ISSUES IN COMMONSENSE SET THEORY

Müjdat Pakkan and Varol Akman

BILKENT UNIVERSITY

Department of Computer Engineering
and
Information Science

Technical Report BU-CEIS-94-06

Issues in commonsense set theory

MÜJDAT PAKKAN and VAROL AKMAN

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

Abstract: The success of set theory as a foundation for mathematics inspires its use in artificial intelligence, particularly in commonsense reasoning. In this survey, we briefly review classical set theory from an AI perspective, and then consider alternative set theories. Desirable properties of a possible commonsense set theory are investigated, treating different aspects like cumulative hierarchy, self-reference, cardinality, etc. Assorted examples from the ground-breaking research on the subject are also given.

Key Words: set theory, commonsense reasoning, knowledge representation, cumulative hierarchy, self-reference, hypersets

"... among all the mathematical theories,
it is just the theory of sets that requires
clarification more than any other."
(Mostowski 1979)

1 INTRODUCTION

Set theory is a branch of modern mathematics with a unique place because various other branches can be formally defined within it (Suppes 1972). For example, Book 1 of the influential works of N. Bourbaki is devoted to the theory of sets, which provides the framework for the whole enterprise. Bourbaki has said in 1949: "... all mathematical theories may be regarded as extensions of the general theory of sets ... on these foundations I can state that I can build up the whole of the mathematics of the present day" (Goldblatt 1984). Indeed, one can represent a natural number as a set, a rational number as a pair of natural numbers, a real number as a set of rationals, and so on (Mac Lane 1986). Hence, most of the mathematical entities may be regarded as sets and set theory can be considered as the fundamental theory underlying mathematics [1]. This brings up the possibility of using set theory in foundational studies in artificial intelligence (AI), particularly in commonsense reasoning. McCarthy (1983) has emphasized the need for foundational research in AI and claimed that AI needs mathematical and logical theory involving conceptual innovations. He stated that one of the key problems is the formalization of commonsense knowledge and reasoning. In his opening address in *IJCAI–85* (McCarthy 1985), he stressed the feasibility of using set theory in AI and invited researchers to concentrate on the subject.

There is, in some sense, great beauty, economy, and naturalness in using sets for modeling and knowledge representation in AI (Akman 1992). This is owing to the fact that sets agree very well with our intuitions (Parsons 1990). Gödel (1947) eloquently states this in the following excerpt:

> "But despite the remoteness from sense-experience, we do have something like a perception of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense-perception."

In this survey, we first give a brief review of classical set theories, trying to avoid the technical details—which the reader can find in classical texts like (Halmos 1974) or (Fraenkel et al. 1973)—and instead focusing on the underlying concepts. We then consider the alternative set theories which have been proposed throughout the century to overcome the limitations of classical theories. Later, we investigate the properties of a possible commonsense set theory, treating different aspects such as urelements, cumulative hierarchy, self-reference, cardinality, well-orderings, and so on. We finally summarize the noteworthy research on the subject and offer our concluding remarks.

2 CLASSICAL SET THEORY

2.1 *Earliest Developments*

G. Cantor's work on the theory of infinite series and related topics should be considered as the foundation of the research in set theory. In Cantor's conception, a *set*, or *aggregate*, is a collection into a whole of definite, distinct objects of our perception or our thought, called the *elements* of the set (Cantor 1883). 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 earlier stages of his research, Cantor did not work from axioms (Suppes 1972). 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 non-empty 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, Russell found a contradiction in Frege's foundational system (Frege 1893) which was developed on Cantor's naive set conception (van Heijenhoort 1967). Frege's reaction to this can be found in the appendix to the second volume of his famous *Grundgesetze der Arithmetik*: "Hardly anything more unfortunate can befall a scientific writer than to have one of the foundations of his edifice shaken after the work is finished. This was the position I was placed in by a letter of Mr. Bertrand Russell." This contradiction could be derived from the Axiom of Abstraction (which was named 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 $\neg(x \in x)$ in the language of first-order logic. ($\neg(x \in y)$ will be denoted as $x \notin y$ from now on.)

The Axiom of Abstraction 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 $y \in y \leftrightarrow y \notin y$. The problematic thing here is the set x with the property $x \notin x$.

Another antinomy occurred with the conception of the "set of all sets," $\mathrm{V} = \{x : x = x\}$. The well-known Cantor's Theorem states that the power set (set of all subsets) of V has a greater cardinality than V itself. This is obviously paradoxical since V by definition is the most inclusive set. This is the so-called Cantor's Paradox (Cantor 1932) and led to discussions on the sizes 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, von Neumann would clarify this problem of size by stating that (Goldblatt 1984) "Some predicates have extensions that are too large to be successfully encompassed as a whole and treated as a mathematical object."

