Philosophy 532 and Philosophy 536 were the two seminars I presented while on leave at the Princeton University Philosophy Department, Fall, 2002.

Harvey M. Friedman

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 1
9/25/02

"Mathematics is the only major subject which has been given a philosophically credible foundation."

This is widely accepted, inside and outside philosophy, but one can spend an entire career clarifying, justifying, and amplifying on this statement. Certainly a graduate student career.

This could entail clarifying what is meant by "mathematics".

Or "major subject". Or for that matter, "subject". Also "given". And of course "philosophically credible foundation", or even just "foundation".

We are almost down to the level of something manageable. In what sense do we have a "foundation" for mathematics? A "foundation" in the sense that we don't have for other subjects such as physics, statistics, biology, law, history, literary/music criticism or philosophy?

Before I get to the principal features of foundations of mathematics, let us first touch on "what is mathematics?"

Of course, we do not expect any kind of remotely definitive answer to such a question. But it is common to focus on just one aspect of this question. How does mathematics compare to the (other?) sciences in terms of its relationship to "reality"?

Let me focus this a bit. Nobody(?) would compare physics with, say, the study of "the causes of the U.S. Civil War". In one case, we are endeavoring to uncover "fundamental" facts about "external reality", which can be "tested" by repeated "objective" experiments. In the other case, we are doing something radically(?) different.

We start with the most extreme(?) view.

1. Mathematics is just like any other science in that it is about objects of "nature", that are not man made, but part of external reality, like physics or other physical sciences. As a consequence, every properly formed assertion about the "objects" of mathematics is true or false in the same sense that a properly formed assertion about physical reality is true or false. This is usually referred to as Platonism. This view was championed by Kurt Goedel.

Look at his "What is Cantor's continuum hypothesis?" and "Russell's mathematical logic".

Most people are uncomfortable with 1. For example, there remain specific mathematical questions of a rather fundamental character, around 100 years old, which continue to resist solution - and the nature of this resistance makes one feel uncomfortable with Platonism. Specifically, the continuum hypothesis:

CH: every set of real numbers can be mapped one-one into a set of natural numbers or onto the set of all real numbers.

Mathematicians have long since dodged this question by saying things like

"we don't care about every set of real numbers - just the relatively explicit sets of real numbers that come up in actual mathematics (the mathematics that we like) like Borel set of real numbers"

"if we restrict to, say, Borel sets of real numbers, and Borel maps, then CH is provable"

"CH is abnormal in its use of arbitrary sets of real numbers, and if you stay within reasonably explicit sets of real numbers, then you will be safe from logical difficulties, and you won't have to do any philosophical thinking"

But it is not reasonable(?) for a philosopher holding Platonism to dodge CH in this way.

What is most disconcerting is that it has been established that CH cannot be settled on the basis of the currently accepted axioms for mathematics (Kurt Goedel and Paul Cohen). In addition, the most favored new axioms for mathematics - the so called large cardinal axioms - are also known to be insufficient for settling CH.

It gets worse. The experimental method plays a completely crucial a primary role in the physical sciences, and reinforces Platonism in the context of the physical sciences. (However, there is now talk of some new physical theories that may in principle be impossible to verify experiment-ally! That will probably be very hard for mainstream physicists to swallow.)

But what takes on the role of "experimentation" or "experimental verification" or "experimental confirmation" in mathematics?

It is now clear with the use of computers, that there is a viable sense of "experimental verification", and more commonly, "experimental confirmation" in play now in mathematics.

However, this has appeared only in the context of very concrete mathematical problems. At present we have no idea how we can "experimentally verify" or even "experimentally confirm" anything even remotely like CH through the results of computation.

There have been attempts to use a much more liberal notion of "experimental confirmation". Maybe the CH allows us to solve other mathematical problems, or some group of mathematical problems, in a direction which somehow strikes us as much more plausibly true than false. This is certainly far weaker than what we have in the physical sciences, where results of experimentation are supposed to be unarguable(?). Despite some continuing efforts by some along these lines, in favor of not(CH), the overwhelming majority of mathematical logicians have not found this persuasive.

For these reasons (and undoubtedly others), most mathematical logicians feel much more comfortable with a hybrid view:

2. Views to the effect that Platonism is correct but only for certain relatively "concrete" mathematical "objects".

Other mathematical "objects" are man made, and are not part of an external reality.

This is a hybrid point of view and is probably the most common point of view among mathematical logicians -- to the extent that they are concerned with philosophical issues. Examples are "finitism" (William W. Tait) and "predicativism" (Solomon Feferman) and "ultrafinitism" (Edward Nelson).

(Assuming time permits today, I will come back to some of the principal isms. This is a substantial topic, and one can easily arrange an entire seminar on such isms).

There are many possible "isms", some of which have been put forth, and some of which appear to be implicit in the "views" of people who do not consider the explication of "views" as part of their professional activities.

We now come to the main tool that Platonists use for defense (attacking the opposition). They ask

"where do you draw the line?"

The point is that wherever you draw the line, there is a natural slightly higher place, and one has to defend why the stuff on one side of the line is OK whereas the stuff just barely on the other side of the line is not OK.

Let me give an example. I have seen some ultrafinitists go so far as to challenge the existence of 2^{100} as a natural number, in the sense of there being a series of "points" of that length. There is the obvious "draw the line" objection, asking where in

$$2^1, 2^2, 2^3, ..., 2^{100}$$

do we stop having "Platonistic reality"? Here this ... is totally innocent, in that it can be easily be replaced by 100 items (names) separated by commas.

I raised just this objection with the (extreme) ultrafinitist Yessenin Volpin during a lecture of his. He asked me to be more specific. I then proceeded to start with 2^1 and asked him whether this is "real" or something to that effect. He virtually immediately said yes. Then I asked about 2^2, and he again said yes, but with a perceptible delay. Then 2^3, and yes, but with more delay. This continued for a couple of more times, till it was obvious how he was handling this objection. Sure, he was prepared to always answer yes, but he was going to take 2^{100} times as long to answer yes to 2^{100} then he would to answering 2^1. There is no way that I could get very far with this.

To recapitulate, this was used as a defense against the defense of Platonists that asks for a place to draw the line between reality and convention(?) - at least in this one context or related contexts involving large finite objects. However, I have never seen this kind of defense used in infinitary contexts in connection with Platonism, but it may be applicable.

Incidentally, there is a formal line of investigation that is very relevant here, with some somewhat surprising outcomes.

I'll first give a formulation using a very weak set theory, T. In this set theory, all objects are sets, and we have membership and equality. There are only three axioms:

1. Two sets are equal if and only if they have the same elements.

2. There is a set with no elements.

3. If x,y are sets then there is a set whose elements are exactly the elements of x together with y itself.

How big a set can you prove exists in T? Obviously any size. E.g., each of Ø,{Ø},{{Ø}},... as long as you want.

However, this misses the point. Consider the one with 2^{100} pairs of braces. One cannot even give a humanly digestible description of this set in the language of T, let alone give a humanly digestible proof of its existence in T. I may be wrong here, but both claims seem likely to be true even if abbreviations are allowed. (I will see if I can prove this).

What we really want to know is if there is a "short" description of a set in T, with a "short" proof in T that it exists, but where the set is extraordinarily large in some way. We will simply ask that it be extraordinarily large in the usual sense.

The answer is yes. The result works for T, but the estimates involved are much nicer if we allow abbreviations in the underlying logic. Use of abbreviations is completely standard in virtually any kind of mathematics. We can make the following definition for $0 \le i \le 8$, in T, in an especially nice way if abbreviations are allowed. A set has "rank at most 0" if and only if it is empty. A set has "rank at most 1" if and only if all of its elements have "rank at most 0". Continue to: a set has "rank at most 8" if and only if all of its elements have "rank at most 7".

In fact, there is a much more direct way to define "rank at most 8" in T, even without any abbreviations. x has "rank at most 8" iff

there is no $y_1 \in y_2 \in y_3 \in y_4 \in y_5 \in y_6 \in y_7 \in y_8 \in y_9$ with $y_1,y_2,y_3,y_4,y_5,y_6,y_7,y_8,y_9 \in x$.

One can write out a proof in T, especially with abbreviations, in a mathematically friendly way, with friendly length and structure. Yet the number of elements in the set of sets of rank at most 8 is 2^2^2^2^65536.

In the case of arithmetic, one can prove related results, although I don't know how to prove a result as satisfying as the above. It is clear how to prove such a result if, say, addition and (base 2) exponentiation are primitives. However, in the above case of finite set theory, no additional primitive was needed to very simply define the huge set in question.

Can we prove such a result in arithmetic with only 0,1, and addition? The answer is no. The sizes do grow quite large, but not the kind of iterated exponential-ity that we see above.

However, it can be done with 0,1,+,x. The example of the huge number is perhaps not as fundamental as the huge set above, but reasonably natural. Use the relation

"every $1 \le i \le n$ divides m"

to bootstrap up starting at 3. To define the relevant 9 integers, we have to take the least m at each stage.

However, it is not obvious that we can prove the existence of such a least m with an appropriately short (or even friendly) proof. However, that becomes part of the result.

It would be interesting to do a thorough proof theoretic investigation of such systems. The only way to get such huge objects out of such small proofs and such weak axioms is to be using the cut rule over and over again.

Whoops! I don't know if you can reduce this to, e.g., a single cut, and at what cost. That is the kind of thing that I mean by a proof theoretic investigation.

Let us return now to "what is mathematics". We have had a brief discussion concerning Platonism and Platonism in its restricted forms.

But even if we settle on such ontological issues, there still remains a question of an entirely(?) different sort. It is clear that mathematicians do not even conceive of the idea of settling the truth value of all sentences about the mathematical objects that they are concerned with. They are only interested in certain sentences on a case by case basis.

Obviously, they cannot be expected to be interested in any specific sentences of inhuman size. However, sentences of inhuman size can be interesting to them if they are somehow incorporated in a sentence of human size.

For example, some mathematicians are quite interested in arbitrarily (first order) sentences about the field of real numbers. There are infinitely many of these, and therefore ones beyond any given finite size, yet all of them come under a celebrated theorem of Tarski with his decision procedure for determining the truth value of all of them. Here arbitrary such sentences get folded into a single overarching theorem.

