show that the equivalence in (1) still holds when that statement is eliminated. Given (1), it follows that

   $$x \in c \tag{2}$$

implies

$$\exists y \exists z [y \in a \;\&\; z \in b \;\&\; x = \langle y, z \rangle]. \tag{3}$$

To prove the converse implication, we must show that (3) implies $x \in P(P(a \cup b))$, since it is obvious by (1) that (3) implies (2). By (3) and the definition of ordered pairs $x = \{\{y\}, \{y, z\}\}$, and by the hypotheses $y \in a$ and $z \in b$, we have:

$$\{y\} \subseteq \{a \cup b\} \text{ and } \{y, z\} \subseteq \{a \cup b\}.$$

Then by the following theorem (which can be proved by the Power Axiom):

$$b \in P(a) \leftrightarrow b \subseteq a, \tag{4}$$

we conclude:

$$\{y\} \in P(a \cup b) \text{ and } \{y, z\} \in P(a \cup b).$$

Thus, $\{\{y\}, \{y, z\}\} \subseteq P(a \cup b)$, i.e., $x \subseteq P(a \cup b)$, and again by (4), we have $x \in P(P(a \cup b))$.

7. The collection of Δ_0 formulas of a language L is the smallest collection Γ containing the atomic formulas of L and inductively defined as:

   (a) If φ in in Γ, then $\neg\varphi$ is also in Γ.

   (b) If φ, ψ are in Γ, then $(\varphi \wedge \psi)$ and $(\varphi \vee \psi)$ are also in Γ.

   (c) If φ is in Γ, then $\forall u \in v\; \varphi$ and $\exists u \in v\; \varphi$ are also in Γ for all variables u and v.

8. A *structure* for a first order language L is a pair $\langle M, I \rangle$, where M is a nonempty set called the *domain* of the structure and I is an *interpretation function* assigning functions and predicates over M to the symbols in L.

9. Aczel uses *tagged graphs* to represent sets, i.e., each childless node in the graph is *tagged* by an atom or $\emptyset$. A *pointed graph* is a tagged graph with a specific node n_0 called its *point*. A pointed graph is *accessible* (denoted as apg) if for every node n there is a path $n_0 \longrightarrow n_1 \longrightarrow \cdots \longrightarrow n$. A *decoration* of a graph is a function $D(n)$ for each node n, defined as:

$$D(n) = \begin{cases} tag(n), & \text{if } n \text{ has no children,} \\ \{D(m) : m \text{ is a child of } n\}, & \text{otherwise.} \end{cases}$$

   A *picture* of a set is an apg which has a decoration in which the set is assigned to the point.

10. The uniqueness property of AFA leads to an intriguing concept of extensionality for hypersets. The classical extensionality paradigm, that sets are equal iff they have the same members, works fine with wellfounded sets. However, this is not of use in deciding the equality of say, $a = \{1, a\}$ and $b = \{1, b\}$ because it just asserts that $a = b$ iff $a = b$, a triviality (Barwise & Etchemendy, 1987). However, in the universe of hypersets, a is indeed equal to b since they are depicted by the same graph. To see this, consider a graph G and a decoration D assigning a to a node x of G, i.e., $D(x) = a$. Now consider the decoration D' exactly the same as D except that $D'(x) = b$. D' must also be a decoration for G. But by the uniqueness property of AFA, $D = D'$, so $D(x) = D'(x)$, and therefore $a = b$.

11. The Solution Lemma is an elegant result, but not every system of equations has a solution. First of all, the equations have to be in the form suitable for the Solution Lemma. For example, the pair

$$\begin{aligned} x &= \{y, z\}, \\ y &= \{1, x\}, \end{aligned}$$

cannot be solved since it requires the solution to be stated in terms of the indeterminate z. As another example, the equation $x = P(x)$ cannot be solved because Cantor has proved (in ZFC$^-$) that there is no set which contains its own power set—no matter what axioms are added to ZFC$^-$.

12. A *state of affair* (a.k.a. *infon*) $\langle R, a, i\rangle$ is a triple where R is an n-ary relation, a is an appropriate assignment of objects, and i is the polarity, 1 if there is at least one instance of R holding of a, and 0 otherwise. By a state of affair, a state that affairs may or may not be in is meant. When $i = 1$, that state of affair is called a *fact* and the polarity is usually omitted. For example, the state of affair $\langle$sleeping, Tom, garden$\rangle$ is a fact if Tom is sleeping in the garden.