Such paradoxes shook the theory to its foundations and were instrumental in new axiomatizations of the set theory or 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. Axiomatization of a theory is important since it provides a concise formulation of the principles of the theory and allows fundamental notions like completeness and consistency to be discussed in a precise way; these would be formulated in an imprecise manner (e.g., in natural language) otherwise.

2.2 *Alternate Approaches and Axiomatizations*

The new axiomatizations took a common step for overcoming the deficiencies of the naive approach by introducing *classes*, a membership-eligible entity corresponding a given condition. NBG, which was proposed by von Neumann (1925) and later revised and simplified by 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 $\mathrm{Y} = \{x : x \notin x\}$ 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.

Strengthening NBG by replacing the axioms of class existence with an axiom scheme, a new theory called *Morse-Kelley* (MK) is obtained (Morse 1965). MK is suitable if one is not interested in the subtleties of set theory. But its strength risks its consistency (Mendelson 1987).

Ackermann (1956) also proposed an axiomatization again employing classes, but in which the central objects are sets. The main point of this axiomatization is that its axioms retain only the weakest consequences of the limitation of size constraint, i.e., a member of a set and a subclass of a set are sets.

Other approaches against the deficiencies of the naive approach alternatively played with its language and are generally called *type-theoretical* approaches. 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 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 $x \in x$ is not stratified.

2.3 *ZF Set Theory*

Zermelo-Fraenkel (ZF) is the earliest axiomatic system in set theory. The first axiomatization was by Zermelo (1908). Fraenkel (1922) observed a weakness of Zermelo's system and proposed a way to overcome it. His proposal was reformulated by Skolem (1922) by introducing a new axiom. This axiomatization 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 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 the empty set and then construct further sets in a stepwise manner by various defined operators. Hence there are no individual objects (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., the empty set $\emptyset$. 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 $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots\}$. 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 [3] 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

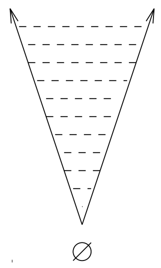

Figure 1: ZF universe extending in a cumulative hierarchy

problematic set $x = \{x\}$ cannot be shown to exist. (This will be demonstrated shortly.) 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 by his Axiom of Abstraction. It only allows 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*. They become axioms when one substitutes a specific description or relational expression in the language of ZF instead of the variable expression $\varphi(x)$. Therefore, we say that 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 led by showing that for any set, the cardinality of its power set must be greater than its cardinality [5].

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 *well-founded* 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 [6]:

- $R(0) = \emptyset$,
- $R(\alpha + 1) = P(R(\alpha))$,
- $R(\alpha) = \bigcup_{\beta < \alpha} R(\beta)$ when α is a limit ordinal,
- $\mathrm{WF} = \bigcup\{R(\alpha) : \alpha \in Ord\}$.

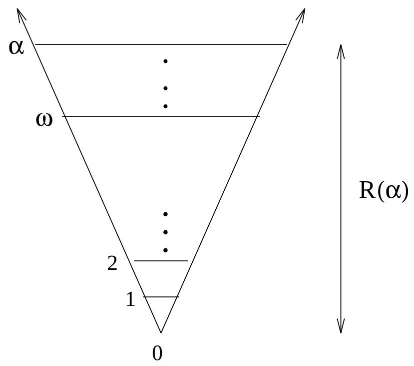

Figure 2: WF universe defined recursively in terms of ordinals

This universe of WF 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 (Kunen 1980).

Let us now 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, $x \in r$ if and only if $x \notin x$. Substituting r for x, we obtained the contradiction.

With the preceding discussion of WF 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.

It can be said that 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 $\models \varphi$ (Mendelson 1987). 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, e.g., research on independency and consistency, is being carried out in ZF. Nevertheless, ZF has its own drawbacks (Barwise 1975). First of all, it is too weak to decide some questions like the *Continuum Hypothesis* (Gödel 1947). Another critical point is that 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 difference disappears in ZF. Similarly, for an arbitrary Abelian group $\langle G, + \rangle$, all of the following subgroups of G are considered as equivalent in ZF (Barwise 1975), while the definitions are increasing in logical complexity [7]:

- $pG = \{px : x \in G\}$ = the left coset of G,
- $T = \{x : nx = 0$ for some integer $n > 0\}$ = the torsion subgroup of G,
- $\bigcup\{H : H$ is a divisible subgroup of $G\}$ = the divisible part of G.

A desirable property, the *Principle of Parsimony*, which states that simple facts should have simple proofs, is quite often violated in ZF (Barwise 1975). For example, the 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 [8].