The sociology of this is more complicated. Since the predicate calculus is too philosophical for comfort, the mathematicians know of Tarski's result in different notation.

So first of all, the mathematicians can only be (directly) interested in mathematical statements of quite small size, with abbreviations allowed. Of course, the number of such statements is absolutely enormous.

Secondly, of even the statements of quite small size (with abbreviations), only a tiny fraction are "mathematically interesting".

So any reasonably complete account of what mathematics is, or what mathematical activity is, must ultimately confront the issue of what mathematicians are trying to accomplish - at least if it is to be relevant to actual mathematical activity.

This is very difficult to get a hold of, especially in light of the fact that mathematicians are not in anything like full agreement as to what they are trying to accomplish.

What makes matters more difficult still, is that writing about "what is mathematics, and what are we trying to accomplish" is not considered normal professional activity among mathematicians. This statement is not so negative. After all, "what is philosophy, and what are we trying to accomplish" is rarely a topic of the leading philosopher's papers either. I gather that philosophers did not like it when Rorty wrote about his. I asked Kripke if he would write about this and related matters, and it he made it clear that he wouldn't touch it with a xxxxxxxx foot pole!

You do sometimes see such writings by mathematicians, but about fragments of mathematics, and aimed at very limited audiences, with very specific purposes in mind. For many reasons, including the way that fundamental issues are (or are not) handled, this proves not to be generally useful for philosophers trying to uncover "what is mathematics, and what are mathematicians trying to accomplish."

I do get the feeling that this situation is better in physics. This may be partly due to the apparent fact(?) that it is relatively clear from the outset "what physics is and what physicists are trying to accomplish" than "what mathematics is and what mathematicians are trying to accomplish".

Nevertheless I think that it is possible to find some overarching powerful motivating principles for mathematics, or at least several "brands" of mathematics that would be illuminating, especially to philosophers. Doing this well for even, say, mathematics before 1950, or even mathematics before 1900, would be a major achievement.

Coming back to more traditional philosophical concerns, let's go back to the hybrid view:

2. Views to the effect that Platonism is correct but only for certain relatively "concrete" mathematical "objects". Other mathematical "objects" are man made, and are not part of an external reality.

Under such a view, what is to be made of the part of mathematics that lies outside the scope of Platonism?

An obvious response is to reject it as utterly meaningless. Let us then call this an exclusionary view. (We will discuss more moderate views later). We have already discussed the "where do you draw the line" attack against the exclusionary view.

But another problem comes up for the exclusionary view.

What if it turns out that the excluded objects can be used to derive information about the blessed objects that cannot be derived just using the blessed objects?

Strictly speaking, we are confusing epistemology with ontology, and so let me rephrase this.

What if it turns out that accepted axioms associated with the excluded objects can be used to derive information about the blessed objects that cannot be derived just from the accepted axioms associated with the blessed objects?

There is just such a result of Kurt Goedel that applies in most such situations. You can call it "Goedel's principle".

GP: you can prove the consistency of accepted axioms about certain objects using accepted axioms about slightly more "potent" objects, but not with just accepted axioms about the original objects.

Here "potency" normally coincides with size, or cardinality in the sense of Cantor.

Incidentally, there are more subtle kinds of "potency" that are very relevant.

We have also seen size considerations play an important role in the finite. However, a full development of GP in this context needs to be done. I got involved in what I call finite Goedel's theorem, which is a decent start.

You can easily argue that GP is not a strong argument against the exclusionary view. Granted, something is to be gained by admitting more potent objects, but the gains may be ill gotten.

Perhaps the "accepted" axioms for the excluded objects are inconsistent (even short inconsistencies). Then they can prove anything. Incidentally, attempts to segregate inconsistent parts of theories from consistent parts of theories in an appropriate way have not been convincing).

NOTE: There may be a reasonably uniform way of passing from objects to "standard axioms about them" emerging out of current research. The idea is to consider all short sentences about them in primitive notation which are true. These may be susceptible to complete classification, and coincide in many cases with the axioms that are intuitively compelling. There may be more subtle considerations than "shortness". This is beginning to be carried out with set theory in primitive notation, and also in arithmetic. END

But the exclusionary view becomes really painful to hold if it turns out that the use of the excluded objects turns out to be so effective that lots of mathematics of the kind valued by mathematicians has to be thrown in the trash.

In fact, implicit in Goedel is the following conjecture:

GP*. you can prove lots of mathematically "interesting" or "beautiful" facts about certain objects using accepted axioms about more "potent" objects, but not with just accepted axioms about the original objects.

In foundations of mathematics, we are at a place where we are just beginning to assemble a variety of results supporting GP*. It is still very early, so it is not yet clear just how pervasive this phenomena is. Obviously this requires substantial feedback from the mathematical community for confirmation.

I gave a four lecture series called "Rademacher Lectures" last week at the U Penn Math Dept, on just this topic. That's why we didn't meet the first week. These lectures will be on the web shortly.

Look at the following interesting article:

Hilary Putnam, "Philosophy of mathematics: why nothing works." In: Words and Life, ed. Conant.

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 2
10/2/02

We now discuss some special features of the foundations of mathematics which illustrate its distinctive power as compared to attempts at the foundations of other major subjects. I will make a number of numbered points about f.o.m.

1. In f.o.m. we have a very simple ontology that suffices to express all mathematical concepts in a uniform manner. This simple ontology is that of set theory, with sets, membership between sets, and equality between sets.

2. There is usually an intuition behind most mathematical concepts that is not preserved when expressing them in terms of sets. For instance, nobody declares that natural numbers are really sets, or that real numbers are really sets.

3. However, no one seems to know what a natural number really is, or a real number really is, least of all mathematicians (smile). The Russellian idea that a natural number really is a special kind of class of sets is certainly not compelling either.

4. So under these circumstances, it is clear that the relationship between mathematical entities is what counts, and nothing else.

5. Attempts to found mathematics through relationships only, and not on actual objects, has been a persistent but elusive goal. Sooner or later, all attempts at such an autonomous foundation have failed, usually on these

grounds: At some point, objects of some sort have to be postulated, and one has to mirror the standard approach through set theory. This particularly applies to so called categorical "foundations" (through category theory, which is undoubtedly at least a useful organizational scheme for a considerable variety of mathematical contexts).

6. Coming back to the standard set theoretic f.o.m., since only the relationship between mathematical entities is what counts, we first need to judge standard f.o.m. in these terms. But before we discuss this, we need to give an overarching description of the nature of this standard f.o.m.

7. As stated earlier, in standard f.o.m., every mathematical object is a set. Two sets are equal if and only if they have the same elements. Thus the only thing that counts about a set is what's in it. Since everything is to be a set, all elements of sets are also sets.

8. The simplest of all sets is the set with no elements, called the empty set, $\emptyset$. We can take the set consisting of just the empty set, written $\{\emptyset\}$. We now have two sets, $\emptyset,\{\emptyset\}$, and so we then have the two sets $\{\{\emptyset\}\},\{\emptyset,\{\emptyset\}\}$, for a total of 4 sets. We can continue in this way, obtaining the so called hereditarily finite sets.

9. The above is supported by that very weak set theory we have encountered earlier, which we called T. This has the axioms of extensionality, emptyset, and the adjoining axiom, x □ {y}.

10. Finitary mathematical objects are identified with hereditarily finite sets. These include natural numbers, integers, rational numbers, finite sets of rational numbers, functions from finite sets of rational numbers into finite sets of rational numbers, Cartesian products of finitely many sets of rational numbers, polynomials of several variables with rational coefficients. But definitely not real numbers and complex numbers, which are infinitary in an essential way. With more subtlety, one also handles certain rather special real and complex numbers, particularly real algebraic numbers and complex algebraic numbers.

11. Set theory really gets going when we go further than the hereditarily finite sets. We can take the set of all

hereditarily finite sets. This is supported by the axiom of infinity, although the usual axiom of infinity is too weak to get this without some additional axioms.

12. The set of all hereditarily finite sets (usually called HF) together with its subsets, is more than enough to handle the most common infinitary objects in mathematics. These include real and complex numbers, and more generally, elements from a complete separable metric space. All of the usual arguments of "countable" mathematics can be easily done with HF and its subsets.

13. With some care, almost all of the statements of current mathematics can be appropriately viewed as "countable" mathematics; the exceptions consist of general formulations, where the generality of the formulations is something leading mathematicians stand ready to sacrifice.

14. But how is "having all subsets of HF" reflected in the axioms? This is by the crucial separation axiom, which says that the set of all elements of HF, or of any given set, obeying any given property (parameters allowed) presented as a formula in the language of set theory.

15. Countable set theory does not directly take care of, say, the set of all real numbers. Though for almost all practical purposes, one can just talk about real numbers without having the set of all such. The set of all real numbers and more generally, the completion of any countable metric space, is supported by another f.o.m. axiom, the axiom of power set. This asserts that the set of all subsets of any given set exists.

16. We now have a very powerful system, more or less what is called Zermelo set theory. This supports HF, S(HF), S(S(HF)), S(S(S(HF))), etc. Also, of course, their subsets. This is a very powerful set theoretic structure. But it does not include itself as a set. I.e., we can take the union of all of these sets just listed, to form a set that is bigger than anything so far. To support this union, we need the use the axiom of replacement.

17. There is one more major axiom missing, and that is the axiom of choice. This was controversial for many years, but now is considered standard. This says that if you have a set of nonempty sets, no two of which have any elements in common, then there is a set that has exactly one element in

common with each of them.

18. The full standard f.o.m. is called ZFC for Zermelo Frankel set theory with the axiom of choice. The axioms are:

1. Extensionality.
2. Pairing.
3. Union.
4. Separation.
5. Power set.
6. Choice.
7. Foundation.
8. Replacement.
9. Infinity.

All of 1-8 hold in HF. The addition of 9 is incredibly powerful.

