Commonsense Set Theory

Donald Perlis

University of Maryland
Department of Computer Science
College Park, Maryland 20742

Abstract: It is argued that set theory provides a powerful addition to commonsense reasoning, facilitating expression of meta-knowledge, names, and self-reference. Difficulties in establishing a suitable language to include sets for such purposes are discussed, as well as what appear to be promising solutions. Ackermann's set theory as well as a more recent theory involving universal sets are discussed in terms of their relevance to commonsense.

keywords: commonsense, sets, reification, reflection, meta-knowledge, defaults

I. Introduction

Set theory has long been known for its phenomenal success in providing a basis for virtually all of mathematics, both in the philosophical sense of a precise foundation, and in the more prosaic sense of naturalness for defining other mathematical concepts with minimum pain. Here we explore the idea that also in commonsense reasoning there is a natural role for sets, although perhaps not as a basis for all such reasoning. Despite the wide variety of knowledge representation schemes that have been proposed, set theory seems not to have been explored as a vehicle for representing commonsense knowledge. However, John McCarthy, in his address at IJCAI-85, spoke of the possible utility of set theory in commonsense reasoning. This, we will argue, is in need of doing. We will first make rather general remarks on the uses of sets, and then work our way toward more concrete and technical proposals in the concluding section.

In particular, we argue that sets provide a convenient tool for the representation of meta-knowledge, in the following way: The formalization of meta-knowledge often is undertaken via an explicit handling of syntax; for instance in [Attardi&Simi 1981], [Haas 1985,1986], [McCarthy 1979], [Moore&Hendrix 1979], [Perlis 1985], [Weyhrauch 1980]. This in turn involves the use of names for wffs, i.e., the *reification* or *reflection* We employ the word ''reflection'' in a technical sense distinct from its standard English sense of a cognitive introspection. However, there is an important parallel: our use provides formal counterpart to the *use and mention* distinction of ordinary language, which in turn seems to be a *sine qua non* of introspection. To go from believing x to ''reflecting'' that one is believing x, one turns x from a use state into a mention state, often signalled by quotation marks. It is surprising that this very simple cognitive facility turns out to be highly non-trivial from the point of view of formalism. The term ''reification'' is more particular as I am using it: it refers specifically to the creation of a syntactic term from a predicate expression. It comes from ''re-ify'': to ''thing-ify''. This is just one version of reflection, as we will see. of wffs into terms. On the other hand, sets can be viewed as vehicles for just this very thing: A set is typically an individual that corresponds to (or is the extension of) a given wff.

Ideas of matching up portions of a formal structure I shall refer to as ''reflection principles''. These can take many forms. Two are of particular interest in the present discussion: (i) what we shall call ''hierarchical'' reflection, and (ii) ''universal'' reflection. The use of hierarchies does have important uses, but so does a universal reflection principle. Our goal here is to explore both of these and seek a way to combine them. The former refers to step-like constructions that themselves are reflected (or reified) into concepts and thereby collections, as in the case of the closure of a constructive ordering such as the ''tower'' of blocks arising from the ''on top of'' relation. The latter is seemingly more ambitious, in that it forms an entity out of (virtually) any concept whatsoever without concern for its origins or possible self-referential aspects; it is more commonly called a ''comprehension'' principle.

Thus, while not totally sharp, there is a distinction between hierarchical and universal principles. Indeed, efforts to date to formalize the latter in such a way as to capture the former (as would be philosophically pleasing), have not been as successful as might have been expected. It appears this is due to the following observation: the formation of hierarchical closures depends for its power on the existence of constructions in the first place, and these seem to hinge on principles that go beyond mere universal reflection in itself. That is, universal reflection is fine, but it doesn't in itself provide enough concepts for the reflection process to yield every hierarchical construct we want. A separate rule (or set of rules) defining important constructions seems needed so that reflection can use these constructions to generate reified hierarchies. Thus without an on-top-of operator, universal reflection will not produce a tower of blocks.

The explicit first-order treatments of self-reference in [Feferman 1984] and [Perlis 1985] involve a kind of universal reflection principle that reifies any wff whatsoever into a name for the wff together with an assertion that

the name bears a strong relation to the wff it names. Hence the set of wffs is matched up with a set of names of wffs, which leads to self-reference in that a wff now may refer to its own name, or even to itself via its name. In particular, a wff may refer to all wffs by universal quantification over names of wffs and an appropriate name-named relation (what amounts to a truth predicate). In effect, this allows for a language in which the concept of ''everything'' is present.