It can also be claimed mathematical practice suffers from the fact that all the mathematical objects are represented as sets in ZF. For example, while one can construct in ZF something isomorphic to the real line, the practicing mathematician is not very interested in this. Representing reals as sets could be considered important from a theoretical viewpoint, but we should hardly ever worry about the fact that $\sqrt{2}$ can be determined by the infinite sequence $\langle 1, 2\rangle, \langle 1.4, 1.5\rangle, \langle 1.41, 1.42\rangle, \ldots$

3 ALTERNATIVE SET THEORIES

3.1 *Admissible Set Theory*

Admissible sets are formalized in a first order set theory called *Kripke-Platek* (KP) (Kripke 1964). Barwise weakened KP to a new theory KPU by readmitting the *urelements* (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 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 with the membership symbol added. 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* [9] 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,
- $\mathrm{V}_M = \bigcup_\alpha V_M(\alpha)$.

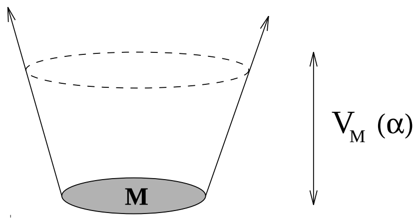

Figure 3: The universe of admissible sets (adapted from (Barwise 1975))

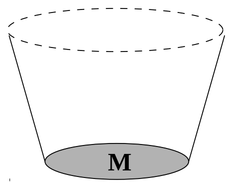

Figure 4: A typical admissible set (adapted from (Barwise 1975))

where P is the power operation, and α and λ are ordinals. This universe can be depicted as in Figure 3. 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 (Barwise 1977).

If M is a structure [10] 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 non-urelements and $\in$ is defined in $M \times A$. Such a typical admissible set over M can be depicted as in Figure 4. A *pure admissible set* is an admissible set with no urelements, i.e., it is a model of KP. Such a set can be depicted as in Figure 5.

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.

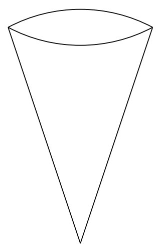

Figure 5: A pure admissible set (adapted from (Barwise 1975))

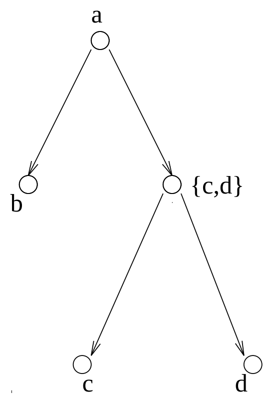

Figure 6: Representation of the set $a = \{b, \{c, d\}\}$ in Aczel's conception

3.2 *Hyperset Theory*

It was Mirimanoff (1917) who first stated the fundamental difference between *well-founded* and *non-well-founded* sets. He called sets with no infinite descending membership sequence well-founded and others non-well-founded. Non-well-founded sets have been extensively studied through decades, but did not show up in notable applications until Aczel. This is probably due to the fact that the classical well-founded universe was a rather satisfying domain for the practicing mathematician ("the mathematician in the street" (Barwise 1985)). Aczel's work on non-well-founded sets evolved from his interest in modeling concurrent processes. He adopted the graph representation for sets to use in his theory. A set like $a = \{b, \{c, d\}\}$ can be unambiguously depicted as in Figure 6 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 well-founded if it has no infinite paths or cycles, and non-well-founded otherwise. Aczel's *Anti-Foundation Axiom*, AFA, states that every graph, well-founded or not, pictures a unique set. Removing the Axiom of Foundation (FA) from the ZFC and adding the AFA results in the *Hyperset Theory*. (ZFC without 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. For example, the graph in Figure 7 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 [11] 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 7 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 the three graphs in Figure 8 depict the unique non-well-founded set $\Omega = \{\Omega\}$.)

Aczel develops his own extensionality concept by introducing the notion of bisimulation. A *bisimulation* between two apg's, G_1 with point p_1 and G_2 with point p_2, is a relation

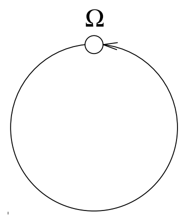

Figure 7: The picture of the circular set $\Omega = \{\Omega\}$

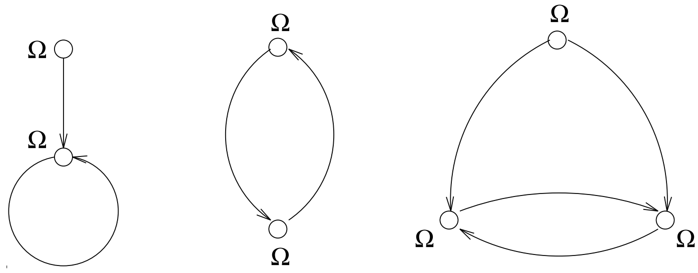

Figure 8: Other graphs depicting Ω

$R \subseteq G_1 \times G_2$ 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.

The uniqueness property of AFA leads to an intriguing concept of extensionality for hypersets. The classical extensionality paradigm, that sets are equal if and only if they have the same members, works fine with well-founded 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 $a = b$ if and only if $a = b$, 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 $a = b$.