19. How adequate is ZFC? It is the de facto standard for correctness, and will be used if a dispute arises in the mathematics community as to correctness that is not resolved informally. But I do not believe that it has ever had to be used for this purpose - at least for many decades. Disputes get resolved informally and semiformally well before ZFC needs to seriously enter the picture.

20. For nearly all mathematical purposes, ZFC is known to suffice. In fact, for nearly all mathematical purposes, very weak fragments of ZFC are known to suffice.

21. Detailed studies of principal fragments of ZFC (including ZFC itself) and what math can be done there and what math cannot be done there is perhaps the main line of investigation in modern f.o.m.

22. There are a variety of problems that we now know cannot be settled in ZFC. However, at least until very recently, there has been something "abnormal" about these examples. The abnormalities are sufficiently pronounced that mathematicians can somewhat defensibly segregate them from the kind of mathematics that they have most traditionally valued.

23. In particular, at least until recently, the examples have been visibly less "concrete" than "normal"

mathematics. We are now just beginning to see examples that are regarded as having a normal level of concreteness, and judged to be "beautiful" by some of the leading mathematicians in the world. "Beauty" is a special buzz word among mathematicians, which is more than enough to get the so described statement admitted to the realm of "real mathematics".

****

We now discuss Russell's paradox for sets and its usual ways out. This results in some important formal systems. We then discuss Russell's paradox for a number of concepts, and consider analogs of the usual ways out for these concepts.

Let us begin with the following formal system FCS (full comprehension for sets). The language is the binary relation symbol $\in$ only. The nonlogical axioms consist of

$$(\exists x)(\forall y)(y \in x \leftrightarrow A)$$

where x is not free in the formula A.

This system FCS is inconsistent, as can be seen by replacing A with $y \notin y$, thereby obtaining

$$(\exists x)(\forall y)(y \in x \leftrightarrow y \notin y).$$

Let x be such that

$$(\forall y)(y \in x \leftrightarrow y \notin y).$$

Then

$$x \in x \leftrightarrow x \notin x$$

which is a contradiction.

The refutation of the once seemingly innocent FCS is called Russell's paradox for sets.

One of the most natural ways out involves using two sorts of variables. Lower case letters for objects, and upper case letters for sets. The atomic formulas are of the form $x \in A$, where x is an object variable and A is a set variable. The appropriate comprehension axiom now reads

$$(\exists A)(\forall x)(x \in A \leftrightarrow \varphi)$$

where φ is any formula in which A is not free.

Of course, this is far too weak by itself to be a good way out since one cannot even prove that there exists more than one object.

Actually that last sentence makes no sense, since we don't have equality. So how do we formulate it? We cannot even prove that there exists two sets, with an object in one that is not in the other.

For any usual purpose for f.o.m., we would like some sort of axiom of infinity. However, there is no way to express anything like the axiom of infinity in this language. More precisely, every consistent extension of this two sorted comprehension axiom scheme has a finite model. This follows from work done on monadic second order theories (Buchi et al).

The road to appropriately powerful set theories of this kind is to introduce a third sort. Thus we will have objects, sets of objects, and sets of sets of objects. We will use superscripts 0,1,2, respectively. The atomic formulas are of the form $x \in y$, where the superscript on x is one less than the superscript on y.

The comprehension axiom scheme we are interested in reads

$$(\exists x^{i+1})(\forall y^{i})(y^{i} \in x^{i+1} \leftrightarrow \varphi)$$

where φ is a formula in the three sorted language in which x_{i+1} is not free. Here i = 0 or 1.

Of course this is still too weak, since it has a finite model.

But now we can formulate a reasonable form of the axiom of infinity. Informally, this asserts that there is a nonempty set of sets of objects in which every element is properly included in an element. We leave it as an exercise to formalize this in our three sorted language.

This gives us a reasonably powerful system which we will call IST(3) + INF, or

impredicative set theory with 3 types
plus the axiom of infinity.

QUESTION: Is this form of the axiom of infinity the "simplest" possible axiom consistent with our comprehension axiom scheme, that forces all models to be infinite?

IST(3) + INF is obviously interpretable with the system Z_3 of third order arithmetic that we will encounter in PHIL 536 in a few weeks. Z_3 also has three sorts, and the intended interpretation is that the objects of the first sort are exactly the natural numbers; of the second sort, the sets of natural numbers; of the third sort, the sets of sets of natural numbers. We also have 0,S,+,• on the first sort, with the usual quantifier free axioms, and with induction for all formulas in the language.

It may seem reasonable that Z_3 is interpretable in IST(3) + INF. However, we are currently checking a proof that this is false. In fact, we are checking a proof that IST(3) + INF is mutually interpretable with Z_2 and not Z_3 (just the part of Z_3 without sets of sets of natural numbers). This is a bit tricky, and I will discuss this further at a later date.

However, let us now strengthen the axiom of infinity in the following way. Let INF* assert that there is a nonempty set of sets of objects in which every element has a least proper superset among the elements.

We are also checking a proof that IST(3) + INF* is mutually interpretable with Z_3.

For each $n \geq 2$, it is clear how to define the n-sorted system IST(n). The result should be that for $n \geq 3$, IST(n) + INF is mutually interpretable with Z_{n-1}, and for $n \geq 3$, IST(n) + INF* is mutually interpretable with Z_n.

Let IST(<•) be the union of the IST(n), $n \geq 2$. Let Z(<•) be the union of the Z_n, $n \geq 2$. Then IST(<•) + INF, IST(<•) + INF*, Z(<•) are mutually interpretable. This is well known without my recent efforts.

IST(<•) + INF or INF* is a streamlined version of what is essentially due to Russell, although he wrote down a far more complicated formalism, dividing the system into a purely predicative part and an axiom of reducibility that allows the derivation of impredicative comprehension axiom schemes.

The formalization of mathematics using such many sorted systems as IST(<•) + INF* is cumbersome and unpleasant. This is why f.o.m. turned to axiomatic set theory.

There is a set theory which corresponds very closely to IST(<•) + INF. This is the so called Zermelo set theory, Z.

The language of Z is $\in$,=. The axioms are

1. Extensionality. If two sets have the same elements then they are equal.
2. Pairing. The set consisting of any two given sets exists.
3. Union. The set consisting of the elements of the elements of any given set exists.
4. Separation. The set consisting of all elements of any given set that satisfy any given condition expressible in $\in$,= exists.
5. Power set. The set consisting of all subsets of any given set exists.
6. Infinity. There is a set A with $\emptyset \in A$ and for all $x \in A$, $x \cup \{x\}$ exists and lies in A.

A remark about Separation. We chose to give this informally, but a slight ambiguity has arisen. The formal version is

$$(\exists y)(\forall z)(z \in y \leftrightarrow z \in x \wedge \varphi)$$

where y is not free in φ and φ is a formula in $\in$,=.

From the informal description above, it was perhaps not clear that parameters are allowed in φ representing arbitrary sets. So we are using conditions expressible with the aid of reference to any sets, regardless of how these sets (parameters) can be themselves defined. This is typical of formal systems for f.o.m. purposes, although parameterless separation and comprehension has come up in important f.o.m. contexts.

IST(<•) + INF* is interpretable in Z. However, Z is not interpretable in IST(<•) + INF*, and in fact Z proves the consistency of IST(<•) + INF*. However, Z is not much stronger than IST(<•) + INF*. To indicate this, let us consider a weakening of Z called BZ (bounded Zermelo).

We call a formula of set theory (∈,=) bounded if and only if all of its quantifiers are bounded in the following sense.

A bounded quantifier is usually written in one of two ways:

$$(\forall x \in y)$$
$$(\exists x \in y).$$

These are of course meant to be abbreviations. I.e.,

$(\forall x \in y)(\varphi)$ abbreviates $(\forall x)(x \in y \rightarrow \varphi)$
$(\exists x \in y)(\varphi)$ abbreviates $(\exists x)(x \in y \wedge \varphi)$.

In BZ, we use exactly the same axioms, except that Separation is weakened to require that the formula used be bounded.

BZ, it turns out, is finitely axiomatizable, as opposed to Z or IST(<•) + INF (INF*).

IST(<•) + INF* is interpretable in BZ, but BZ is not interpretable in IST(<•) + INF*. The later is clear because otherwise, BZ would be interpretable in IST(n) + INF* for some n, which contradicts that BZ proves the consistency of IST(n) + INF*.

Here are the results I am getting at. Firstly, BZ, IST(<•) + INF*, IST(<•) + INF* are equiconsistent in that a weak fragment of arithmetic suffices to prove that the consistencies of these three individual systems are equivalent.

Secondly, BZ and BZ, (IST(<•) + INF*)', (IST(<•) + INF*)' are mutually interpretable, where ' is as discussed in PHIL 536.

Thirdly, there is a natural interpretation of $Z^{<\bullet}$ (the union of the Z_n) into each of these three systems. The result is that if φ is a sentence in the language of $Z^{<\bullet}$, and T is any of the three systems, then φ is provable in $Z^{<\bullet}$ iff the interpretation of φ is provable in T.

This result certainly does not hold of Z.

The extension of Z to the current gold standard foundation for mathematics, ZFC, arose out of the glaring incompleteness of Z.

Look at the series □, S(□), S(S(□)), etc., where S is the power set operation, and □ is the least set obeying the condition in the axiom of infinity. One cannot obtain the union of these sets in Z. To remedy this, the axiom of Replacement was introduced:

7. Replacement. Suppose that an operation assigning a unique set to any element of a given set is expressed in □,=. Then the set of all sets so assigned exists.

Another glaring incompleteness is more conceptual. We will come to this after we have discussed the so called ***cumulative hierachy of sets***.

The idea behind IST(<•) is that the sets are arranged in an organized linearly ordered hierarchy, where every set consists entirely of lower sets.

This conceptual picture is easy to maintain in connection with Z. In fact, one gets a cumulative hierarchy over □, indexed by the natural numbers. First the elements of □, then the sets of those, then the sets of those, etc. Since □ is a transitive set (every element of an element is an element), this cumulates. I.e., every set appearing at any given stage appears at all subsequent stages.

