Circumscribing with Sets

Donald Perlis

University of Maryland Department of Computer Science College Park, Maryland 20742 perlis@maryland.arpa (301) 454-7931

Abstract: Sets can play an important role in circumscription's ability to deal in a general way with certain aspects of commonsense reasoning. A result of Kueker indicates that sentences that intuitively one would want circumscription to prove, are nonetheless not so provable in a formal setting devoid of sets. Furthermore, when sets are introduced, first-order circumscription handles these cases very easily, obviating the need for second-order circumscription. The *Aussonderungs* axiom of ZF set theory plays an intuitive role in this shift back to a first-order language.

descriptors: commonsense, circumscription, sets

I. Introduction

Commonsense requires for many purposes the notion of sets [McCarthy 1985]. Second-order circumscription in particular exploits this fact, in that it is a kind of weak set theory. But it does not introduce sufficient set theory for commonsense reasoning in general. On the other hand, if we take as our underlying language for commonsense reasoning one that has a rich (first-order) set theory, then second-order circumscription is superfluous, and first-order circumscription (with variable predicates) is sufficient. Thus there is no need to pass to a second-order language at all. This in turn has the well-known advantages of familiarity, ease of comparison with other first-order formalisms, and (relative) computational facility. Moreover, other technical benefits accrue from the use of sets in circumscription, as will be seen¹.

An example of Kueker will be analyzed as implicitly involving a notion of set. This in turn suggests that sets be made explicit in a formal language for commonsense reasoning. We explore this and show that it overcomes the problem Kueker found, and also appears to be more general than the approach of second-order circumscription.

¹Thus the present work is closer to the commonsense side of circumscription research, as in [McCarthy 1986], as contrasted with certain proof-theoretic and semantical studies such as [Etherington&Mercer&Reiter 1985], [Lifschitz 1985,1986] and [Perlis&Minker 1986].

We will proceed as follows: In Section II we recall Kueker's example and discuss it relative to commonsense and to existing solution routes. In Section III we propose another solution, namely the introduction of a (fi rst-order) set-theoretic language for commonsense, and argue its merits. In Section IV we return to Kueker's example, and in Section V we present a "pure" set-theoretic version of circumscription. In Section VI we discuss briefly the infusion of "massively set-theoretic"² notions into commonsense reasoning.

II. Kueker's example

To initiate our discussion, we begin with Kueker's example. Let K[P] be the theory given below, in which P is a predicate letter, b and c are constants, and s is a function symbol.

Pb
b≠sx
$c \neq sx$
b≠c
$sx = sy \rightarrow x = y$
$Px \leftrightarrow Psx$

Then it turns out, as Kueker has shown, that the sentence

¬Pc

is not a theorem of the circumscribed theory Circum(K[P])³. However, the intended semantics providing

 $A[Z_1,...,Z_n] \& (\forall x)(E[Z_1,...,Z_n] \to E) \to (\forall x)(E \to E(Z_1,...,Z_n))$

²I am punning on the oft-heard phrase "massively parallel".

³I use Circum for fi rst-order variable circumscription, as described in [Perlis&Minker 1986], namely,

for wffs $Z_1,...,Z_n$ where $E = E[P_1,...,P_n]$ is a formula in which $P_1,...,P_n$ may appear, and $E[Z_1,...,Z_n]$ is obtained from E by substituting Z_i for each P_i . Here the idea is that E is a formula involving predicates P_i whose extensions may fail to be fully specified by axioms, so that several interpretations of E are possible. By considering various interpretations Z_i of those predicates, it may be found that certain choices lead to smaller extensions than others. By insisting, via the schema, that E have no strictly smaller interpretation

the motivation for circumscription *does* have \neg Pc as a circumscriptive consequence of K[P], so that in the language of K[P] predicate circumscription does not accomplish its intended goal. In effect, circumscription attempts to minimize the extension of designated predicates by replacing them with wffs of potentially smaller extensions. But in the language of K[P], there is no wff that singles out the transitive closure of s on {b}, namely, the set of successors of b.