The AFA universe can be depicted as in Figure 9, extending around the well-founded universe, because it includes the non-well-founded sets which are not covered by the latter.

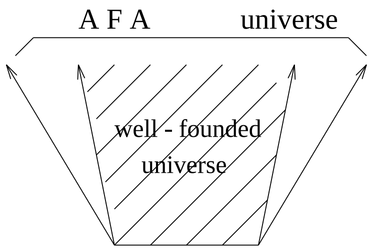

Figure 9: AFA universe extending around the well-founded universe (adapted from (Barwise & Etchemendy 1987))

3.2.1 *Equations in the AFA Universe*

Aczel's theory includes another important useful feature: *solving 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

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

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

$$\begin{aligned} x &= \{\mathrm{C}, y\}, \\ y &= \{\mathrm{C}, z\}, \\ z &= \{\mathrm{M}, x\}. \end{aligned}$$

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.,

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

The *Solution Lemma* can now be stated (Barwise & Moss 1991). If a_x is an X-set, then the system of equations $x = a_x \, (x \in X)$ 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. 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

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

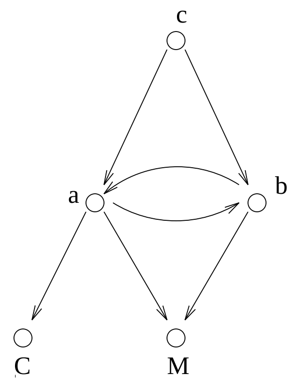

Figure 10: The solution to a system of equations (adapted from (Barwise & Etchemendy 1987))

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

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

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, a pair equations such as

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

cannot be solved since it requires the solution to be stated in terms of the indeterminate z. (These are analogous to the Diophantine equations.) 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$^-$.

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 10 with $x = a$, $y = b$, and $z = c$.

This technique of solving equations in the universe of hypersets can be very useful in modeling information which can be cast in the form of equations (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.

FatherOf

John	Bill
John	Kitty
Tom	Tim

MotherOf

Sally	Tim
Kathy	Bill
Kathy	Kitty

BrotherOf

Bill	Kitty

Figure 11: A relational database consisting of three binary relations

SizeOf

FatherOf	3
MotherOf	3
BrotherOf	1
SizeOf	4

Figure 12: The SizeOf relation defined for the database in Figure 11

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 example adapted from (Barwise 1990), the database in Figure 11 includes three binary relations: FatherOf, MotherOf, and BrotherOf. (Binary relations can be represented as sets of ordered pairs such that if an object a stands in relation R to another object b, denoted by aRb, then $\langle a, b\rangle \in R$.) 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 12. Now it may be taken for granted 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 $\text{SizeOf}\,\text{SizeOf}\,n$. 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

4.1 *Motivation*

The success of set theory in mathematics owes to the fact that all compound entities and the relations between their parts can be represented in terms of sets. We claim that this also applies to 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*. Common sense covers the fields of experience in which we all reason the same way and to the same effect. Any intelligent task requires it to some degree and designing programs with common sense is one of the most important problems in AI. McCarthy (1969) 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," and proposes a program, the Advice Taker (McCarthy 1959).

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. We can face commonsense ideas in a variety of ways: by example, counter-example, analogy, or partial description (Perlis 1988). Even then we may not consider them in terms of indivisibles but in somehow composed ways. For example, consider the following definition of "society" (adapted from the *Webster's Ninth New Collegiate Dictionary* with some modifications):

> "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, sets may still be useful in commonsense reasoning. 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 well-formed formula 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. It helps us name the unities we have formed out of entities and use those names for further reference to those unities.

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.

4.2 *Desirable Properties*

A theory proposed for commonsense reasoning should be examined from a variety of angles. We first begin with the general principles of set formation. The first choice that comes to mind is 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 13, where we have bricks as our individuals, and make towers out of bricks, and then make walls out of towers, and so on. 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 very previous stage, as in Russell's Type Theory). This provides talking about collections of previously formed objects as a new object in a safe manner and prevents entities of very large size to be formed. Because a possible commonsense set theory also needs to be mathematically precise, it should take care of the question when a collection formed of previously formed sets will be considered as a set.

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, non-profit organizations are sets of individuals and the set of all non-profit organizations is also a set; all these are expressible in the cumulative hierarchy. But what if the set of all non-profit organizations wants to be a member of itself, since it also is a non-profit organization? This is not an unexpected event (Figure 14) because this umbrella organization would probably benefit from having the status of a non-profit organization (e.g., tax exemption, etc.) (Perlis 1988).