For conceptual homogeneity, it has been standard to not use □ as the starting point, but rather ∅. For infinitely many stages, one has only finitely many sets at each stage. I.e., we are looking at

$$\emptyset,\ S(\emptyset),\ S(S(\emptyset)),\ \text{etc.}$$

Note that they grow very fast, and in fact the sizes successively exponentiate to base 2.

At the end of these first infinitely many stages, one takes the union of the sets in this hierarchy, the union being called HF = the set of all hereditarily finite sets.

NOTE: HF does not stand for "Harvey Friedman", as much as I would like it to (smile).

However, Replacement is needed to construct this hierarchy

$$\emptyset,\ S(\emptyset),\ S(S(\emptyset)),\ \text{etc}$$

and prove that HF exists. A good alternative is to sharpen the axiom of infinity to assert the existence of a nonempty set containing $\emptyset$, and closed under the operation that sends x,y to $x \cup \{y\}$.

Coming back to the standard cumulative hierarchy, we first begin gently with $V(0) = \emptyset$, $V(n+1) = S(V_n)$, indexed along the natural numbers. It is usual to define $V(\omega) = HF$ to be the union of these $V(n)$. Then we continue with $V(\omega+1) = S(V(\omega))$, $V(\omega+n+1) = S(V(\omega+n))$, $n < \omega$.

The union of the $V(\omega+n)$ is written $V(\omega+\omega)$.

The construction of $V(\omega+\omega)$ cannot be done in Z, even if we were to sharpen the axiom of infinity in the way indicated above so that we would be able to get $V(\omega) = HF$.

$V(\omega+\omega)$, or more precisely $(V(\omega+\omega),\in)$, forms what is generally called the standard model of Z.

The full cumulative hierarchy of sets is defined in terms of the notion of ordinal in set theory. These are the epsilon connected transitive sets. The hierarchy is given by $V(0) = \emptyset$, $V(\alpha+1) = S(V(\alpha))$, $V(\lambda) = \cup_{\alpha<\lambda}V(\alpha)$. Here $0 = \emptyset$, $\alpha+1 = \alpha \cup \{\alpha\}$, and λ is a limit ordinal.

There is nothing present in the axioms of Z = Zermelo set theory that bears on the question of whether all sets appear in the cumulative hierarchy of sets.

The intention behind the axiom of Foundation is that using it, one can prove that every set lies in (some stage of) the cumulative hierarchy.

8. Foundation. Every nonempty set has an element no element of which is in the given nonempty set. I.e., every nonempty set has an epsilon minimal element.

Thus we now have what is called ZF:

1. Extensionality.
2. Pairing.

3. Union.
4. Separation.
5. Power set.
6. Infinity.
7. Replacement.
8. Foundation.

The final glaring(?) incompleteness is the axiom of choice. The axiom of choice is very useful for stating mathematical theorems in convenient full generality. It is known that any suitably concrete mathematical theorem that can be proved with the axiom of choice can be proved without the axiom of choice. Nevertheless, for the purposes of creating a reasonably "conceptually complete picture" of the cumulative hierarchy of sets, we definitely need to have the axiom of choice.

An alternative would be to construct a conceptually attractive and "reasonably complete" picture in which the axiom of choice is contradicted. However, we are very far from being able to do this.

9. Choice. For every set of pariwise disjoint nonempty sets, there is a set which contains exactly one element from each.

We have arrived at ZFC, the gold standard:

1. Extensionality.
2. Pairing.
3. Union.
4. Separation.
5. Power set.
6. Infinity.
7. Replacement.
8. Foundation.
9. Choice.

It seems like ZFC ought to be complete in some tangible sense. It has stood the test of time, having emerged in clear form by 1920. It provides a lot of comfort. Not only are mathematicians generally comfortable with what it says, and therefore its presumed consistency, they are also generally comfortable with its role as the de facto gold standard. They like to keep it in the background as an ultimate arbiter, but they rarely need to pull it out for this purpose. I have heard recently, however, that some people have run into trouble using very abstract category theory which they never bothered to see how to formalize in ZFC (or minor extensions of ZFC). They used what amounts to inconsistent theories, and their results had to be retracted. (Serves them right! (smile)).

It is also true that mathematicians now don't think much about ZFC, and very few can recall the axioms. This is partly from its great success, as it lies in the background. But another aspect of this is that in normal mathematics, one uses only a very tiny fragment of ZFC. Transforming this situation has been one of my major goals.

Coming back to the feeling of "conceptual completeness" of ZFC, there is a result that makes it perhaps all the more difficult to get at this.

We found a surprisingly simple statement in primitive notation ($\in$,=) which cannot be decided in ZFC.

For those of you who are familiar with class theory, the statement in class theory is a bit simpler in that it saves a quantifier.

PROPOSITION. Every transitive proper class has a four element chain under inclusion. I.e., there are distinct elements x,y,z,w such that $x \subset y \subset z \subset w$.

This is independent of even the Morse Kelley theory of classes with the global axiom of choice (for those of you who know what I am talking about).

Here is the set theoretic version.

PROPOSITION. The transitive sets that do not have a four element chain under inclusion form a set.

This is neither provable nor refutable in ZFC. It uses 7 quantifiers (in $\in$,=), but there is a way to modify it so as to use only 5 quantifiers. We also know that every 3 quantifier sentence in $\in$,= can be proved or refuted in ZFC. What about 4? That seems to be a difficult open question.

Nevertheless, I am still optimistic about finding a result of the following kind for systems like Z, ZC, and ZFC:

"The suitably simple set theoretic statements (including schemes) in primitive notation that are not obviously inconsistent with the set theoretic framework form an axiomatization of ZFC."

PHILOSOPHY 532
PHILSOOPHCIAL PROBLEMS IN LOGIC
LECTURE 3
10/9/02

We have discussed Russell's paradox for sets, and the impredicative theory of types as a first way out. This was replaced for purposes of the foundations of mathematics, by the closely associated Zermelo set theory (Z). Then Z, in turn, is then "completed" (in a sense we don't yet understand) by the gold standard Zermelo Frankel set theory with the axiom of choice, ZFC.

We now want to discuss several other ways out (of Russell's paradox for sets), some of which are far more powerful than ZFC in terms of interpretation power. There is an on-going search for new ways out, based on new unifying principles, which may have compelling philosophical stories that are superior to the heterogenous and specialized stories that have been given for ZFC (and the so called large cardinal axioms).

There is a constant adjustment and modification of these new ways out according to possible philosophical stories, as well as interpretation power. One ultimately wants (and will undoubtedly get) massive interpretation power sufficient to interpret at least the largest of the large cardinal axioms studied by set theorists, together with intriguing and compel-ling philosophical stories that will have counterparts throughout analytic philosophy. One should be able to get a range of interpretation powers, with a corresponding range of philosophical stories, of ever increasing boldness.

Here are the ways out that we will discuss.

1. Two sorted theory of classes with sets and classes as separate sorts.
2. Single sorted set/class theory.

3. Small/large set distinction, leading to a range of new issues.
4. Two set theoretic universes, leading to the ZFC level.
5. Three and more set theoretic universes.

1. Class/set theory.

The standard theory of this kind is NBG, after von Neumann, Bernays, Gödel.

There are two sorts of variables, lower case over sets, and upper case over classes. The atomic formulas are of the form

$$x = y$$
$$x \in y$$
$$x \in A.$$

Note that we can't even ask the question of whether a given class is an element of a given set, or even whether two classes are equal. However, one can, if one wishes, make perfectly good definitions of these concepts. For the latter, just that they have the same set elements. For the former, that there is a set element of the given set that has the same set elements as the given class.

We first give the axioms of NBG so that they most closely resemble those of ZF. We then give a well known simplification of them.

1. Extensionality. If two sets have the same elements then they are equal.
2. Pairing. The set consisting of any two given sets exists.
3. Union. The set consisting of the elements of the elements of any given set exists.
4. Separation. The set consisting of all elements of any given set that satisfy any given condition expressible in our two sorted language without class quantifiers, exists.
5. Power set. The set consisting of all subsets of any given set exists.
6. Infinity. There is a set z with $\emptyset \in z$ and for all $x \in z$, $x \cup \{x\}$ exists and lies in z.
7. Replacement. Suppose that an operation assigning a

unique set to any element of a given set is expressed in our two sorted language without class quantifiers. Then the set of all sets so assigned exists.
8. Foundation. Every nonempty set has an element no element of which is in the given nonempty set. I.e., every nonempty set has an epsilon minimal element.
9. Class comprehension (without class quantifiers). The class of all sets satisfying any condition expressible in our two sorted language without class quantifiers, exists.

In light of 9, we can simplify the remaining axioms of NBG as follows.

1. Extensionality.
2. Pairing.
3. Union.
4. Separation. The common elements of any given class and any given set forms a set.
5. Power set.
6. Infinity.
7. Replacement. The values of any univalent class of ordered pairs with arguments from a given set forms a set.
8. Foundation.
9. Class comprehension (without class quantifiers).

Note that this formulation has infinitely many axioms because of the scheme 9. However, it is known how to replace 9 with finitely instances, thereby giving a finite axiomatization of NBG. However, this has not been done in any elegant way.

Here we can utilize one of our main themes. We can show that we can use all instances of 9 with at most one quantifier and at most a few set variables and at most a few class variables. We have not gone into the details of this. It may be quite difficult to see what minimum sizes suffice.

We now come to the issue of the axiom of choice. First of all, there is the ordinary Set Choice, or AxC:

10. Choice (AxC). For any given set of pairwise disjoint nonempty sets, there is a set which contains exactly one from each of the nonempty sets.

However, there is the obvious version for classes, called Class choice, or Global choice (GC):

10'. Global choice (GC). For any given class of pairwise disjoint nonempty sets, there is a class which contains exactly one from each of the nonempty sets.

GC seems as powerful in class theory as Set choice is in set theory. Let me amplify on this.

It is universally believed that AxC is the strongest "choice principle" one can formulate in set theory.

However, nobody knows what this last sentence means! There should be some general criterion on what a choice principle is, with a proof that AxC implies any such choice principle. This has not yet been done.

Evidence of the power of the AxC is that it suffices, within even Z, to prove the set well ordering principle

every set can be well ordered.