13. According to the *Liar Paradox* (also known as the *Epimenides Paradox*), Epimenides, the Cretan, said "All Cretans are liars." Now this statement cannot be true since this would make Epimenides a liar, leading to the falsity of his statement. The statement cannot be false either, since this would imply that Cretans are not liars, hence what Epimenides says should be true, leading to a contradiction.

14. The predicate of a, b, u, which is defined as "u is an ordered pair $\langle y, z\rangle$ with $y \in a$ and $z \in b$" in KPU, is Δ_0. Hence, Δ_0-Separation can be used once it is known that there exists a set c with $\langle y, z\rangle \in c$ for all $y \in a$ and $z \in b$. This follows from Δ_0-Collection as follows. Given $y \in b$, there exists a set $d = \langle y, z\rangle$. So, by Δ_0-Collection, there exists a set w_y such that $\langle y, z\rangle \in w_y$ for all $z \in b$. Applying Δ_0-Collection again, we have:

$$\forall y \in a\, \exists w\, \forall z \in b\, \exists d \in w\, (d = \langle y, z\rangle)$$

so there is a c_1 such that for all $y \in a$, $z \in b$, $\langle y, z\rangle \in w$ for some $w \in c_1$. Thus, if $c = \bigcup c_1$, then $\langle y, z\rangle \in c$ for all $y \in a$ and $z \in b$.

15. The $*$ is the notation for first writing all $\rightarrow$ symbols in terms of & and $\neg$, then passing negations in α through to predicate letters, and finally replacing each occurrence of a subformula $\neg x \in [y]$ in the result by $x \in [\neg y]$.

16. The *transitive closure* of a set, denoted by $TC(a)$, is defined by recursion as follows:

    - $\bigcup^0 a \;=\; a$,
    - $\bigcup^{n+1} a \;=\; \bigcup(\bigcup^n a)$,
    - $TC(a) \;=\; \bigcup\{\bigcup^n a \;:\; n = 1, 2, \ldots\}$.

    Hence, $TC(a) = a \cup \bigcup a \cup \bigcup^2 a \ldots$ For any infinite cardinal m, the set $H(m)$ is defined as $H(m) = \{x : |TC(x)| < m\}$. The elements of $H(m)$ are said to be *hereditarily* of cardinality less than m. $\{H(n) : n = 1, 2, \ldots\}$ is the set of *hereditarily finite* sets. Hence, every element of a hereditary set is a hereditary set.

References

Aczel, P. (1988). *Nonwellfounded Sets.* Number 14 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.

Akman, V., and Pakkan, M. (1993). HYPERSOLVER: A Graphical Tool for Commonsense Set Theory. In Gün, L., Önvural, R., and Gelenbe, E. (eds.) *Proceedings of Eight International Symposium on Computer and Information Sciences*, Istanbul, Turkey.

Barwise, J. (1975). *Admissible Sets and Structures.* Springer-Verlag, Berlin.

Barwise, J. (1977). An Introduction to First Order Logic. In Barwise, J. (ed.) *Handbook of Mathematical Logic*, 5–46, North-Holland, Amsterdam.

Barwise, J. (1985). Model-Theoretic Logics: Background and Aims. In Barwise, J., and Feferman, S. (eds.) *Model-Theoretic Logics*, 3–23, Springer-Verlag, New York.

Barwise, J. (1989a). Situations, Sets, and the Axiom of Foundation. In Barwise, J. *The Situation in Logic*, 177–200, Number 17 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.

Barwise, J. (1989b). Situated Set Theory. In Barwise, J. *The Situation in Logic*, 289–292, Number 17 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.

Barwise, J. (1989c). AFA and the Unification of Information. In Barwise, J. *The Situation in Logic*, 277–283, Number 17 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.

Barwise, J. (1989d). On the Model Theory of Common Knowledge. In Barwise, J. *The Situation in Logic*, 201–220, Number 17 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.