Yet as just mentioned, although ''everything'' sounds like a lot, it isn't so much if one has few ideas about what there is. So constructions are needed to enrich the world and give more body to ''everything''. This is where hierarchical reflections come in. For example, a principle of infinity in the context of a theory of numbers amounts to reifying or reflecting the implicit infinitude of symbols 0,1,2,... into an explicit totality **N**, based on the construction ''plus 1''. The well-known Zermelo-Fraenkel set theory (*ZF*) consists largely of such constructions. See Appendix for the axioms of *ZF*.

Note that circumscription is also a kind of implicit reflection principle in that it provides a (somewhat) computationally tractable means to isolate inner models of a given set of axioms. That is, a given universe is reflected into itself (with respect to the axioms) so as to provide a minimal interpretation of that universe. See [McCarthy 1980,1984]. Thus what might otherwise be formalized as a meta-principle reduces, in the case of circumscription, to first- or second-order formulas. In effect, one is saying via such formulations that the very set of axioms (all one knows) is a totality to be taken account of in one's reasoning. However, circumscription does this implicitly, in that the set of axioms is not formally reified. This is equally true of other standard formalisms that aim at representing default knowledge, for use is made of an implicit accounting of all one knows, as in Reiter's Closed-World Assumption [1978], Reiter's Default Logic [1980], or McDermott and Doyle's Non-Monotonic Logic [1980]. That is, there is a strict hierarchical distance kept between the intended topic of discourse and the meta-principle of minimization provided by circumscription.

Sets were exploited in [Perlis 1987] in an attempt to render more explicit the minimization properties of circumscription; there McCarthy's scheme was revised to refer directly to containment properties of sets as individual entities, including the very set of axioms with respect to which the minimization is to occur. The advantage of explicit representation is clear: the meta-knowledge that is being employed then itself can be the object of analysis. For instance if a question should arise about the appropriateness of some prior conclusion (e.g., Tweety flies) made on the basis of meta-knowledge (e.g., birds fly unless known otherwise), this can be reasoned about with the same tools as for the rest of one's reasoning. One can say that it was drawn on the basis of incomplete knowledge, or that it was a guess, or that one's meta-rule is not strictly true. Such considerations are elementary for us, yet formulating then *implicitly* amounts to denying our formal models any corresponding commonsense about their own mechanisms. Some particularly vexing aspects of this are described in [Perlis 1986].

Here we will be concerned with the problem of constructing theories suitable for broad-scale reifications so that meta-knowledge can exist formally on a par with the rest of one's knowledge. As mentioned, [Feferman 1984] and [Perlis 1985] take steps in this direction, in that paradoxes of self-reference that can arise in such an undertaking were shown to be surmountable within a first-order setting. But this left untouched the issue of strengthening the axioms to allow powerful expression of relations between reified commonsense concepts, i.e., what are essentially commonsense set-theoretic notions. In effect, the analogue of a truth-schema allowing self-reference is needed in a commonsense set-theory to provide for universal sets, i.e., sets that can contain themselves. Although considered somewhat in work of Gilmore and Feferman, this must be carried out explicitly in a commonsense setting.

The rest of our presentation is as follows: in section II we set out some very general precepts indicating areas in which set theory might be expected to play a role; in section III we review particular axioms that appear to have usefulness in commonsense reasoning; in section IV we extend these to hierarchical sets along lines due to Ackermann; in section V we observe some further desiderata (''universal sets'') that arise in the use of sets in commonsense reasoning, and suggest a solution; and in section VI we provide some technical suggestions toward combining the above into a single theory that has both hierarchical and universal (full reflection) features.

II. General Arguments for Sets

The usefulness of sets in mathematics surely rests on the following observation: Apart from the indivisible units of any abstract discussion, all other entities are compound and therefore consist of structures made of other entities. But structures are in essence collections of their parts, if by ''parts'' we include whatever relationships (between other parts) pertain to the given structure. Since the very notion of a relationship is readily represented in terms of sets, the general applicability of sets to mathematics follows.

It should be pointed out that there are some areas of mathematics in which this argument falls down, specifically in precisely certain foundational considerations involving self-reference. Thus Cantor's paradox and, more recently, its analogue in category theory (the category of all categories), do not appear to have natural and satisfactory representations in traditional set theory. Nevertheless, the success of set theory as a basis for mathematics is impressive.

We are led then to ask whether the same general considerations are likely to apply to commonsense reasoning. It would appear that a *prima facie* case could be made in the above vein: in commonsense reasoning, as in mathematics, an item of discussion or thought is either considered, for the moment, to be an indivisible unit, or composed of other items, where again we view the structural relationships of such composition to be part of the compound set description. Thus democracy might be considered by some to be a relationship between a set of persons, laws, customs.