It is also true that GC proves the class well ordering principle

the class of all sets
can be well ordered.

It would seem that this is the strongest choice principle in class theory. However this is not true! There is a choice principle that cannot be derived from GC.

For instance, the choice principle

$$(\forall A)(\exists B)(\varphi(A,B)) \rightarrow (\exists A_1,A_2,...)(\text{each } \varphi(A_n,A_{n+1}) \text{ holds}),$$

or if one wants to avoid use of the natural numbers,

$$(\forall x)(\exists A)(\varphi(x,A)) \rightarrow (\exists B)(\forall x)(\varphi(x,B_x)).$$

Presumably these choice principles cannot be proved in NBG + GC, even if we require that φ have no class quantifiers.

The reason that NBG has become so standard, despite its glaring incompletenesses discussed below, is its relationship to ZF. Also the relationship between NBG + GC with ZFC.

THEOREM. NBG and ZF prove the same sentences without class variables. ZF and NBG are equiconsistent. NBG is not interpretable in ZF. NBG and ZF' are mutually interpretable.

Recall that NBG is finitely axiomatizable, and ZF is not finitely axiomatizable. That is what is behind why NBG is not interpretable in ZF.

Here equiconsistency is taken to mean that

$$\text{Con(ZF)} \leftrightarrow \text{Con(NBG)}$$

is provable in a weak fragment of Peano arithmetic (PA). In fact, the fragment associated with superexponentiation suffices.

Note that ZF' is the result of the general prime operator on theories that we have discussed in PHIL 536.

THEOREM. NBGC = NB + GC and ZFC prove the same sentences without class variables. ZFC is a subsystem of NBG + GC. ZF and NBG are equiconsistent. NBG + GC is not interpretable in ZFC. In fact, even NBG is not interpretable in ZFC. NBG + GC and ZFC' are mutually interpretable.

Except for the first claim, this second theorem follows from the first theorem since ZF and ZFC are mutually interpretable and equiconsis-tent, and NBG and NBG + GC are mutually interpretable and equiconsistent.

We now come to the glaring incompleteness of NBG. This is of course the restriction that there be no class quantifiers in class comprehension.

If we remedy this, we get the system known as MK for the Morse Kelley theory of classes:

1. Extensionality.
2. Pairing.
3. Union.
4. Separation.
5. Power set.
6. Infinity.
7. Replacement.
8. Foundation.
9'. Class comprehension. The class of all sets satisfying

any condition expressible in our two sorted language, exists.

MKGC is a "reasonably complete" theory of classes, whatever that means. It is not "as complete" as ZFC, because of the fact that GC does not derive some choice principles mentioned above.

2. Single sorted set/class theory.

This is simpler than the theory of classes with sets and classes in that it is single sorted. It is also quite convenient.

There are two common ways of doing this. The first uses the language of ordinary set theory with $\in$,=,M, where M is a unary predicate symbol.

M(x) is defined to mean that x is an element of some set. These are the sets that correspond to the sets in class theory. The x such that not M(x) correspond to the classes in class theory that do not have the same elements as any set.

The second way of doing this is essentially the same. We use $\in$,= only. We define M(x) to mean "x is an element of a set", and repeat the same axioms as we use for the first way. The idea is that they are to be officially expanded into primitive notation.

We leave it to you to actually write down the theories under the two approaches, and to formulate the relationships between these theories and NBG, MK.

3. Small/large set theory.

Here we have a single sorted theory with $\in$,=, and the unary predicate symbols SM(x) and LA(x), for "x is small" and "x is large". Every set is either small or large, but not both.

As opposed to the previous section, we will have large sets that are elements of sets. In fact, we will have large sets that are elements of small sets - namely their singleton.

The idea is very obvious. Smallness is a matter of size, and so it is perfectly natural for a large set to be an element of a small set, or a large set.

We want to give a preferred model(s) for this conception. To do this, let us back up and give the corresponding prefer-red model(s) for ZFC and for MKGC (Morse Kelly with the global axiom of choice).

Let □ be a strongly inaccessible cardinal. This is a von Neumann cardinal □ such that

i) for all □ < □, $|2^{\square}| < \square$;
ii) □ > □;
iii) any subset of □ of cardinality < □ is bounded above.

Then (V(□),□) is a model of ZFC. Also (V(□),□,V(□+1)) is a model of MKGC.

The metatheory in which this discussion is carried out can be taken to be ZFC + "there exists a strongly inaccessible cardinal", or even ZC + "there exists a strongly inaccessible cardinal".

The preferred model(s) for this conception is as follows.

The sets consist of all sets that are included in a transitive set of cardinality □ □. The small sets are such sets of cardinality < □. The large sets are such sets of cardinality □.

The natural axioms that hold in this structure are as follows.

1. Every set is small or large, but not both.
2. Every subset of a small set is small.
3. Extensionality. If two sets have the same elements then they are equal.
4. Pairing. The set consisting of any two given sets exists, and is small.
5. Union. The set consisting of the elements of the elements of any given set exists.
6. Union'. The set consisting of the elements of the small elements of any given small set exists and is small.
7. Large set. The set consisting of all sets that are a subset of a small transitive set is a large set.
8. Separation. The set consisting of all elements of any given set that satisfy any given condition expressible in

our language, exists.
9. Power set. The set consisting of all small subsets of any given set exists.
10. Power set'. The set consisting of all subsets of any small set exists and is small.
11. Infinity. There is a small set z with $\emptyset \in z$ and for all $x \in z$, $x \cup \{x\}$ exists and lies in z.

12. Replacement. Suppose that an operation assigning a unique set to any element of a given set is expressed in our language. Then the set of all sets so assigned exists.
13. Replacement'. Suppose that an operation assigning a unique set to any element of a given small set is expressed in our language. Then the set of all sets so assigned exists and is a small set.
14. Foundation. Every nonempty set has an element no element of which is in the given nonempty set. I.e., every nonempty set has an epsilon minimal element.
15. Choice. For any given set of pairwise disjoint nonempty sets, there is a set which contains exactly one from each of the nonempty sets.
16. Choice'. For any given small set of pairwise disjoint nonempty sets, there is a small set which contains exactly one from each of the nonempty sets.

We can interpret MKGC in this system as follows. The sets in MKGC are the sets that are subsets of small transitive sets. The classes in MKGC are the sets whose elements are not subsets of small transitive sets.

We can interpret this system in MKGC as follows. The sets in this system are identified with the extensional well founded relations on the class of all sets in MKGC. The small sets are those whose domain is a set in MKGC.

There are a number of tricky issues concerning sharp conservative extension results between set and class theories, versus the above system and its variants.

This axiomatization is an example of where a lot of axioms are written down about a clearly conceived structure, and where one can easily omit important axioms by accident. After intensive study, one generally gets to a finished product that has some sort of feel of "completeness" about it.

In every such case, one would like to know just what kind of "completeness" one has, or what kind of "completeness" one can at least establish. All of the main formal systems emanating out of f.o.m. should be subject to such an investigation.

4. Two set theoretic universes.

We now discuss a system with two set theoretic universes, and two very simple axiom schemes. This system is equiconsistent with ZFC. ZFC is interpretable in the following system.

We consider the following system S in $L(\in,W)$. Here W is a constant symbol. The system is related to what is called Ackermann set theory, but it is simpler, and also it lacks extensionality. We are trying to be minimalistic.

1. SS. $x \in W \rightarrow (\exists y \in W)(\forall z)(z \in y \leftrightarrow (z \in x \& \varphi))$, where φ is a formula in $L(\in,W)$ in which x is not free.

2. WIT. $x_1,\ldots,x_k \in W \rightarrow ((\exists y)(\varphi) \rightarrow (\exists y \in W)(\varphi))$, where $k \geq 0$ and φ is a formula in $L(\in)$ with at most the free variables $x_1,\ldots,x_k,y$.

SS stands for "subworld separation". We think of W as a subworld. WIT stands for "witness". I used to write "RED" for reducibility, in vague connection with Russell's axiom of reducibility, but I now prefer WIT.

We can derive that W is a transitive set from these axioms. The significance of this is that every set in the first world; i.e., W, does not pick up elements from the second, "outer", world.

We first considered the above system with the axiom

3. EXT. $(\forall z)(z \in x \leftrightarrow z \in y) \rightarrow (x \in z \leftrightarrow y \in z)$.

We first derived in K(W) all axioms of ZFC except AxC and foundation, and derived a sharp form of replacement called reflection. NOTE: in the presence of foundation, replacement and reflection are equivalent, but not so without foundation. These derivations are pretty straightforward and fun! This was enough to conclude that ZFC is interpretable in SS + WIT + EXT.

We then showed how to inter-pret SS + WIT + EXT in SS +

WIT. This shows that ZFC is interpretable in SS + WIT.

5. Three and more set theoretic universes.

Having three or more set theoretic universes, with an indistinguishability principle gives some significant large cardinals (small large cardinals).

We present the system $K(W_1,W_2)$. The language is $L(\in,W_1,W_2)$, whose nonlogical symbols are $\in,W_1,W_2$, where W_1,W_2 are constant symbols.

1. $W_1 \in W_2$.
2. Subworld separation. $x \in W_1 \rightarrow (\exists y \in W_1)(\forall z)(z \in y \leftrightarrow (z \in x \wedge \varphi))$, where φ is a formula in $L(\in,W_1,W_2)$ in which y is not free.
3. Resemblance. $x_1,...,x_k \in W_1 \rightarrow (\varphi \leftrightarrow \varphi[W_1/W_2])$, where $k \geq 0$ and φ is a formula in $L(\in,W_1)$ with at most the free variables $x_1,...,x_k$.
4. Extensionality. $(\forall z)(z \in x \leftrightarrow z \in y) \rightarrow (x \in z \leftrightarrow y \in z)$.

Obviously we can derive 2 with W_1 replaced by W_2.

$K(W_1,W_2)$ is very considerably stronger than ZFC and K(W). A cardinal κ is extremely indescribable iff for all $R \subseteq V(\kappa)$ and first order sentence φ, if $(V(\kappa+\kappa),\in,R)$ satisfies φ, then $\exists \lambda < \kappa$ such that $(V(\lambda+\lambda),\in,R \cap V(\lambda))$ satisfies φ.