Barwise, J. (1990). Consistency and Logical Consequence. In Dunn, J. M., and Gupta, A. (eds.) *Truth or Consequences*, 111–122, Kluwer, Dordrecht, Holland.

Barwise, J. (1992). *Applying AFA: Notes for a Tutorial on Nonwellfounded Sets*, Manuscript, Department of Mathematics, Indiana University, Bloomington, IN.

Barwise, J. and Etchemendy, J. (1987). *The Liar: An Essay on Truth and Circularity*, Oxford University Press, New York.

Barwise, J. and Moss, L. S. (1991). Hypersets. *Mathematical Intelligencer*, *13*, 31–41.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*, MIT Press, Cambridge, MA.

Boyer, R., Lusk, E., McCune, W., Overbeek, R., Stickel, M. and Wos, L. (1986). Set Theory in First Order Logic. *Journal of Automated Reasoning*, *2*, 287–327.

Brachman, R. J., and Schmolze, J. G. (1985). An Overview of the KL-ONE Knowledge Representation System. *Cognitive Science*, *9*, 171–216.

Brown, F. M. (1978). Towards the Automation of Set Theory and its Logic. *Artificial Intelligence*, *10*, 281–316.

Cowen, R. (1993). Some Connections Between Set Theory and Computer Science. In *Proceedings of Third Kurt Gödel Colloquium*, Number 713 in Lecture Notes in Computer Science, Springer-Verlag, Berlin.

Dionne, R., Mays, E., and Oles, F. J. (1992). A Nonwellfounded Approach to Terminological Cycles. In *Proceedings of AAAI-92*, San Jose, CA.

Feferman, S. (1984). Toward Useful Type-Free Theories, I. *Journal of Symbolic Logic*, *49*, 75–111.

Fraenkel, A. A., Bar-Hillel, Y., and Levy, A. (1973). *Foundations of Set Theory*, North-Holland, Amsterdam.

Gilmore, P. (1974). The Consistency of Partial Set Theory without Extensionality. In Jech, T. J. (ed.) *Axiomatic Set Theory*, 147–153, American Mathematical Society, Providence, RI.

Gilmore, P. (1993). Logic, Sets, and Mathematics. *The Mathematical Intelligencer*, *15*, 10–19.

Goldblatt, R. (1984). *Topoi: The Categorial Analysis of Logic*, North-Holland, Amsterdam.

Halmos, P. R. (1974). *Naive Set Theory*, Springer-Verlag, New York.

van Heijenhoort, J. (1967). *From Frege to Gödel*, Harvard University Press, Cambridge, MA.

McCarthy, J. (1959). Programs with Common Sense. In Lifschitz, V. (ed.) *Formalizing Common Sense: Papers by John McCarthy*, 9–16, Ablex, Norwood, NJ.