Thus we conclude that a possible commonsense set theory should also allow circular sets to be expressed. 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 well-formed formula into a name for the well-formed formula asserting that the name has strong relationship with the formula, is presented. In this way, any set of well-formed formulas are matched with a set of names of well-formed formulas, thereby allowing self-reference by the use of names. In (Feferman 1984), the urgent 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

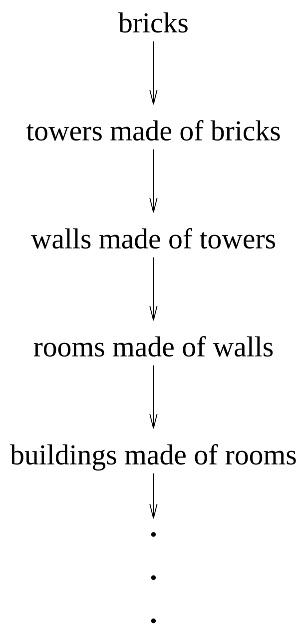

Figure 13: A hierarchical construction exemplifying the cumulative hierarchy (adapted from (Perlis 1988))

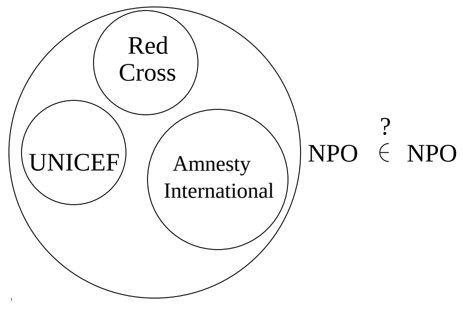

Figure 14: Can the set of non-profit-organizations be a member of itself?

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.

As an application of such a theory, we see the *Situation Theory* (Barwise & Perry 1983). Situations are parts of the reality that can enter into relations with other parts. 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 very exciting 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 public 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 then proves some theorems, including the one which states 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 15, 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 for the figure if we take E to be the property of "being expressible in English in ten words").

The model theory of common knowledge can also be studied using self-reference and situation theory (Barwise 1989d). This will be our next subject. The discussion on common knowledge will be followed by two discussions on membership and counting, respectively.

4.2.1 *Common Knowledge*

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

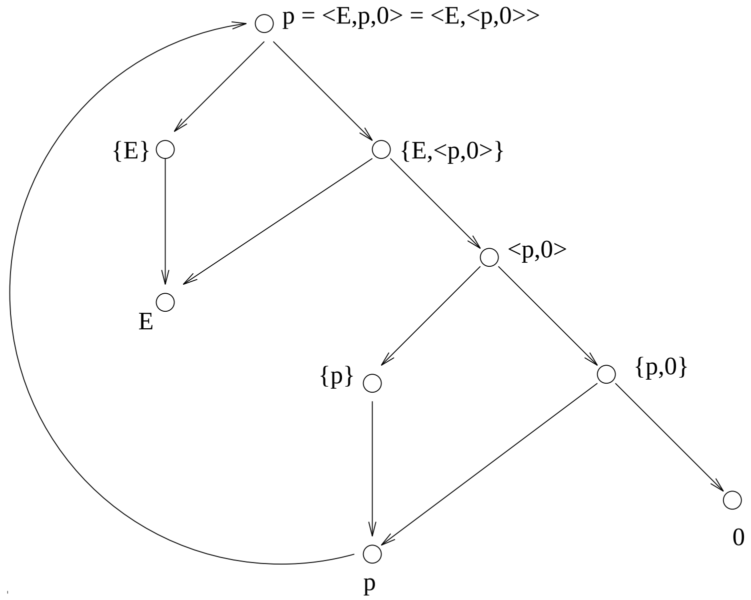

Figure 15: The picture of the statement "This sentence is not expressible in English in ten words" (adapted from (Barwise & Etchemendy 1987))

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.

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

$$\begin{aligned} x &= \{\langle K, P_1, y\rangle, \langle K, P_2, y\rangle\}, \\ y &= s \cup \{\langle K, P_1, y\rangle, \langle K, P_2, y\rangle\}, \end{aligned}$$

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

4.2.2 *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. There is another notion called *clarification*, which is a kind of general information-theoretic ordering that 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.

4.2.3 *Cardinality and Well-Ordering*

There are other set-theoretical aspects like cardinality and well-ordering issues to be considered for a commonsense set theory. We have previously stated that classical set theory does provide a precise framework for mathematics. This assertion is arguable for commonsense reasoning. Minsky (1981), for example, had mentioned that the proof of the consistency of modern set theory indicates that it is inadequate for AI purposes and he criticized the popularity of formal logic in AI arguing that some important properties of logic, e.g., consistency and completeness, may not be desirable for knowledge representation. Indeed, as McCarthy (1977) pointed out, since there is no general agreement on the fundamental structure of the world, the need for precise representations might lead to the use of imprecise or inconsistent formalizations.

The following example illustrates this point (Zadrozny 1989). Imagine a box of 16 black and 10 white balls (Figure 16). 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. But this is not true in classical set theory, because a set with finite cardinality must have a well-ordering.