THEOREM. ZFC + "there exists an extremely indescribable cardinal" is interpretable in $K(W_1,W_2)$.

A cardinal κ is subtle iff for all sets $A_\alpha \subseteq \alpha$, $\alpha < \kappa$, and closed unbounded $C \subseteq \kappa$, there exists $\alpha < \beta$ from C such that $A_\alpha = A_\beta \cap \alpha$.

THEOREM. $K(W_1,W_2)$ is interpretable in ZFC + "there exists a subtle cardinal".

We showed that these results still hold if we cut $K(W_1,W_2)$ down considerably as follows.

Let $T(W_1,W_2)$ be the following system, in the same language $L(\in,W_1,W_2)$.

1. Parameterless comprehension. $(\exists x)(x \in W_2 \wedge (\forall y)(y \in x \leftrightarrow (y \in W_1 \ \& \ \varphi)))$, where φ is a formula in $L(\in,W_1)$ with at most

the free variable y.
2. Resemblance. x □ W_1 □ (□ □ □[W_1/W_2]), where □ is a formula in L(□,W_1) with at most the free variable x_1.
3. Extensionality. (□z)(z □ x □ z □ y) □ (x □ z □ y □ z).

The importance of comprehension without parameters is that it makes sense for what we call pure predication. Thus instead of thinking of sets, we can instead think of pure predicates, which do not allow reference to particular objects. General predication is what most axiom systems are based on, as well as mathematics itself. Reference is allowed to any finite list of objects, regardless of whether these objects can be "defined" in any way.

We never got a chance to see if we can remove extensionality. It is extremely likely that we can, in a way that is similar to our removal of extensionality from K(W).

Going further, we present the system K(W_1,W_2,W_3). The language is L(□,W_1,W_2,W_3), whose nonlogical symbols are □,W_1,W_2,W_3, where W_1,W_2,W_3 are constant symbols.

1. W_1 □ W_2 □ W_2 □ W_3.
2. Subworld separation. x □ W1 □ (□y □ W_1)(□z)(z □ y □ (z □ x □ □)), where □ is a formula in L(□,W_1,W_2,W_3) in which y is not free.
3. Resemblance. x_1,...,x_k □ W_1 □ (□ □ □[W_1,W_2/W_2,W_3]), where k ≥ 0 and □ is a formula in L(□,W_1,W_2) with at most the free variables x_1,...,x_k.
4. Extensionality. (□z)(z □ x □ z □ y) □ (x □ z □ y □ z).

THEOREM. ZFC + "there exists a subtle cardinal" is interpretable in K(W1,W2,W3), which is interpretable in ZFC + "there exists a 2-subtle cardinal".

We can define the systems K(W_1,...,W_n), n ≥ 2, as follows.

1. W_1 □ W_2 □ ... □ W_{n-1} □ W_n.
2. Subworld separation. x □ W_1 □ (□y □ W_1)(□z)(z □ y □ (z □ x □ □)), where □ is a formula in L(□,W_1,...,W_n) in which y is not free.
3. Resemblance. x_1,...,x_k □ W_1 □ (□ □ □[W_1,...,W_{n-1}/W_2,..., W_n]), where k ≥ 0 and □ is a formula in L(□,W_1,...,W_{n-1}) with at most the free variables x_1,...,x_k.

4. Extensionality. $(\forall z)(z \in x \leftrightarrow z \in y) \rightarrow (x \in z \leftrightarrow y \in z)$.

This would be trapped in interpretation power between the existence of an (n-2)-subtle cardinal and an (n-1)-subtle cardinal.

Presumably, again extensionality can be dropped. Also we should be able to work with a version based on parameterless comprehension.

PHILOSOPHY 532
PHILSOOPHCIAL PROBLEMS IN LOGIC
LECTURE 4
10/16/02
10/20/02

NOTE: The lecture on 10/16/02 is entirely superceded by the lecture on 10/23/02. Much progress has taken place. So these notes address a topic that was not actually covered on 10/16/02.

1. The comparable element principle.

This work is contained in a finished manuscript "A Way Out" on the preprint server http://www.mathpreprints.com.

The comparable element principle came about as a way out of Russell's paradox for sets with an unexpectedly simple escape clause.

We can think of Russell's paradox for sets in the following way.

Informally, the full comprehension axiom scheme in the language $L(\in)$ with only the binary relation symbol $\in$ and no equality, is, in the context of set theory,

Every virtual set forms a set.

We use the term "virtual set" to mean a recipe that is meant to be a set, but may be a "fake set" in the sense that it does not form a set. The recipes considered here are of the form $\{x: \varphi\}$, where φ is any formula in $L(\in)$.

We say that $\{x: \varphi\}$ forms a set iff there is a set whose elements are exactly the y such that □. Here y must not be free in □ (and must be different from x). Thus {x: □} forms a set is expressed by

(□y)(□x)(x □ y □ □).

Russell showed that

{x: x □ x} forma a set

leads to a contradiction in pure logic.

This way put of Russell's Paradox via the comparable element principle is to modify the inconsistent Fregean scheme in this way:

Every virtual set forms a set, or ____.

We refer to what comes after "or" as the "escape clause".

The escape clause that we use involves only the extension of the virtual set and not its presentataion.

We are now ready to present the comprehension axiom scheme.
\
NEWCOMP. Every virtual set forms a set, or, outside any given set, has two inequivalent elements, where all elements of the virtual set belonging to the first belong to the second.

To avoid any possible ambiguity, we make the following comments.

1. For Newcomp, we use only the language L(□), which does not have equality.

2. Here "inequivalent" means "not having the same elements".
3. The escape clause asserts that for any set y, there are two unequal sets z,w in the extension of the virtual set, neither in y, such that every element of z in the extension of the virtual sets is also an element of w.

THEOREM. Newcomp is provable in ZFC + "there exists arbitrarily large subtle cardinals". Newcomp is interpretable in ZC + "there exists a subtle cardinal". A shade less than ZFC "there exists a subtle cardinal" is interpretable in NEWCOMP.

PHILOSOPHY 532
PHILSOOPHCIAL PROBLEMS IN LOGIC
LECTURE 5
10/23/02
10/25/02

NOTE: THIS IS AN INCOMPLETE DRAFT.

We begin with the impredicative theory of types with the axiom of choice, which accommodates a unary predicate R whose extension forms a proper subuniverse. We augment this theory by various relativization principles which assert (schematically) that every sentence not mentioning R is equivalent to the result of relativizing only certain specified quantifiers to R. Augmentation by two such (categories of) relativization principles of particular simplicity results in a system which is mutually interpretable with ZFC together with certain well studied large cardinal axioms that are substantially stronger than the existence of measurable cardinals. This particular augmentation is given a reaxiomatization in more familiar terms (elementary substructure and restricted completeness). A substantive step towards a general theory of relativization principles in this context is offered in the form of an analysis of all relativization principles of a certain form. Strong conjectures are made about relativization principles in general.

1. Type theory with choice.

In the intended interpretation of type theory, we have objects of type 0, classes of type 1 consisting of elements of type 0, classes of type 2 consisting of elements of type 1, etc. We can also replace "classes" by "predicates" or "properties".

The language of type theory, L(TT), is a many sorted predicate calculus with a sort for each $n \geq 0$. We use variables x_i^n, $i,n \geq 0$, of type n. These variables range over objects of type n.

It will be convenient to write t(x) for the type of the variable x of L(TT).

The atomic formulas are the expressions x □ y, where t(y) = t(x)+1.

Formulas are built up from atomic formulas in the usual way using the connectives □,□, ,□ ,□ , and the quantifiers □,□.

There are standard logically complete axioms and rules for first order predicate calculus in this many sorted language. For exposition, we will give a version here. In 1-6, A,B are any formulas of L(TT), and x,y are variables of the same type.

1. All tautologies.
2. □xA □ A[x/y], where y is substitutable for x in A.
3. A[x/y] □ □xA, where y is substitutable for x in A.
4. From A □ B derive A □ (□x)(B), where x is not free in A.
5. From A □ B derive (□x)(A) □ B, where x is not free in B.
6. From A and A □ B, derive B.

For the nonlogical axioms for type theory with choice, below A is any formula of L(TT), and $n \geq 0$.

7. Comprehension. There is a class of type n+1 whose elements are exactly those objects of type n obeying any given condition expressible in the language L(TT). The sentences (□x)(□y)(y □ x □ A), where A is a formula of L(TT) in which x is not free.

It is standard to define equality (at each type) using all sentences

x ≡ y □ (□z)(x □ z □ y □ z).

It is implicit in Russell that under this definition of equality, all of the usual axioms for equality are derivable from 1-7. Furthermore Russell saw the equivalence of the above definition of equality with

x ≡ y □ (□z)(x □ z □ y □ z).

8. Choice. For every family of nonempty classes, where any two classes from the family with a common element have the same elements, there is a class containing exactly one element from each of the nonempty classes. All sentences ((∀y ∈ x)(∃z)(z ∈ y) ∧ (∀y,z ∈ x)((∃w)((w ∈ y ∧ w ∈ z) → (∀w)(w ∈ y ↔ w ∈ z)) → (∃u)(∀y ∈ x)(∃v)(v ∈ u ∧ v ∈ y ∧ (∀w)((w ∈ u ∧ w ∈ y) → w ≡ v)).

We call the formal system 1-8, TTC (type theory with choice).

2. Type theory with choice and restriction.

In this section, we straightforwardly expand L(TT) and give the obvious axioms that correspond to this expansion of the language.

We only introduce a single unary predicate symbol R. The language L(TT,R) is the language of type theory with restriction. The atomic formulas are the expressions x ∈ y, R(z), where t(y) = t(x)+1. Formulas are built up from atomic formulas in the usual way using the connectives ¬,∧,∨,→,↔, and the quantifiers ∀,∃.

In 1-6, A,B are any formulas of L(TT,R), and x,y are variables of the same type.