This however hints at a deeper issue, not so persistently present in the domain of mathematics. Mathematics, being largely synthetic, affords precise definitions agreed upon by the community of mathematicians. But commonsense ideas, such as democracy, appear to us not so much in terms of definitions as by example, counter-example or partial description, and even then in terms not of an underlying realm of indivisibles but in intertwined ways. This calls to mind the qualification problem, the frame problem, and natural kinds; see [McCarthy&Hayes 1969], [McCarthy 1980], and [Drapkin&Miller&Perlis 1986]. Thus one might be given the following ''definition'':

> Democracy gives people choices as to the organization and behavior of their social structures, whereas a dictatorship directs those structures without there being inputs or other recourse by the people.

Such a ''definition'' does not easily lend itself to the approach we are exploring here. The notions of country and social structure appear to be as complex as that of democracy, so that it is difficult to isolate indivisibles, or even relatively simple compounds, out of which to build a definition. The above description is perhaps best viewed as an appeal to the listener's experiences with these complex entities rather than as a definition in the mathematical sense.

Nevertheless, set theory seems to hold out at least two strengths for us. One is simply that even though many aspects of commonsense reasoning may lie beyond the usual use of sets in forming definitions, other aspects remain that do fit this traditional definitional form rather well, as we will see in later sections. The second is that whether or not a set-theoretical *definition* is provided (or possible) for many terms such as democracy, still sets are enormously useful for conceptualizing such terms. For instance, countries of the world include a number of democracies; these are often the topic of discussion, i.e., we focus on the *set* of democracies, and view it as disjoint from the set of dictatorships, even though we may not have a well-formed formula whose extension is usefully identified with either of these sets. Not only is the familiar visual picture of a set as (part of) a Venn diagram an undoubtedly important one, but also the very *formation* of internal expressions of any sort that can be used to collect imaginatively a variety of entities at once for further thought. This latter phenomenon corresponds in certain respects to the set formation process present on most set theories, often as a so-called *comprehension principle*. Its chief functions are two-fold: to provide a supply of *names* for the totalities of interest, and to state that these names in fact name what they should.

This second use of sets, as aids in conceptualizing terms when we are not provided in advance with a precise definition, is suggestive of the problem of inductive reasoning: a set or concept is supposed to exist, and a name is provided for it, but beyond that only examples, counterexamples, and analogies are given. Often in fact the problem is even harder: one is not informed of the existence of a concept worth learning, but rather is left to ferret out what concepts might or might not exist in a given realm of experience. Set theoretically one might say that many sets exist but only a few have useful or succinct descriptions. The rich language that sets afford for making contrasts between other entities would seem to be useful here. For instance, entities can be contrasted in terms of which sets they belong to, as in ''Some dictatorships are supported by democracies.'' One can picture this in the mind's eye in terms of overlapping sets, and the further issue readily comes to mind as well: are there many such instances?

We will not delve more deeply into this particular line of thought here. Our purpose in this section is to simply delineate some plausible considerations for further thought. In what follows we turn instead to areas where we are able to present more technical arguments as to the usefulness of sets on commonsense reasoning.

III. A first attempt at a commonsense set theory

In [Perlis 1987] we proposed an axiom schema in a first attempt at formalizing a naive set theory, CST_0 (''Commonsense Set Theory''), with the caveat that additional axioms will be needed for more sophisticated applications. We quickly review CST_0 here for comparison with more ambitious theories we will study. CST_0 is noncommittal with regard to hierarchies or universal reflection. That is, it does not have an axiom of universal reflection, and yet it does not specify hierarchies either except in a very weak sense, much like circumscription, simply by focussing on a fixed domain of discourse from above.

The most important notion to axiomatize is that of set formation. This also is perhaps the subtlest axiom of the standard versions of formal set theory, since care must be taken to avoid Russell's paradox. It appears that certain aspects of commonsense reasoning actually require a very strong axiom of set formation. However, in CST_0 we confine our attention to very limited kinds of set formation.

The choice for a set formation axiom in CST_0 is a weak version of the *Aussonderungs* axiom of *ZF* set theory, and indeed is equivalent to adopting a second-order theory over a ''set'' of individuals. Namely, we postulate:

$$(y)(x)(x \in y \leftrightarrow \phi(x) \& Ind(x))$$

where ϕ is any formula and Ind is a predicate symbol with the intended extension of ''individuals''. I.e., this is really a schema, saying intuitively that for any formula ϕ, there is a set consisting of all individuals having the property ϕ. Here individuals are whatever one wants them to be, Within reason. Letting *all* entities be individuals allows Russell's paradox back in the door; we either explicitly restrict the intended range of Ind, or simply observe that not everything is an individual (which in any case is provable by attempting to form the paradox). so that CST_0 can be applied in a broad range of situations. It was shown in [Perlis 1987] that this works nicely in conjunction with (first-order) circumscription to produce intuitively correct answers in situations that may be awkward or impossible for (second-order) circumscription alone. Now in CST_0 we can establish many familiar set-theoretic notions. For instance, the wff x∈s & x∈t for ϕ(x) in the above schema guarantees the existence of a set y that is the intersection of (the Individuals of) s and t. Many other such constructions follow in like fashion. Absent however, is the key notion that a set is defined by its elements, i.e., the notion of *extensionality*, which says that two sets are equal if they have the same elements. Although we could include this as an axiom of CST_0, we will not for now, as this issue will arise in a more technical setting shortly.

[Perlis 1987] points out a further objection to CST_0, namely that the richness of Ind is a critical matter in many situations, so that a theory guaranteeing a hierarchical extension for Ind is desirable. In fact, the suggestion was made of incorporating *ZF* wholesale into CST_0. Now we wish to discuss whether this is reasonable for a commonsense theory of sets, and also what further set-theoretic notions may be useful in this regard.

IV. Hierarchical Set Formation

Instead of arguing directly that *ZF* constructions such as power sets and infinite sets are needed for commonsense, we instead take a more basic approach, inquiring into general principles about set formation that have such a fundamental nature as to be difficult to exclude from ordinary reasoning. Let us begin by considering whether an intuitive set but powerful theory can be devised on the basis of CST_0, in which we experiment with the Ind predicate. One's initial impulse might be to allow everything to be an individual, but this leads immediately to Russell's paradox. So the question arises, just which things can be individuals? Consider again *Aussonderungs*:

$$(y)(x)(x \in y \leftrightarrow \phi(x) \& Ind(x))$$

This guarantees us the existence of a ''class'' y that contains all individuals satisfying ϕ. But we are not told whether y itself is an individual, and that is the whole difficulty. Clearly, for example, if ϕ refers to Ind itself then y may be ''too large'' to be an individual. But suppose that ϕ does *not* refer to Ind (does not contain the predicate symbol Ind). Then when is it reasonable to have Ind(y)? More generally, when is it reasonable for all ''things'' satisfying ϕ to form a collection that itself is an individual?

Since we are discussing individuals here as things constructed out of previously constructed things, let us use the notation HC (for *hierarchical construct*) Often the phrase ''cumulative hierarchy'' is used for the class of such constructs. instead of Ind. That is, HC(x) has intuitive interpretation of ''x can be built up as a collection from previously obtained entities''. We reserve Ind, vaguely, for any entity we wish to regard as part of a totality, whether or not built up constructively. Thus a column of blocks is a hierarchical construct, as is a wall of columns, a room of walls, etc. We ask the question: what is it that allows certain ϕ-things to be so collected from below? The obvious answer is that all ϕ-things are already hierarchical. Once all blocks are granted to be HC's, a column of blocks can be collected into a new HC; and once columns are thus made into HC's, they can be collected in turn.

Ackermann [1956] has provided a very streamlined set theory *A* based on a schema that can be seen as elevating this notion into a formal principle, which we formulate here in terms of our predicate HC:

$$HC(y_1)\& \ldots \& HC(y_n)\&(x)(\phi(x) \rightarrow HC(x)) \; . \rightarrow . \; (z)[HC(z)\&(x)(x \in z \leftrightarrow \phi(x))]$$

for all ϕ not containing the symbol HC and having free variables among x, y_1, ..., y_n. Ackermann's full theory includes as well the axiom of transitive completeness of HC That is, all elements of elements of HC, and all subsets of HC, are themselves elements of HC. extensionality (for all sets, not only those in HC), and a comprehension principle like that of CST_0 except for HC instead of for Ind.

The biconditional subformula of the above schema supplies a rich variety of collections which may or may not be well-behaved enough to be among the hierarchical constructs HC. The remaining portion of the schema to its right gives a further condition sufficient for such a collection to in fact be in the hierarchy. Note that intuitively this tells us that if we are thinking of a property ϕ, and if ϕ is not self-referential (well-behaved with respect to previously-collected entities) then it makes sense to speak of ϕ as a finished thing, i.e., reified into an existing hierarchy of discourse. In particular, *A* guarantees a very rich domain of individuals as described in [Perlis 1987]: given blocks, it will generate towers of blocks, walls of towers, rooms of walls, buildings of rooms, cities of buildings, etc. Again, we can easily form many familiar set constructions, such as intersections and powers, simply by choosing the appropriate wff for ϕ. But now there is an added feature, namely, the results of such choices are sets *in HC* ! So we get an

ever-expanding supply of new sets in the hierarchy. While this is interesting, the real significance is that these new sets, being in HC, then are fodder for another round of application of the schema. Thus *A* becomes quite a powerful tool. Let us use the name CST_1 for the theory consisting of the above version of Ackermann's schema together with CST_0 and the axiom HC(x) → Ind(x).