A commonsense "blocks world" interpretation of this situation is given in [Perlis&Minker 1986]: A king has decreed that all marble blocks in the kingdom are to be painted in certain ways, such that any block situated in the exact center 'b' of the floor of the royal mint is to be purple, purple blocks may lie on or below only purple blocks (or on the floor), and--since purple is a royal and exclusive color--the number of purple blocks is to be minimized. Here the universe of discourse can be thought of as positions x that blocks could occupy, Px as saying that any block at x is to be purple, and sx as the position immediately above x. Then Pb guarantees that only a purple block may be at position b, and similarly for the other axioms.

Now let 'c' be another location on the floor. Commonsense would seem to immediately show that no block on the floor at positions other than b should be purple, and in particular that $\neg P(c)$. Yet Kueker's result shows that $\neg Pc$ is *not* provable by (predicate) circumscription, although one would hope otherwise, since only blocks in the "column" over b are intuitively forced to be purple. Indeed, in the intended ("P-minimal") models of Circum(K[P]), only these blocks will be purple, and $\neg Pc$ therefore is true there. Note that a column corresponds intuitively to a set of blocks.

An obvious solution would be to specify that c is not among the blocks of the column. This could be expressed as $c\neq b$, $c\neq sb$, $c\neq ssb$, etc., which is an infinite set of formulas and therefore not suitable for circumscription. What we need is a way to refer to the entire column at once, i.e., in a single (finite) formula. No first-order formula in the language given is available for this purpose. Perlis and Minker [1986] introduce a new symbol "tower" into the original language that is intended to apply precisely to the positions b

than itself, we guarantee that E is already minimal. Of course, this guarantee is only as good as the range of choices for the Z_i that is afforded by the language employed. This is the point being explored here in terms of a rich language of sets.

and above (and in particular to no positions on the fbor other than at b). In effect, they introduce a single set into the theory. However, as they point out, "this is not a fully satisfactory solution, since it was necessary to augment the theory in a non-obvious way; in effect we had to mentally circumscribe in order to see how to alter the theory." Note that this solution "works" precisely because it enriches the language so that the needed concept is expressible.

Now, McCarthy [1986] introduced a more powerful variation of circumscription, called *formula* circumscription, which has at least two advantages over the original formulation: it allows so-called *variable* predicates in addition to the predicate being circumscribed, and also is couched in a second-order language.⁴ It is the second feature that is directly relevant to Kueker's example, for in a second-order language one can exhibit an explicit formula⁵ that expresses the notion of column as above. That is, certain sets of blocks are definable in this theory, so that this approach also amounts to introducing a certain amount of set theory into the formalism.

However, Kueker's example serves to raise the prospect that no matter what devices we build into a language for the specific expression of particular sets of some level, there may be higher levels--e.g., walls (of columns), rooms (of walls), buildings (of rooms), etc.--so that third- or higher-order languages may be needed, throwing some doubts on the primacy of second-order languages for expressing minimal closures of functions.

In fact, a "tower" predicate is a reasonable thing to have. The trick is to provide a language broad enough so that it will be there when needed rather than added by us in an *ad hoc* manner. Second-order circumscription can be assessed in this light. It does provide automatically a supply of notions: sets of individuals. But as we have suggested in the preceding paragraph, this is not nearly enough.

Indeed, the introduction of a second-order language into circumscription seems to gain its power not because of a revised version of the circumscription schema *per se*, but because the language in which

⁴Formula circumscription also allows *formulas*, rather than simply predicate letters, to be circumscribed, whence it's name. However, this feature appears not to afford any real generalization beyond that provided by variable predicates [see Perlis&Minker 1986], unless advantage is taken of names for formulas as is suggested below.

 $^{{}^{5}}E.g.$, the second-order wff ($\forall Q$)[Qa & ($\forall y$)(Qy \rightarrow Qsy) . \rightarrow Qx] where Q is a second-order variable.

formulas in general are written is more expressive (and indeed amounts to a weak set theory). It seems appropriate to separate these two issues: on the one hand, one wants a circumscriptive formalism that states the notion of minimality in intuitively correct terms--and here fi rst-order circumscription seems to be adequate as long as variable predicates are allowed--and on the other hand, a sufficiently broad language is needed to produce formulas to express whatever concepts may be useful in a given domain.