Counting is an important activity to be mentioned at this point. 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 16, 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

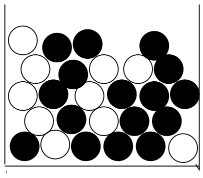

Figure 16: Balls with a finite cardinality but without a well ordering

a system which can represent a statement like "A group of kids are shouting" should probably not 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 [8]. 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).

4.3 *Some Interesting Attempts*

There is relatively little work on the use of set theory in AI. McCarthy (1980) exploited sets in his nonmonotonic reasoning method *circumscription* [15]. A weak set theory has been proposed as a specification language called SETL by Schwartz et al. (1986). Allgayer (1990) proposed an approach to introduce ways to talk and reason about sets into term languages like KL-ONE, which are widely used in natural language processing. Set theory has also been the subject of research in automated theorem proving. Data on the use of inference rules in student-constructed proofs in axiomatic set theory is presented in (Suppes & Sheehan 1981). The computer programs that are used in the computer-aided set theory course of Suppes and Sheehan represent perhaps the largest production programs created thus far for instructional purposes. 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 set theory, 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 mention some essential research efforts, by Perlis, Zadrozny, Mislove et al., and Barwise towards a possible commonsense set theory.

4.3.1 *Perlis's Commonsense Set Theory*

Perlis's approach was to develop a series of theories towards a complete commonsense set theory. He first proposed an axiom scheme of set formation for a naive set theory which he named CST_0 (Perlis 1987):

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

Here ϕ is any formula and Ind is a predicate symbol with the intended extension "individuals." This theory lacks further axioms, like an axiom of extensionality, which can be easily added. 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 Ackermann's Scheme (Ackermann 1956) which is a formal principle of this hierarchy:

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

where $HC(x)$ 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 is hierarchical and hence not able to 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, and the hierarchical theory of Ackermann (1956). GK set theory has the following axiom scheme where each well-formed formula $\alpha(x)$ has a *reification* (name) $[\alpha(x)]$ with variables free as in α and distinguished variable x

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

where y does not appear in α [16]. 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:

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

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

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

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

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 non-profit organization membership described earlier (Perlis 1988). But Perlis could not prove the consistency of CST_2 yet because this requires linking two notions of membership of the two theories.

4.3.2 *Zadrozny on Cardinalities and Well-Orderings*

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 $x \in y$ if and only if

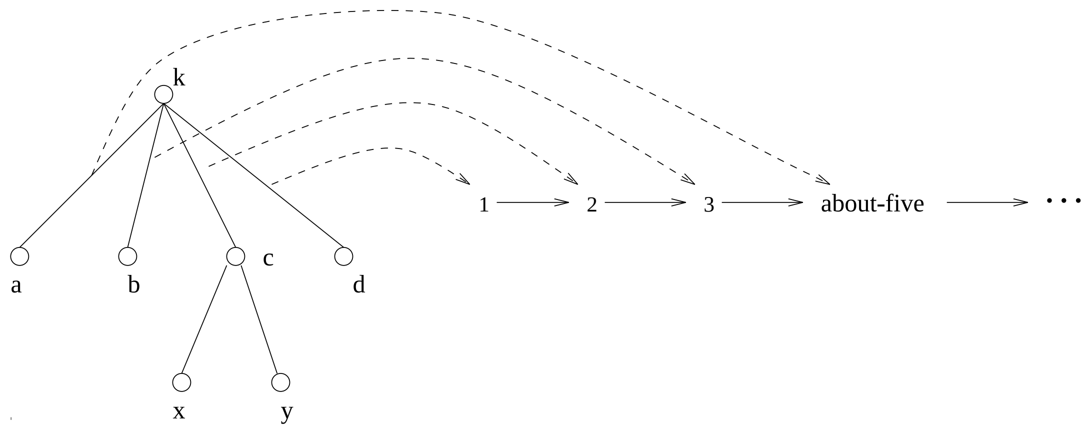

Figure 17: The one-to-one order preserving cardinality function of Zadrozny (adapted from (Zadrozny 1989))

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 17. 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 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).

4.3.3 *Protosets of Mislove et al.*

Mislove, Moss, and Oles (1990) developed a partial set theory, ZFAP, based on *protosets*, which is a generalization of HF—the set of well-founded hereditarily finite sets [17]. A protoset is like a well-founded 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.

4.3.4 *Barwise's Situated Set Theory*

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 $V_0 \subset V_1 \subset V_2 \subset \ldots$ 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 non-parametric 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.)

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 non-parametric 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 CONCLUSION

We conclude by stating that set theory can be useful in commonsense reasoning. The methodology may change, of course. A universal commonsense set theory can be developed by means of proposing new axioms or modifying existing ones. Alternatively, 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 promising and may even lead to a "mathematical metaepistemology analogous to metamathematics," as pointed out by McCarthy (1988).