I propose that CST_1 captures the intuitive sense of collecting previously gotten individuals into a new individual, in a way appropriate to commonsense reasoning. However, experience with set-theoretic paradoxes shows us that we should be cautious about adopting new axioms. Nevertheless, in this case we are on safe ground. It has been shown by Levy [1959] that Ackermann's theory *A* is consistent relative to *ZF*. Indeed, results of Levy and of Reinhardt [1970] show that when combined with a further axiom (Regularity), See Appendix. *A* is strongly equivalent to *ZF* ! Although there are important differences. For one thing, *ZF* was devised principally as a foundation for everyday mathematics, while A seems to have served largely as a spring-board to large cardinals, as in [Perlis 1972] and [Reinhardt 1974]. In fact, A is usually motivated in terms of a reflection principle concerning the universe of classes: that the class HC of sets reflects all ''HC-free describable'' properties of the universe. In this respect it resembles circumscription; and it also is only a hierarchical reflection, in that one is forbidden to reflect the extension of HC itself; the treatment exploits hierarchies to the utmost but does not tread beyond that into self-reference or full reflection of *all properties*. Thus whether or not infinite sets, power sets, and so on, are explicitly asserted, they are consequences of Ackermann's ''collection-principle''.

Theorem: CST_1 is consistent (relative to *ZF*).
Proof: We know from Levy that *A* is consistent (relative to *ZF*). But CST_1 is consistent relative to *A*, since we can interpret Ind to be HC and then any model of *A* will be one of CST_1 as well.

Have we then found in CST_1 the commonsense set theory sought? While it does bear on issues, such as the acceptability of *ZF*-style axioms, and while it certainly is powerful in that it supplies a very rich interpretation for HC, it unfortunately suffers from being totally hierarchical That is, Ind is given no teeth to distinguish it from HC, although that possibility is left open. and so does not provide for a set that can contain itself. This latter phenomenon however is critical for explicit representation of meta-knowledge, and corresponds to the concerns addressed in [Feferman 1984] and [Perlis 1985]. We consider this in the following section.

V. Difficulties and possibilities

Here we argue that a much stronger set theory is needed for commonsense reasoning, and that a plausible approach is possible on the basis of ideas in [Feferman 1984] and [Perlis 1985]. The theory CST_1 developed above is useful in situations involving ''levels'' of sets, i.e., sets of sets or of individuals. But for many purposes this is insufficient. For instance, if people are viewed as individual elements, then organizations can be viewed as (or associated with) Of course, organizations are more complex than mere collections of individual members. sets of individuals, e.g., the Boy Scouts of America (*BSA*) as the set of boy scouts; and then *non-profit* organizations together form (or can be associated with) a set of such sets. Such situations can be reasonably expressed in CST_1. However, the situation can be even more complicated. For we can imagine an organization consisting of (or having as members) various non-profit organizations (say, The Non-Profit Organizations of America, or NPOA), itself also being a non-profit organization. It is then reasonable to consider that NPOA might be one of its own member organizations. Although it might appear to be an unexpected event, it is not outright impossible nor even without conceivable practicality. (For instance, NPOA might benefit by law from being a member of a non-profit organization.)

It is therefore appropriate to view the expression of such a possibility as part of commonsense reasoning, and to require a formal language for such reasoning to allow such expressions. This however can lead into the treacherous waters of Russell's paradox. A similar issue was addressed in [Feferman 1984] and [Perlis 1985], in which it was found that a suitably arranged treatment of self-reference can be both powerful and consistent. This suggests the following approach.

We propose a theory, which we shall call *GK* (''Gilmore-Kripke set theory''). Much the same theory is called *PST´* by Gilmore, and $CA_{+/-}$ by Feferman. One (inessential) difference is that these latter employ *two* membership notions, corresponding to a∈[b] and a∈[¬b] in *GK*. *GK* has the following axiom schema, where each wff $\alpha(x)$ has a corresponding reification (name) $[\alpha(x)]$ with variables free as in α and distinguished variable x:

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

Here y does not appear in α and the * operator is a short-hand for the result of first writing all → symbols in terms of & and ¬, then passing negations in α through to predicate letters, and finally replacing each occurrence of a subformula ¬a∈[b] in the result by a∈[¬b]. (This corresponds exactly to the schema True($[\alpha]$) ↔ α^* Recall from [Perlis 1985] that the sole function there of the * is to interchange ¬ and True when they occur suitably juxtaposed in α. as given in [Perlis 1985], so that we have here a set theory based on a theory of truth.) We also include in *GK* the definitional axiom

$$\text{Equiv: } s \approx t \leftrightarrow (x)(x \in s \leftrightarrow x \in t)$$

which provides notation to express when two names correspond to the same underlying collection.