Our proposal then is to seek a very expressive (first-order) language, and then use (first-order) variable circumscription. In particular, our contention here is that Kueker's situation arises because, firstly, the notion of a column is a natural ("commonsense") one based on an underlying mental picture of transitive closure (itself perhaps based on something like naive set theory), and secondly, set theory is not currently part of formal approaches to commonsense reasoning. Our proposed solution then is to incorporate a sufficient amount of set theory into formal commonsense reasoning, to capture intuitions such as those providing the construct of transitive closures. The first-order circumscriptive schema can then be retained (in the form having variable predicates).

III. An Elementary Commonsense Set Theory

We propose below an axiom schema in a first attempt at formalizing a naive set theory, CST_0 , for commonsense reasoning, with the caveat that additional axioms will be needed for more sophisticated applications.

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 the present section we will confi ne our attention to very limited kinds of set formation.

Our initial choice then for a set formation axiom is a weak version of the *Aussonderungs* axiom of Zermelo-Fraenkel (ZF) set theory, and indeed is equivalent to adopting a second-order theory over a "set" of individuals. Namely, we postulate:

$$(\forall x)(\forall x)(x \in y \equiv \phi(x)\&Ind(x))$$

where ϕ is any formula. 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 . Thus if we take countries as our individuals, there is a set of all countries, since we may take ϕ to be an identically true property; and there is a set of all democratic countries, since we may take ϕ to be the property of being a democracy.

IV. Kueker's example reconsidered

We note that CST_0 is much like a second-order logic in that all sets are of individuals. We have some freedom in deciding what are the individuals, however, and this does put us in a better position to handle what otherwise would require still higher-order languages. (More on this later.)

Let J[P] be the theory K[P]+ CST_0 , where we interpret the predicate Ind of CST_0 as the property of being the potential position of a block, so that the following additional axioms are implicitly assumed as well:

```
Ind(c)Ind(x) \rightarrow Ind(sx)P(x) \rightarrow Ind(x)
```

Now the rest is easy. We simply consider the formula ϕ as follows:

$$\phi(\mathbf{x}) \equiv P(\mathbf{x}) \& (\forall \mathbf{w})([\mathbf{b} \in \mathbf{w} \& (\forall \mathbf{z})(\mathbf{z} \in \mathbf{w} \rightarrow \mathbf{s} \mathbf{z} \in \mathbf{w})] \rightarrow \mathbf{x} \in \mathbf{w})$$

Note that the clause $(\forall w)([b \in w \& (\forall z)(z \in w \rightarrow sz \in w)] \rightarrow x \in w)$ says that x is contained in every set that is closed under succession and contains b, i.e., x is in the column over b. The formula $\phi(x)$ then intuitively defines the column of purple block positions above b, and also serves our purpose perfectly.

We employ first-order circumscription of P by J[P]:

$$[J[Z] \& (\forall x)(Z(x) \to P(x))] \to (\forall y)(P(y) \to Z(y))$$

Substituting ϕ for P in J[P] produces J[ϕ] which is readily provable from J[P]. Since also $\phi(x) \rightarrow P(x)$, first-order circumscription yields at once that $P(x) \rightarrow \phi(x)$. It remains only to show $\neg \phi(c)$ in order to conclude $\neg P(c)$. But it is easily seen that $\neg \phi(c)$, since if any w in the bracketed formula above were to contain c, then let w' be w - {c} (this new set is guaranteed by our axiom of set formation). But w' also obeys the bracketed formula, since we know $c \neq sx$ for any x. But clearly c is not in w', so $\phi(c)$ cannot hold. This then gives Kueker's intuitive but missing conclusion $\neg P(c)$.