ACKNOWLEDGMENTS

We would like to thank the anonymous referees of *Artificial Intelligence Review* whose comments were crucial in revising the content of this paper. We are also indebted to Patrick Suppes (CSLI), Wlodek Zadrozny (IBM T. J. Watson), Rohit Parikh (CUNY), and Ramesh Patil (USC-ISI) for their moral support. Finally, Melanie Willow (Kluwer) deserves our gratitude for her kind assistance beyond the call of duty.

Akman's research is supported in part by a grant (TBAG–992) from the Scientific and Technical Research Council of Turkey (TÜBİTAK).

DEDICATION

This survey is offered as a tribute to John McCarthy whose stimulating and challenging ideas have had and continue to have a strong and lasting effect on AI.

NOTES

1. Note, on the other hand, that another great mathematician of this century, R. Thom, has 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 well-formed formula 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. In every subset of the set of natural numbers, ordered according to increasing magnitude of its elements, there is a first element. This also holds for arbitrary ordered sets as follows. A binary relation $<$ over a set a is a *partial ordering* of a if (i) $x \not< x$ for any $x \in a$, and (ii) if $x < y$ and $y < z$, then $x < z$. (a is called a *partially ordered*

set.) A partial ordering $<$ of a is a *linear ordering* if $x < y$, or $x = y$, or $y < x$, for all $x, y \in a$. If a is a partially ordered set and b is a nonempty subset of a, and $v \in a$, then v is called the *least element* of b if $v \in b$ and $\forall x \in b(v \leq x)$, and w is called a *minimal element* of b if $w \in b$ and $\forall x \in b(x \leq w \rightarrow w = x)$.

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* if and only if there is a finite $T' \subseteq T$ such that for every ψ in T, $T' \models \psi$.

5. Cantor (1883) stated that two sets are *equipollent*, or have the same *power*, if and only if they can be put into one-to-one correspondence. The axiom for cardinal numbers states that two sets have the same *cardinality* if they are equipollent, denoted by $|x| = |y|$. m is a *cardinal number* if and only if there is a set a such that $|a| = m$.

6. 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 if and only if $\alpha \in \beta$. 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 (cf. Figure 2).

7. A group $\langle G, +, 0\rangle$ (where G is a nonempty set, $+$ is a function mapping $G \times G$ into G, and $0 \in G$) satisfies the following axioms:

$$\forall x \forall y \forall z[x + (y + z) = (x + y) + z],$$

$$\forall x(x + 0 = x),$$

$$\forall x \exists y(x + y = 0).$$

An *Abelian group* is a group G satisfying:

$$\forall x \forall y(x + y = y + x).$$

An Abelian group is *divisible* if

$$\forall n \geq 1 \forall x \exists y(ny = x).$$

It should be noted that this axiom is not first-order since there is a quantification over the set of positive integers. Hence, a divisible Abelian group is not finitely axiomatizable in first-order logic. An Abelian group is *torsion* if

$$\forall x \exists n \geq 1(nx = 0).$$

The last remark regarding axiomatizability applies to torsion groups, too.

8. 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 $x \in P(P(a \cup b))$.

9. 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.

10. 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 (Mendelson 1987).

11. 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.

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 indeed sleeping in the garden.

13. The word *paradox* refers to any statement which is contrary to common sense and intuition. The *Liar Paradox* (also known as the *Epimenides Paradox*) is one of the most famous paradoxes which disturbed philosophers, mathematicians, and linguists throughout history. According to this 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. Circumscription is a tool to jump to certain conclusions. The main point of circumscription is that the objects that can be shown to have a certain property φ by reasoning from certain facts are all the objects that satisfy φ. Thus one *circumscribes* the set of objects.

16. 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 $x \in [\neg y]$.

17. 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 (Kunen 1980).

REFERENCES

Ackermann, W. (1956). Zur Axiomatik der Mengenlehre. *Mathematische Annalen* **131**:336–345.

Aczel, P. (1988). *Non-Well-Founded Sets*. Center for the Study of Language and Information, Stanford, CA.

Akman, V. (1992). Undaunted Sets. *SIGACT News* **23**:47–48.

Allgayer, J. (1990). SB-ONE$^+$—Dealing with Sets Efficiently. In Aiello, L. C. (ed.) *Proceedings of Ninth European Conference on Artificial Intelligence*, 13–18, Pitman, London.

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 *The Situation in Logic*, 177–200, Center for the Study of Language and Information, Stanford, CA.

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

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

Barwise, J. (1989d). On the Model Theory of Common Knowledge. In *The Situation in Logic*, 201–220, 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. and Etchemendy, J. (1987). *The Liar: An Essay on Truth and Circularity*. Oxford University Press, New York.

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

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

Bernays, P. (1937). A System of Axiomatic Set Theory, I. *Journal of Symbolic Logic* **2**:65–77. (II. *ibid.* **6**:1–17, 1941. III. *ibid.* **7**:65–89, 1942. IV. *ibid.* **7**:133–145, 1942, V. *ibid.* **8**:89–104, 1943. VI. *ibid.* **13**:65–79 1948. VII. *ibid.* **19**:81–96, 1954.)