There is the immediate question as to whether *GK* is consistent, for it is not obvious that *GK* can be viewed as a subtheory of a standard set theory such as *ZF*. However, here our worries can be laid to rest:

Theorem: *GK* is consistent (relative to *ZF*).
Proof: *GK* is simply a version of the theory of truth given in [Perlis 1985], which was proven there to be consistent (using methods easily formalized in *ZF*).

In fact, CST_1 together with the axiom of regularity can prove consistency of *GK*, so that in some sense *GK* is a weaker theory; but this is illusory, for *GK* embodies principles not provable in CST_1, as we shall see in the next section. Once more, the theory in question (*GK*) allows many familiar results to be established, such as the existence of intersections. However, a new twist occurs, that has features difficult to incorporate into a familiar setting, and this is due to the self-reference that is allowed. For now, certain constructions must be handled (automatically by *GK*) in a way that prevents Russell's Paradox, and this has the consequence of allowing sets to exist that nonetheless do not behave as expected. One severe instance is that of the power set. Although we can define what *seems* to be the power set p of a set t, using the wff $(x)(x\in z \rightarrow x\in t)$ for $\phi(z)$, p still may not in fact contain all subsets of t since this depends on the * of ϕ rather than simply ϕ itself. In cases where these are the same (as will happen for ''ordinary'' sets) we do get p being the full power set of t; but in other cases, there may be no set which is exactly the power set, and p as *-power set may be the best we can do. Another odd feature of *GK* is that it deals not with sets *per se* but with names of sets, or attributes, so that extensionality fails. This we address more later.

An example

Suppose we are given a predicate $NP(x)$ for non-profit organizations x. Then the existence of the collection of non-profit organizations follows automatically from *GK*. For we simply take that very predicate expression and reify it: $[NP(x)]$. It is a name, of which *GK* formally asserts that it applies to (or has as members) those organizations that satisfy NP: $y\in[NP(x)] \leftrightarrow NP(y)$. We have left off the * operator here since we have taken *NP* to be atomic and not explicitly involving the notion of truth or membership. Now, whether or not this new entity $[NP(x)]$ itself corresponds to an organization (which then may or may not itself be non-profit) is not a matter for logic to decide; it is contingent on events of the world. However, it provides us with the *concept* of such a collection, and from there an enterprising organizer might try to convince others that such an idealization would be useful in practice. Discussions that might ensue would involve serious talk about properties of $[NP(x)]$ in its ideal as well as potentially organizational states.

Suppose now that an organization with the name *NPOA* has been created and that many (and only) non-profit organizations have become members. We may have reason to suppose that, by and large, $NP(x) \rightarrow x\in NPOA$. Let us treat this as a default rule, so that a non-profit organization is assumed to belong to *NPOA* unless known otherwise. Under plausible conditions, we can actually prove (without risking inconsistency or paradox) that $NPOA\in NPOA$. Of course, we will not be shocked if we learn that this is not the case; yet we should be able to see when our axioms have this implication. The theory *ORG* will now be formalized out of the above precepts for organizations, among others.

If we further take as a default that an organization is profit-making unless known otherwise, it is reasonable to inquire whether circumscription can produce the intuitively correct conclusions. But General Motors (*GM*) is not known to be a member of a profit-making organization, nor is it known to be non-profit, so by default we wish to assume it is for-profit. Circumscribing *non-profit* should do the trick. Since we will already consider the circumcription of $NP(x)$ & $x\notin NPOA$ for the first default above, there is no need for a further (multiple) circumscription to achieve this second goal as well.

Specifically, let *ORG* have the following axioms:

$$NP(x)\&x\in y.\rightarrow NP(y)$$

Let us suppose by legal definition any member organization of a profit-making organization is itself a profit-making organization. For instance, if *BSA* is non-profit and a member of *NPOA*, then so is *NPOA* non-profit.

$$NP(BSA)$$
$$BSA\in NPOA$$
$$\neg GM\in NPOA$$

Let *CIRC* be first-order formula circumscription of the formula $NP(x)$ & $x\notin NPOA$ with respect to *ORG*. Then the following are readily established in *GK*+*ORG*+*CIRC*:

$NPOA\in NPOA$.
This we obtain by first deriving $NP(NPOA)$, and then using the wff $x=NPOA \vee x=BSA$ in the circumscriptive

schema.