V. Recovering and extending full second-order circumscription

Here we point out that CST_0 restores the power of second-order circumscription when suitably applied to any theory A[P]. Recall in particular that second-order circumscription can be formalized as a single second-order wff. This is a pleasant feature of that version of circumscription, even though (as McCarthy has noted) it so far has had no applications. This feature will also emerge in a suitable fi rst-order version of circumscription using sets, as will be seen. Let A'[P] be in a second order language L', and let L be a first-order language in which all predicate variables and constants of L' have been "translated" as individuals⁶, and in particular in which each predicate P_i has been replaced with a set constant p_i , except for \in and =. If any P_i is not monadic, p_i will then be a set of tuples which we also assume expressible in L. We assume the axioms of CST_0 to be in a theory T in the language L. We also take the L-translations of all wffs α of A'[P] to be axioms of T; each such translation we call α 's " \in version". In particular, the given axioms A'[P_1, \ldots, P_n] now are A(p_1, \ldots, p_n). Now we introduce the following predicates of *set circumscription*:

$$C_A(p_i): \neg (x_1) \dots (x_n) [A(x_1, \dots, x_n) \& x_i \subset p_i]^7.$$

It follows that from $C_A(p_i)$ we can prove the \in versions of all the theorems of second-order circumscription of P_i (including the circumscriptive axiom itself, for this is simply $C_A(p_i)$) since the rules and axioms for second-order logic have \in versions that are theorems in the theory T.

There are two added bonuses to the set approach that we point out now. First, the "external" character of circumscription can be somewhat avoided. That is, the wff P_i that is circumscribed is chosen for minimization independently of the formal theory. Someone simply decides to include the schema or axiom of circumscription. However, ideally, there should be a formal deduction showing that a predicate letter P is to be minimized under certain conditions. The obvious way to do this would be to include *all* instances of circumscription, but prefixed by the condition that the given wff be one that is to be minimized. However, we have no first-order predicates available to state that a predicate letter P is to be minimized.⁸ On the other hand, if we replace P by a set p, then we may introduce a first-order predicate Min for this purpose; we simply conjoin the axioms

$$\operatorname{Min}(p_i) \to \operatorname{C}_A(p_i)$$

for each "predicate" set symbol p_i . Then we need not decide in advance which sets are to be minimized; if

⁶See pp. 157-158 of [van Dalen 1983] for details.

⁷Multiple-predicate circumscription is then just

 $[\]neg (x_1) \dots (x_n) [A(x_1, \dots, x_n) \& x_{m_1} \subset p_{m_1} \& \dots \& x_{m_k} \subset p_{m_k}].$

We are using \subset for \subseteq and \neq .

⁸However, the techniques in [Perlis 1985] would seem to allow this via a quotation device. We will employ this below.

it turns out that we can prove that a given set satisfies Min, then automatically circumscription will come into play. Also, this allows the choice of substitute predicate (set) x_i for p_i to be formally made⁹. (Actually, it would be more elegant to have instead something like

$$(\forall p)(\operatorname{Min}(p) \to C_A(p)).$$

However, then it is problematic to get the intended p to be selected out of A's list of arguments. This will be treated below.)

For example¹⁰, let $p_1 = white$ and $p_2 = heavy$, and let A be $w \in white \&h \in heavy$. Then if it should be proven that *heavy* is to be minimized, i.e., if *Min(heavy)* is a theorem, the appropriate circumscription of heavy objects will be done; similarly for *Min(white)*.

The second bonus is that also to some extent the non-monotonic character of circumscription can be made more explicit, in that the presence of the axiom A that provides the context for circumscription of p, can be formally manipulated as well. Then when new axioms are added, A will not shift meanings in a formally unspecified way. That is, instead of A we may use a set 'a' of (names of) axioms such that when new axioms are introduced, or when focus of attention is shifted, a new set (possibly a superset 'a+' of 'a') may be chosen for circumscription; but now the former apparent non-monotonicity simply becomes a theorem within the theory: for some sets a, a+, and p, where a \subset a+, Circum(p,a) will provably not be a subset of Circum(p,a+).

More precisely, consider a language having the usual symbols for set theory, and also a name (constant) 'c' for each expression c, and an unquotation function symbol \sim so that 'c' = c.¹¹ Then if 'a' is a set of wff names (axioms), 'e' a set-expression name (the property to be minimized, written as a term in set notation), and 'p' a set of set-constant names (the items allowed to vary in order to effect the minimization), we define Circum(a,e,p), *full set circumscription of e by a with respect to p*, to be

⁹And possibly even to be relaxed so that, for instance, cardinalities could allow the choice of a slightly larger than minimal interpretation of p_i if called for, or for that matter the least among various minimal interpretations.