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.

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

Cantor, G. (1883). Fondaments d'une théorie générale des ensembles. *Acta Mathematica* **2**:381–408.

Cantor, G. (1932). *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts* (ed. by E. Zermelo). Springer-Verlag, Berlin.

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

Fraenkel, A. A. (1922). Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen* **86**:230–237.

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

Frege, G. (1893). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet (Volume I)*. Jena.

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.

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

Gödel, K. (1940). *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*. Princeton University Press, Princeton, NJ.

Gödel, K. (1947). What is Cantor's Continuum Problem? *American Mathematical Monthly* **54**:515–525.

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

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

Kripke, S. (1964). Transfinite Recursion on Admissible Ordinals, I–II *Journal of Symbolic Logic* **29**:161–162.

Kunen, K. (1980). *Set Theory: An Introduction to Independence Proofs*. North-Holland, Amsterdam.

Mac Lane, S. (1986). *Mathematics: Form and Function*. Springer-Verlag, New York.

McCarthy, J. (1959). Programs with Common Sense. In *Proceedings of Teddington Conference on Mechanization of Thought Processes*, 75–91, Her Majesty's Stationary Office, London.

McCarthy, J. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In Meltzer, B. and Michie, D. (eds.) *Machine Intelligence 4*, 463–502, Edinburgh University Press, Edinburgh, UK.

McCarthy, J. (1977). Epistemological Problems of Artificial Intelligence. In *Proceedings of Fifth International Joint Conference on Artificial Intelligence*, 1038–1044.

McCarthy, J. (1980). Circumscription—A Form of Non-Monotonic Reasoning. *Artificial Intelligence* **13**:27–39. (Addendum: 171–172.)

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.

McCarthy, J. (1988). Mathematical Logic in Artificial Intelligence. *Dædalus* **117**:297–311.

Mendelson, E. (1987). *Introduction to Mathematical Logic*. Wadsworth & Brooks/Cole, Belmont, CA.

Minsky, M. (1981). A Framework for Representing Knowledge. In Haugeland, J. (ed.) *Mind Design*, 95–128, MIT Press, Cambridge, MA.

Mirimanoff, D. (1917). Les antinomies de Russell et de Burali-Forti et le probléme fondamental de la théorie des ensembles. *L'Enseignment Mathematique* **19**:37–52.

Mislove, M. W., Moss, L., 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.

Morse, A. (1965). *A Theory of Sets*. Academic Press, San Diego, CA.

Mostowski, A. (1979). Thirty Years of Foundational Studies. In Kuratowski, K. et al. (eds.) *Andrzej Mostowski: Foundational Studies, Selected Works*. North-Holland, Amsterdam.

von Neumann, J. (1925). Eine Axiomatisierung der Mengenlehre. *J. für Math.* **154**:219–240. (Corrections: *ibid.*, **155**:128, 1926.)

Pakkan, M. (1993). *Solving Equations in the Universe of Hypersets*. M. S. Thesis, Dept. 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.

Parsons, C. (1990). Mathematical Intuition. In Yourgrau, P. (ed.) *Demonstratives*, 195–214, Oxford University Press, Oxford.

Perlis, D. (1985). Languages with Self-Reference I: Foundations. *Artificial Intelligence* **25**(3):301–322.

Perlis, D. (1987). Circumscribing with Sets. *Artificial Intelligence* **37**:201–211.

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.

Schwartz, J. T., Dewar, R. B. K., Dubinsky, E., and Schonberg, E. (1986). *Programming with Sets: An Introduction to SETL*. Springer-Verlag, New York.

Shoenfield, J. R. (1977). Axioms of Set Theory. In Barwise, J. (ed.) *Handbook of Mathematical Logic*, 321–344, North-Holland, Amsterdam.

Skolem, T. (1922). Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Wiss. Vorträge gehalten auf dem 5. Kongress der Skandinav. Mathematiker in Helsingförs*, 217–232.

Suppes, P. (1972). *Axiomatic Set Theory*. Dover, New York.

Suppes, P., and Sheehan, J. (1981). CAI Course in Axiomatic Set Theory. In Suppes, P. (ed.) *University-Level Computer-Assisted Instruction at Stanford: 1968–1980*, 3–80, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA.

Tiles, M. (1989). *The Philosophy of Set Theory*. Basil Blackwell, Oxford, UK.

Whitehead, A. N. and Russell, B. (1910). *Principia Mathematica (3 Volumes)*. Cambridge University Press, Cambridge, UK.

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 Part-Of Relations for Multimedia Indexing. In *Bar-Ilan Symposium on the Foundations of AI (BISFAI–93)*, Israel.

Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre. *Mathematische Annalen* **65**:261–281.