$\neg NP(GM)$.

Now this follows from the same circumscription, since it can be proven that $GM \neq NPOA$ and $GM \neq BSA$.

VI. A Proposed Synthesis

Finally we wish to suggest a synthesis of a universal (reflection) theory *a la* Gilmore-Kripke and a hierarchical theory *a la* Ackermann. This we shall call CST_2; its axioms consist of those of *GK* together with:

Ext1: $ext\ s = ext\ t \leftrightarrow s \approx t$

Ext2: $s \approx ext\ s$

Ext3: $z \in HC \rightarrow (w)(z = ext\ w)$

$A_{ext:}\quad y_1, \ldots, y_n \in HC \& (x)(\phi x \rightarrow x \in HC)\ .\rightarrow.\ ext[\phi] \in HC \& (x)(x \in [\phi] \leftrightarrow \phi(x))$

We take the notational liberty of using HC both as predicate and as term.

Here A_{ext} and Ext1-3 provide extensional constructions i.e., collections determined solely by their members. while still further sorts of collections are permitted as well due to *GK*'s contribution. The two formations are related, since a collection formed as an extension of a formula ϕ *a la GK* could turn out to have as well a hierarchical construction in which case the name given it by *GK* should be related to the latter. This is provided in A_{ext}, specifically, by the presence of $[\phi]$ and ext $[\phi]$. That is, whereas *GK* is essentially name-oriented, HC has the intended orientation of underlying sets rather than their names. So ext $[\phi]$ is the underlying set corresponding to $[\phi]$. It is as if someone (in the role of *GK*) postulates an abstract entity, and someone else (in the role of A_{ext}) shows that it is concretely present. This is indeed a familiar form of rational experience, both in scientific investigations and also in everyday deliberations. Put differently, *GK* provides a sort of philosophical architecture, and A_{ext} some constructive engineering techniques, for sets.

The following now are easily established:

Lemma: CST_2 |-- $(ext\ s \approx ext\ t)\ .\leftrightarrow.\ (ext\ s = ext\ t)\ .\leftrightarrow.\ (s \approx t)$

Proof: The second $\leftrightarrow$ equivalence is simply axiom Ext1. For the first $\leftrightarrow$ equivalence, suppose the left-hand side. Then if z∈ext s, we also have (by definition of ≈) z∈ext t, and conversely, hence the right-hand side holds. The other direction follows immediately from the usual first-order axioms of equality.

Theorem: CST_2 |-- $(x \in HC)(y \in HC)(x \approx y \leftrightarrow x = y)$

Proof: By axiom Ext3, every element of HC is itself an extension, and by Ext1, extensions are determined solely by their members. So, extensionally equivalent elements of HC are equal. More formally, if x=y then we have $x \approx y$ as in the Lemma, from axioms of equality. Conversely, if $x \approx y$, then ext x = ext y from Ext1; but from Ext3, we have x = ext u and y = ext v for some u and v, so ext ext u = ext ext v. But from the Lemma we then find ext u = ext v, i.e., x=y.

Thus in HC we have a universe of sets not unlike the classical sets of everyday mathematics, in which sethood is determined solely by abstraction (extension) rather than by name (intension). This has close ties to commonsense in that we dance a fine line between intensions and extensions in our reasoning, as witness the famous *de dicto/de re* distinction.

We have not specified CST_2 as an extension of CST_1; however, this is not an oversight, since each instance of the schema of CST_1 is easily proven in CST_2. Thus CST_1 is a subtheory of CST_2. The (relative) consistency of CST_2 is an open question. It would appear that a model *a la* Gilmore/Kripke built on top of a model for *A* ought to show CST_2 consistent relative to *A*; this would however involve linking two notions of membership: that of *A* and that of the *GK*-extension of *A*.

It would be premature to make grand claims for CST_2. However, I hope to have illustrated that set theory has a role to play in commonsense reasoning, and that it will be a fairly sophisticated one, rivaling the best current foundational theories of sets in mathematical logic. This should be no surprise. The (set-theoretic) foundations of mathematical logic are deeply and appropriately tied to the philosophy of reasoning in general.

Acknowledgements

This research has been supported in part by the following institutions:
The U.S. Army Research Office (DAAG29-85-K-0177)
The Martin Marietta Corporation.

I would like to thank W. Ian Gasarch for helpful comments.

Appendix: The axioms of Zermelo-Fraenkel set theory.