1. All tautologies.
2. ∀xA → A[x/y], where y is substitutable for x in A.
3. A[x/y] → ∃xA, where y is substitutable for x in A.
4. From A → B derive A → (∀x)(B), where x is not free in A.
5. From A → B derive (∃x)(A) → B, where x is not free in B.
6. From A and A → B, derive B.

For the nonlogical axioms for type theory with choice, below A is any formula of L(TT,R), and n ≥ 0.

7. Comprehension. There is a class of type n+1 whose elements are exactly those objects of type n obeying any given condition expressible in the language L(TT,R). The sentences (∃x)(∀y)(y ∈ x ↔ A), where A is a formula of L(TT,R) in which x is not free.

8. Choice. For every family of nonempty classes, where any

two classes from the family with a common element have the same elements, there is a class containing exactly one element from each of the nonempty classes.

We add the axiom of restriction, that tells us that the restriction is nontrivial.

9. Restriction. RE. R fails of some object of type 0. $(\exists x^0)(\neg R(x^0))$.

We call the formal system 1-9, TTCR (type theory with extensionality, choice, and restriction).

A model of TTC, or TTCR, is said to be extensional if and only if the interpretation of ≡ is identity.

Clearly TTC has an extensional model with exactly one object of type 0, that is unique up to isomorphism. The axiom of choice is not needed for this statement.

It is also obvious that TTCR has an extensional model with exactly one object of type 0, and no objects of any type that fall under R, that is unique up to isomorphism.

We close this section with a general comment about restriction.

In the system TTCR + IR + TR that we primarily focus on (see below), it is derivable that the class of all type zero objects will itself fall under R. However, by Restriction, this class has elements that do not fall under R.

This makes sense if we think of classes as a "way" of separating out certain objects from others. That "way" - which might be viewed as an intellectual procedure - can be in some restricted universe, where it maintains its character but gets applied, from the point of view of the restricted universe, only to objects in the restricted universe.

The same considerations apply if we use "predicates" or "properties" throughout instead of classes.

Suppose we think of sets rather classes, in the sense of "completed totalities". Then we have a set in the restricted universe with elements outside the restricted universe. This invites serious objections in a way that using classes does not.

Note we have relied on the well known fact that the theory of types admits more than one closely related informal interpretation (sets as one, and classes as another. These correspond to notions to which Russell's paradox naturally applies.

3. Pure Relativization.

Let A be any formula of L(TT). We construct the formula A^R of L(TT,R) by relativizing all quantifiers (Qx) to R. I.e., replace each

(Qx)

by

(Qx|R)

and expand this out to a formula of L(TT,R) in the obvious way.

The most straightforward relativization principle is

Pure Relativization. PR. $A \leftrightarrow A^R$, where A is a sentence of L(TT).

We consider the formal system TTCR + PR.

The usual formulation of the axiom of infinity in type theory is the existence of a nonempty class of type 2 with no maximal element under inclusion.

THEOREM 3.1. For each $k \geq 1$, TTCR + PR proves "there exists at least k objects of type 0". However, TTCR + PR does not prove the axiom of infinity.

Proof: Fix k and assume in TTCR + PR that there are exactly k objects of type 0. By PR, there must be exactly k objects of type 0 satisfying R. From this it follows logically that every object of type 0 falls under R. This contradicts Restriction.

For the second claim, for each k consider the extensional model M[k] of type theory whose type 0 objects are exactly

{1,...,k}. This is unique up to isomorphism. Let us fix the finite fragment T of PR that uses exactly the sentences $A_1,...,A_p$ from L(TT).

We claim that T has a finite extensional model (finite in each type). By obvious combinatorics, we can find k < r such that M[r] and M[k] satisfy the same A_i's. Make M[r] into a model M[r]* of TTCR + T by taking the extension of R to be M[k]. QED

4. Two principles of impure relativization.

Let A be a sentence of L(TT). We will be interested in relativizing some quantifiers in A to R and leaving other quantifiers unchanged. When we speak of relativizing a quantifier, we will always mean relativizing the quantifier to R.

We now extend axiom system TTCR to include the following impure relativization principles.

10. Initial Relativization. IR. A → A', where A is a sentence of L(TT) and A' is the result of relativizing any initial segment of quantifiers.

We will also consider the following weakening of IR.

10'. IR'. A → A', where A is a sentence of L(TT) and A' is the result of relativizing any initial segment of at most 3 quantifiers.

It turns out that TTCR + IR' is enough to obtain the axiom of infinity, and so this system get close in logical strength to Zermelo set theory. In fact, this system (also with IR) is equiconsistent with BZ = bounded Zermelo set theory. When we add TRP below, the system explodes into hefty reaches of the large cardinal hierarchy.

11. Type Relativization. TR. A → A'', where A is a sentence of L(TT) and A'' is the result of relativizing all quantifiers except the quantifiers of highest type.

We will be particularly interested in TTCR + IR + TR. We will show that this is logically equivalent to TTCR + IR' + TR.

5. Additional principles and logical equivalences.

For formulas A in L(TT), let A^R be the full relativization of A; i.e., the result of relativizing every quantifier to R.

12. Elementary Substructure. ES. $(R(x_1) \wedge \ldots \wedge R(x_k) \rightarrow (A \leftrightarrow A^R)$, where A is a formula of L(TT) whose free variables are among $x_1,\ldots,x_k$.

13. R Completeness. RC. Every class has the same restricted elements as some restricted class. The sentences $(\forall x)(\exists y|R)(\forall z|R)(z \in x \leftrightarrow z \in y)$.

14. All sentences $(\forall x_1,\ldots,x_k|R)(\exists y|R)(\forall z)(z \in y \leftrightarrow A)$ of L(TT,R), where A is any formula of L(TT) in which y is not free.

15. $(\exists x|R)(\forall y|R)(y \in x \leftrightarrow A)$, where A is any formula of L(TT,R) in which x is not free.

LEMMA 5.1. TTCR + IR proves ES.

Proof: Suppose that we can derive ES for all formulas A of L(TT) by induction on the formula A. This is obvious for atomic A.

case 1. A is $\neg B$. Then ES for A follows immediately from ES for B.

case 2. A is B op C, where op is $\wedge, \vee, \rightarrow$, or $\leftrightarrow$. Then TTCR + IRP proves

$(R(x_1) \wedge \ldots \wedge R(x_k)) \rightarrow (B \leftrightarrow B^R)$
$(R(x_1) \wedge \ldots \wedge R(x_k)) \rightarrow (C \leftrightarrow C^R)$

where the free variables of B are among $x_1,\ldots,x_k$. Then TTCR + IR proves

$(R(x_1) \wedge \ldots \wedge R(x_k)) \rightarrow ((B \text{ op } C) \leftrightarrow (B^R \text{ op } C^R))$.

This case is completed since $(B \text{ op } C)^R$ is B^R op C^R.

case 3. A is $(\forall y)(B)$, where the free variables of A are among $x_1,\ldots,x_k$. By the induction hypothesis,

*) TTCR + IR proves $(R(x_1) \wedge \ldots \wedge R(x_k) \wedge R(y)) \rightarrow (B \leftrightarrow B^R)$.

We need to show that TTCR + IR proves

$(R(x_1) \wedge \ldots \wedge R(x_k)) \rightarrow ((\exists y)(B) \rightarrow (\exists y|R)(B^R))$.

Now TTCR (or even just logic) proves

$(\forall x_1,\ldots,x_k)(\exists y)((\exists y)(B) \rightarrow B)$.

Hence TTCR + IR proves

$(\forall x_1,\ldots,x_k|R)(\exists y|R)((\exists y)(B) \rightarrow B)$.

By *), TTCR + IR proves

$(\forall x_1,\ldots,x_k|R)(\exists y|R)((\exists y)(B) \rightarrow BR)$.

Hence TTCR + IR proves

$(R(x_1) \wedge \ldots \wedge R(x_k)) \rightarrow ((\exists y)(B) \rightarrow (\exists y|R)(B^R))$

as required.

case 4. A is $(\forall y)(B)$. By induction hypothesis, we have ES for B, and hence for $\neg B$. Repeating case 3 with $\neg B$ replacing A, we have ES for $(\exists y)(\neg B)$. ES for $(\forall y)(B)$ follows immediately. QED

We wish to prove Lemma 5.1 with IRP replaced by IR'. This requires the development of what amounts to ordered tuples of objects, perhaps from different types.

LEMMA 5.2. Let $x_1,\ldots,x_k,y_1,\ldots,y_k,z$ be distinct variables, where $t(z) = \max(t(x_1),\ldots,t(x_k))+2$, and each $t(x_i) = t(y_i)$. There is a formula B whose free variables are exactly $x_1,\ldots,x_k,z$, in which $y_1,\ldots,y_k$ do not appear, such that TTCR + IR' proves

i) $(\forall x_1,\ldots,x_k)(\exists z)(B)$;
ii) $(\forall x_1,\ldots,x_k|R)(\exists z|R)(B)$;
iii) $(\forall z|R)(B \rightarrow (\exists x_1,\ldots,x_k|R)(B))$;
iv) $(B \wedge B[x_1,\ldots,x_k/y_1,\ldots,y_k]) \rightarrow (x_1 \equiv y_1 \wedge \ldots \wedge x_k \equiv y_k)$.

Proof: (To be polished later). The idea is to first show that every $x|R$ has a $\{x\}$ with R, and every $x,y|R$ of the same type has an $x \cup y$ with R. This shows that every $x_1,\ldots,x_k|R$ of the same type has a $\{x_1,\ldots,x_k\}$ with R.

Next, develop an ordered k-tuple of distinct objects of the same type, where if the objects have R then some k-tuple has R, and also if two k lists of distinct objects have a common k-tuple, then they are ≡. This is done by taking $<x_1,...,x_k>$ to be $\{\{x_1\},\{x_1,x_2\},...,\{x_1,...,x_k\}\}$, which means a set of all sets that look like one of these k terms. Beware that we do not have extensionality, and so each of these k terms may have many realizations, all with the same elements. And then the whole expression also has perhaps a multitude of realizations, all with the same elements. But all realizations of the whole expression must have the same elements.