¹⁰I am indebted to V. Lifschitz for suggesting this simple but illustrative example, which will be used again below.

¹¹For more on quotation and unquotation devices we refer the reader to [Perlis 1985]. Note that McCarthy [1986] also advocates the use of names in treating the *unique names hypothesis* of Reiter [1980].

 $\neg(\underline{f})((\forall b \in a) True(interp(f,b,p)) \& interp(f,e,p) \subset e).$

Here, interp is a function symbol with associated axioms guaranteeing that for suitable arguments interp(f,e,p) equals the expression resulting from prepending f to each occurrence in e of elements of p.

The idea behind this is that if e is minimal then there should not exist an interpretation f of the symbols in p that produces a smaller version interp(f,e,p) of e, and that still respects the axioms $b \in a$ when each is so interpreted (as interp(f,b,p)). Thus, interp(f,x,y) specifies a new reading of x, given by reading each symbol s of x that happens to be in the set y, as f(s) instead of s.

To modify the previous example, if $a=\{ w\in white', h\in heavy'\}, p=\{ white', heavy'\}, e='white\cup heavy', and g=\{<white, \{w\}>, <heavy, \{h\}>\}, then$

interp('g', 'w \in white', p)) = 'w \in g(white)'

interp('g', 'h \in heavy', p)) = 'h \in g(heavy)'

and

$$interp('g', 'white \cup heavy', p)) = 'g(white) \cup g(heavy)'$$

Then we use the function-definition of g (in terms of ordered pairs in the form <argument,value>), and find $g(white)=\{w\}, g(heavy)=\{h\}$. We get as a result

$$True(interp('g', 'w \in white', p)) \equiv True('w \in g(white)') \equiv w \in g(white) \equiv w \in \{w\}, w \in \{w\},$$

and

$$True(interp('g', h \in heavy', p)) \equiv True('h \in g(heavy)') \equiv h \in g(heavy) \equiv h \in \{h\},\$$

so that True(interp('g',b,p)) is provable for each $b \in a$. Then minimizing 'white \cup heavy' via Circum(a, 'white \cup heavy',p) produces

$$\neg$$
[interp('g', 'white \cup heavy',p) \subset 'white \cup heavy'],

i.e., $\{w\} \cup \{h\} \not\subset white \cup heavy$. But since $\{w\} \cup \{h\} \subseteq white \cup heavy$ it follows that $\{w\} \cup \{h\} = white \cup heavy$, i.e., anything other than w or h is neither white nor heavy. On the other hand, Circum(a, 'white',p) would produce that nothing other than w is white, and Circum(a, 'heavy',p) that nothing other than h is heavy.

Now we can write

 $\forall p(Min(a,e,p) \rightarrow Circum(a,e,p))$

solving the problem indicated earlier. That is, we do not actually have to minimize anything at the outset; we do not circumscribe, so much as provide a *definition* of circumscription, which then can be invoked when needed (when it is proven that a particular expression e is to be minimized). In fact, the formula defining full set circumscription of e by a with respect to p, given earlier, should be introduced precisely as a definition of the predicate symbol Circum, within the formalism. To some extent this already can be accomplished within McCarthy's formula circumscription, since that formalism involves a single second-order formula; however, this is so only separately for each choice of axiom A, whereas our approach incorporates any axiom selection in one formula. Moreover, now the axiom set 'a' can be infinite, whereas in predicate and formula circumscription A must be a single formula (or finite conjunction).

VI. An Objection

The reader may have noticed the following objection. We have complained about second-order circumscription on the basis of potential needs for still higher-order constructs. Yet our solution was to introduce in its place an essentially single-level ("flat") set theory that will be just as poor at representing cities of buildings of rooms of walls of bricks as is second-order logic. We mentioned this earlier, but did not discuss how to solve the problem.

Note that CST_0 generalizes trivially to handle such "tougher" cases, simply by allowing certain sets to be individuals. Thus, letting every set of blocks be an individual immediately leads to sets of sets of blocks (and hence to sequences of columns). This still is an *ad hoc* solution, however. On the other hand, all this can be done at the outset by starting with a theory such as ZF itself (as a subtheory of one's commonsense theory, relativized to the Ind predicate). Then fairly arbitrary collections are expressible. Now, is ZF a reasonable set theory for commonsense reasoning? Many of its axioms, such as Powers or Replacement or Infi nity, may seem unpalatable to some, and of little use in commonsense.