1. Extensionality: $(z)(z \in x \leftrightarrow z \in y) \rightarrow x = y$

2. Regularity or Foundation: $(y)(y \in x) \rightarrow (y)(y \in x \& \neg(z)(z \in x \& z \in y))$

3. Aussonderungs or Subset schema: $(z)(y)(x)(x \in y \leftrightarrow x \in z \& \phi(x))$

4. Power set: $(y)(x)(x \in y \leftrightarrow (z)(z \in x \rightarrow z \in w))$

5. Infinity:
$$(w)(y)[\neg y \in w \ \& \ (x)(x \in w \rightarrow (z)(z \in w \ \& \ (u)(u \in z \leftrightarrow u = x \ or \ u \in x)))]$$

6. Replacement schema:
$$(x)(y)(z)(\phi(x, y) \& \phi(x, z) \rightarrow y = z) \ . \rightarrow . \ (w)(y)(y \in w \leftrightarrow (x)(x \in s \ \& \ \phi(x, y)))$$

7. Sum or Union set: $(y)(x)(x \in y \leftrightarrow (z)(x \in z \ \& \ z \in w))$

8. Empty set: $(y)(x) \neg x \in y$

9. Pair set: $(y)(x)(x \in y \leftrightarrow x = s \ or \ x = t)$

Axioms 7, 8, and 9 are redundant, i.e., they follow from 1-6. Moreover, the concepts (and vast bulk of results) of modern mathematics are definable (or provable) in *ZF*. See [Shoenfield 1967] for a concise development of some basic notions, and [Drake 1974] for general background on set theory related to large cardinals. One further ''axiom'' often employed together with *ZF* is the axiom of choice. This has important uses in mathematics and logic. Its significance for commonsense is unclear.

Bibliography

Ackermann, W [1956] Zur Axiomatik der Mengenlehre. *Math. Annalen*, v. 131, 336-345.

Attardi, G. and Simi, M. [1981] Consistency and completeness of OMEGA, a logic for knowledge representation. IJCAI-81, 504-510.

Drake, F. [1974] *Set Theory*. North-Holland.

Drapkin, J, Miller, M., and Perlis, D [1986] The two frame problems. Draft.

Feferman, S. [1984] Toward useful type-free theories, I. *J. Symb. Logic*, 49, 75-111.

Haas, A. [1985] Possible events, actual events, and robots. *Comp. Intelligence*, v. 1, 59-70.

Haas, A. [1986] A syntactic theory of belief and action. *Artif. Intelligence*, v. 28, 245-292.

Levy, A. [1959] On Ackermann's set theory. *J. Symb. Logic*, v. 24, 154-166.

Lifschitz, V. [1985] Computing circumscription. IJCAI-85, 121-127.

McCarthy, J. [1979] First-order theories of individual concepts and propositions. *Machine Intelligence*, 9.

McCarthy, J. [1980] Circumscription -- a form of non-monotonic reasoning. *Artificial Intelligence*, v. 13, 27-39.

McCarthy, J. [1985] Acceptance Address, IJCAI Award for Research Excellence.

McCarthy, J. [1986] Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, v. 28, 89-118.

McDermott, J. and Doyle, J. [1980] Non-monotonic logic I. *Artif. Intelligence*, v. 13, 41-72.

Moore, R. and Hendrix, G. [1979] Computational models of belief and semantics of belief sentences, Tech. Report 187, SRI Intl.

Perlis, D. [1972] An extension of Ackermann's set theory. *J. Symb. Logic*, v. 37, 703-704.

Perlis, D. [1985] Languages with self-reference I. *Artificial Intelligence*, v. 25, 301-322.

Perlis, D. [1986] On the consistency of commonsense reasoning. *Comp. Intelligence*, 2, 180-190.

Perlis, D. [1987] Circumscribing with sets. *Artificial Intelligence* 31, pp 201-211.

Reinhardt, W. [1970] Ackermann's set theory equals *ZF*. *Annals of Math. Logic*, v. 2, 189-249.

Reinhardt, W. [1974] Set existence principles of Shoenfield, Ackermann, and Powell. *Fundamenta Mathematica*, v. 84.

Reinhardt, W. [1986] Some remarks on extending and interpreting theories with a partial predicate for truth. *J. Phil. Logic*, v. 15, 219-252.

Reiter, R. [1978] On closed world databases, in: H. Gallaire and J. Minker (eds.) *Logic and Databases* (Plenun, New York) 55-76.

Reiter, R. [1980] A logic for default reasoning. *Artificial Intelligence*, v.13, 81-132.

Shoenfield, J. [1967] *Mathematical Logic*. Addison-Wesley.

Weyhrauch, R. [1980] Prolegomena to a mechanized theory of formal reasoning, *Artificial Intelligence*, v. 13, 133-170.