McCarthy, J. (1983). Artificial Intelligence Needs More Emphasis on Basic Research: President's Quarterly Message. *AI Magazine*, *4*, 5.
McCarthy, J. (1984). We Need Better Standards for Artificial Intelligence Research: President's Message. *AI Magazine*, *5*, 7–8.
McCarthy, J. (1985). Acceptance Address, *International Joint Conference on Artificial Intelligence (IJCAI)* Award for Research Excellence, Los Angeles, CA.
Mislove, M. W., Moss, L. S., and Oles, F. J. (1990). Partial Sets. In Cooper, R., Mukai, K., and Perry, J. (eds.) *Situation Theory and Its Applications I*, 117–131, Number 22 in CSLI Lecture Notes, Center for the Study of Language and Information, Stanford, CA.
Mislove, M. W., Moss, L. S., and Oles, F. J. (1991). Nonwellfounded Sets Modeled as Ideal Fixed Points. *Information and Computation*, *93*, 16–54.
Pakkan, M. (1993). *Solving Equations in the Universe of Hypersets*, M.S. Thesis, Department of Computer Engineering and Information Science, Bilkent University, Ankara, Turkey.
Parsons, C. (1977). What is the Iterative Conception of Set? In Butts, R. E., and Hintikka, J. (eds.) *Logic, Foundations of Mathematics, and Computability Theory*, 339–345, Kluwer, Dordrecht, Holland.
Perlis, D. (1985). Languages with Self-Reference I: Foundations. *Artificial Intelligence*, *25*, 301–322.
Perlis, D. (1988). Commonsense Set Theory. In Maes, P., and Nardi, D. (eds.) *Meta-Level Architectures and Reflection*, 87–98, Elsevier, Amsterdam.
Quaife, A. (1992). Automated Deduction in von Neumann-Bernays-Gödel Set Theory. *Journal of Automated Reasoning*, *8*, 91–147.
Quine, W. V. (1937). New Foundations for Mathematical Logic. *American Mathematical Monthly*, *44*, 70–80.
Shoenfield, J. R. (1977). Axioms of Set Theory. In Barwise, J. (ed.) *Handbook of Mathematical Logic*, 321–344, North-Holland, Amsterdam.
Suppes, P. (1972). *Axiomatic Set Theory*, Dover, New York.
Tiles, M. (1989). *The Philosophy of Set Theory*, Basil Blackwell, Oxford, UK.
Whitehead, A. N., and Russell, B. (1910). *Principia Mathematica*, Three Volumes, Cambridge University Press, Cambridge, UK.
Woods, W. A. (1991). Understanding Subsumption and Taxonomy: A Framework for Progress. In Sowa, J. F. (ed.) *Principles of Semantic Networks*, Morgan Kaufmann, San Mateo, CA.
Yasukawa, H., and Yokota, K. (1991). *Labeled Graphs as Semantics of Objects*, Manuscript, $\mathcal{QUIXOTE}$ Project, Institute for New Generation Computer Technology (ICOT), Tokyo.
Yokota, K., and Yasukawa, H. (1992). Towards an Integrated Knowledge-Base Management System: Overview of R&D on Databases and Knowledge-Bases in the FGCS Project. In *Proceedings of Fifth Generation Computer Systems*, Tokyo.
Zadrozny, W. (1989). Cardinalities and Well-Orderings in a Common-Sense Set Theory. In Brachman R. J., Levesque H. J., and Reiter, R. (eds.) *Proceedings of First International Conference on Principles of Knowledge Representation and Reasoning*, 486–497, Morgan Kaufmann, San Mateo, CA.
Zadrozny, W., and Kim, M. (1993). Computational Mereology: A Prolegomenon Illuminated by a Study of PartOf Relations for Multimedia Indexing. In *Proceedings of Bar-Ilan Symposium on the Foundations of AI (BISFAI–93)*, Israel.

Authors' biographies

Varol Akman is associate professor of Computer Engineering and Information Science at Bilkent University, Ankara, Turkey. His research is concentrated in mathematical logic and commonsense reasoning, and additionally, philosophy and natural language semantics. Prior to Bilkent, he held a senior researcher position with the Centrum voor Wiskunde en Informatica, Amsterdam (1986–1988) and a post-doctoral position with the Vakgroep Informatica, Rijksuniversiteit Utrecht, the Netherlands (1985–1986). Akman received a B.S. degree in Electrical Engineering (1979, high honors) and an M.S. degree in Computer Engineering (1980) from Middle East Technical University, Ankara. In 1980–1985, he was a Fulbright scholar and a research assistant at Rensselaer Polytechnic Institute, New York, and received a Ph.D. degree in Computer and Systems Engineering. Akman was conferred the 1989 research prize of the Scientific and Technical Research Council of Turkey. He also received the Prof. Mustafa N. Parlar Foundation's scientific research award in 1990. Akman's doctoral dissertation was published in 1987 by Springer-Verlag. He is also the senior editor of a 1989 Springer-Verlag volume on design. He is an editorial board member of the journal *Computers & Graphics* and the *Bilkent University Lecture Series in Engineering* (research monographs marketed by Springer-Verlag). He is in the program committees of various conferences and is a member of American Association for Artificial Intelligence, and European Foundation for Logic, Language, and Information.

Müjdat Pakkan is a doctoral student in the Department of Computer Engineering, Bosphorus University, Istanbul, Turkey. His areas of interest include knowledge representation and natural language semantics. Pakkan received a B.S. degree in Computer Engineering (1990, honors) from Aegean University, Izmir, Turkey. He then received an M.S. degree in Computer Engineering and Information Science (1993) from Bilkent University, Ankara. Pakkan's master's thesis studies nonwellfounded sets from a computational perspective.