I claim that such powerful set theories are needed for commonsense, and that a set theory due to Ackermann [1956] serves to provide the right approach to this. However, this takes us a little off the main topic of circumscription, so we will leave it now with the observation that by employing still "fancier" set theories having universal sets (e.g., Gilmore's NaDSet [1986]) one can even have self-membered sets. Then *all* entities may be considered individuals, but with *Aussonderungs* altered to avoid paradox. This too I claim has significance for commonsense, but is treated in another paper [Perlis, to appear].

VII. Conclusions

Our main points are as follows:

- (1) Formula circumscription has two separate features, namely variable predicates and second-order wffs. These are largely unrelated to each other. Variable predicates address directly a circumscriptive issue, in that the very concept of minimization which circumscription is designed to capture also depends on variability of predicates other than the one minimized.
- (2) On the other hand, second-order formulas do not bear directly on the idea of circumscription except insofar as they extend the expressive power of the underlying language. Indeed, fi rst-order (variable) circumscription works as well as the expressive power of its underlying language. We should fi x the circumscriptive mechanism and enrich the fi rst-order language as needed without changing the former. Going to higher-order logics simply invites more problems, as Kueker's example hints.
- (3) The apparent gains of second-order logic result from its being effectively a (weak) set theory; so why not open the door wide? Set theory is already needed for commonsense reasoning. Indeed, more set theory is needed for commonsense than mere second-order logic.
- (4) There are very powerful and well-understood fi rst-order set theories, going well beyond the power of second-order predicate logic. A powerful fi rst-order set theory with fi rst-order (formula) circumscription gives at least all that second-order formula circumscription does. This includes allowing use of a single formula instead of a schema.

Acknowledgement

I would like to thank the following people for constructive criticism: Jennifer Drapkin, Vladimir Lifschitz, Michael Miller, Jack Minker, John Schlipf.

This research has been supported in part by the following organizations:

The Martin Marietta Corporation

U. S. Army Research Offi ce (DAAG29-85-K-0177)

Bibliography

- (1) Ackermann, W [1956] Zur Axiomatik der Mengenlehre. Math. Annalen, v. 131, 336-345.
- (2) Davis, M. [1980] The mathematics of non-monotonic reasoning. *Artificial Intelligence*, v. 13, 73-80.
- (3) Etherington, D., Mercer, R., and Reiter, R. [1985] On the adequacy of predicate circumscription for closed-world reasoning. *Computational Intelligence*, v. 1, 11-15.
- (4) Gilmore, P. [1986] Natural deduction based set theories. J. Symb. Logic., v. 51, 393-411.
- (5) Kueker, D. [1984] Another failure of completeness for circumscription. Photocopied notes, Week on Logic and Artifi cial Intelligence, Univ. of Maryland, Oct. 22-26, 1984.
- (6) Lifschitz, V. [1985] Computing circumscription. IJCAI-85, 121-127.
- (7) Lifschitz, V. [1986] On the satisfi ability of circumscription. *Artificial Intelligence*, v. 28, 17-28.
- (8) McCarthy, J. [1985] Acceptance Address, IJCAI Award for Research Excellence.
- (9) McCarthy, J. [1986] Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, v. 28, 89-118.
- (10) Perlis, D. [1985] Languages with self-reference I. Artificial Intelligence, v. 25, 301-322.
- (11) Perlis, D. [to appear] Commonsense set theory. In P. Maes and D. Nardi (eds.) *Meta-Architectures and Reflection*, North-Holland.
- (12) Perlis, D., and Minker, J. [1986] Completeness results for circumscription. *Artificial Intelligence*, v. 28, 29-42.
- (13) Reiter, R. [1980] Equality and domain closure in first order data bases. J. ACM, v. 27, 235-249.
- (14) van Dalen, D. [1983] Logic and structure, 2nd edition. Springer-Verlag, Berlin.