Assume $x_1,...,x_k$ have R. We have to prove by induction on $1 \leq i \leq k$ that some class of all representations of $\{x_1,...,x_i\}$ has R. We know this by IR' for i = 1. Suppose true for i. We have to show that some class of all representations of $\{x_1,...,x_{i+1}\}$ has R. Let S be a class of all representations of $\{x_1,...,x_i\}$ that has R. Let S' be the result of tacking on x_{i+1} to every element of S. Of course by comprehension, some S' exists. I.e., we can write

$(\forall x)(\forall S)(\exists S')$(S' is the result of tacking on x to the elements of S).

By IR',

$(\forall x|R)(\forall S|R)(\exists S'|R)$(S' is the result of tacking on x to the elements of S).

So this gives us the kind of ordered k-tupling of objects all of the same type that we need.

For each k, there are only finitely many "kinds" of ordered k-tuples, where "kinds" reflect the pattern of repetitions. For each k, we name these kinds using consecutive integers > k.

So let $x_1,...,x_k$ be all of type n, with repetitions allowed. Let p be the name of its "kind". Then the k-tuple of $x_1,...,x_k$ is taken to be a class consisting of all representations of $\{x_1\}$, of $\{x_1,x_2\}$, ... , of $\{x_1,...,x_k\}$, and all classes of type n+1 with exactly p elements. Of course, counting is with respect to ≡. It is easily verified as above that this tupling has the required properties.

We still need to have tupling of objects of varying types. The idea is to first lift them all up to the highest type of the objects. After they are so lifted, we then apply the tupling procedure we just developed.

Let k □ n. Objects of type k are lifted up to type n by n-k iterated singletons. Of course, we must take into account that the singleton operation is many valued in the sense of ≡. But this does not cause problems. QED

LEMMA 5.3. TTCR + IR' proves $(\Box x_1,\ldots,x_k)(\Box y)(A)$ □ $(\Box x_1,\ldots,x_k|R)(\Box y|R)(A)$, provided the formula to the left of □ is a sentence of L(TT).

Proof: Let B be the formula given by Lemma 5.2 using the variables $x_1,\ldots,x_k,y_1,\ldots,y_k,z$. We can assume that $y_1,\ldots,y_k,z$ do not appear in A. We will display all free variables for clarity.

We claim that TTCR + IR' proves

$(\Box x_1,\ldots,x_k)(\Box y)(A(x_1,\ldots,x_k,y)$ □
$(\Box z)(\Box y)((\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)\ \Box\ A(x_1,\ldots,x_k,y))$ □
$(\Box z|R)(\Box y|R)((\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)\ \Box\ A(x_1,\ldots,x_k,y)).$

To see this, first assume $(\Box x_1,\ldots,x_k)(\Box y)(A(x_1,\ldots,x_k,y))$, and let z be given. If $(\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)$ then let $x_1,\ldots,x_k$ be such that $B(x_1,\ldots,x_k,z)$. Choose y such that $A(x_1,\ldots,x_k,y)$. We have to verify that $(\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)\ \Box\ A(x_1,\ldots,x_k,y))$.

Let $y_1,\ldots,y_k$ be such that $B[x_1,\ldots,x_k/y_1,\ldots,y_k]$. By Lemma 5.2, $x_1 \equiv y_1$ □ ... □ $x_k \equiv y_k$. Hence $A[x_1,\ldots,x_k/y_1,\ldots,y_k]$, and our verification is complete.

Now assume $(\Box z)(\Box y)((\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)$ □ $A(x_1,\ldots,x_k,y))$, and let $x_1,\ldots,x_k$ be given. By Lemma 5.2, let z be such that $B(x_1,\ldots,x_k,z)$. Let y be such that $(\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)\ \Box\ A(x_1,\ldots,x_k,y))$. Then $A(x_1,\ldots,x_k,y)$, as required. This establishes the claim.

To complete the proof, it suffices to show that TTCR + IR' proves

$(\Box z|R)(\Box y|R)((\Box x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)\ \Box\ A(x_1,\ldots,x_k,y))$ □
$(\Box x_1,\ldots,x_k|R)(\Box y|R)(A(x_1,\ldots,x_k,y)).$

Assume the left side, and let $R(x_1),\ldots,R(x_k)$. By Lemma 5.2, let z be such that $R(z)$ and $B(x_1,\ldots,x_k,z)$. Let y be such that $R(y)$, $(\forall x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z) \rightarrow A(x_1,\ldots,x_k,y))$. Then $A(x_1,\ldots,x_k,y)$.

Assume the right side, and let $R(z)$.

case 1. $(\exists x_1,\ldots,x_k)(B(x_1,\ldots,x_k,z)$. By Lemma 5.2, let $x_1,\ldots,x_k$ be such that $R(x_1),\ldots,R(x_k),B(x_1,\ldots,x_k,z)$. Let y be such that $R(y),A(x_1,\ldots,x_k,y)$. Let $y_1,\ldots,y_k$ be such that $B[y_1,\ldots,y_k]$. By Lemma 5.2, $x_1 \equiv y_1 \wedge \ldots \wedge x_k \equiv y_k$. Hence $A(y_1,\ldots,y_k,y)$.

case 2. $\neg(\exists x_1,\ldots,x_k)(Bx_1,\ldots,x_k,z)$. Then set y to be arbitrary such that $R(y)$. This follows from IR'.

LEMMA 5.4. TTCR + IR' proves ES.

Proof: Examination of the proof of Lemma 5.1 reveals that IRP' can be used throughout except for one step. This step is

$$(\forall x_1,\ldots,x_k)(\exists y)(A) \leftrightarrow (\forall x_1,\ldots,x_k|R)(\exists y|R)(A)$$

where the left side of the biconditional is a sentence of L(TT). I.e., we must prove this within TTCR + IR'. This is provided by Lemma 5.3. QED

LEMMA 5.5. TTCR + TR proves RC.

Proof: We need to prove the sentences

$$(\forall x)(\exists y|R)(\forall z|R)(z \in x \leftrightarrow z \in y)$$

in TTCR + TR. By TTCR, it suffices to prove

$$*)\quad (\forall u \subseteq R)(\exists y|R)(\forall z|R)(z \in y \leftrightarrow z \in u)$$

in TTCR + TR, where $u \subseteq R$ abbreviates $(\forall x \in u)(R(x))$.

By TTCR,

$$(\forall x)(\exists y)(\forall z)(z \in y \leftrightarrow (\exists w)(w \in x \wedge z \in w)).$$

By TR, since x is the variable whose type is greater than all other types of variables, we have

(□x)(□y|R)(□z|R)(z □ y □ (□w|R)(w □ x □ z □ w)).

Let u □ R. Let x be such that

(□v)(v □ x □ (□b □ u)(v ~ {b})).

Let y be such that

R(y) □ (□z|R)(z □ y □ (□w|R)(w □ x □ z □ w)).

Let z be such that R(z). Then

z □ y □ (□w|R)(w □ x □ z □ w).

Since we are trying to prove *), we need only

(□w|R)(w □ x □ z □ w) □ z □ u.

For the reverse direction, let z □ u. Then R(z). By Lemma ?, let v ~ {z}, R(v). Then z □ v □ x.

For the forward direction, let R(w), w □ x, z □ w. Let b □ u, w ~ {b}. Hence z ≡ b. Therefore z □ u. QED

LEMMA 5.6. TTCR + IR' + RC proves 14 and 15.

Proof: Clearly TTCR proves all sentences

(□x_1,...,x_k)(□y)(□z)(z □ y □ A)

of L(TT), where y is not free in A. By Lemma 5.3, TTCR + IR' proves all sentences

(□x_1,...,x_k|R)(□y|R)(□z)(z □ y □ A)

of L(TT,R), where y is not free in A.

Now let A be any formula of L(TT,R) in which x is not free. Let x be such that (□y)(y □ x □ A). By Lemma 5.5, let x' be such that R(x') and (□y|R)(y □ x □ y □ x'). We can assume that x' is a variable not free in A. Then (□y|R)(y □ x' □ A) since x,x' are not free in A. QED

LEMMA 5.7. TTCR + ES proves IR.

Proof: We will prove the following strengthening of IR. Let A be a formula of L(TT) with free variables among x_1,...,x_n, n ≥ 0. Let A' be an initial relativization of A; i.e., the result of relativizing an initial segment of the quantifiers in A. Then

$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (A \ \square \ A')$$

is provable in TTCR + ES.

We prove this by induction on the formula A. This is obvious for atomic A.

case 1. A is □B. Let A have free variables among $x_1,\ldots,x_n$. Let A' be an initial relativization of A. Let B' be the corresponding initial relativization of B. Then A' is □B'. By the induction hypothesis,

$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (B \ \square \ B')$$

is provable in TTCR + ES. Hence

$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (A \ \square \ A')$$

is provable in TTCR + ES.

case 2. A is B op C, where op is □, ,□.□ , or □ . Let A have free variables among x1,...,xn. Let A' be an initial relativization of A. Let B' be the corresponding initial relativization of B and C' be the corresponding initial relativization of C. Then A' is B' op C'. By the induction hypothesis,

$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (B \ \square \ B')$$
$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (C \ \square \ C')$$

are provable in TTCR + ES. Hence

$$(R(x_1) \ \square \ \ldots \ \square \ R(x_n)) \ \square \ (A \ \square \ A')$$

is provable in TTCR + ES.

case 3. A is (□y)(B). Let A have free variables among $x_1,\ldots,x_n$. Let A' be an initial relativization of A. Without loss of generality, we may assume that the first quantifier (□y) gets relativized. Let B' be the corresponding initial relativization of B. Then A' is (□y|R)(B').

Note that the free variables of B are among $x_1,\ldots,x_n,y$. By the induction hypothesis,

$$*)\ (R(x_1) \square \ldots \square R(x_n) \square R(y)) \square (B \square B')$$

is provable in TTCR + ES. We wish to prove

$$(R(x_1) \square \ldots \square R(x_n)) \square ((\square y)(B) \square (\square y|R)(B'))$$

in TTCR + ES. Assume $R(x_1) \square \ldots \square R(x_n)$. By ES and *), we have

$$(\square y)(B) \square (\square y|R)(B^*) \square (\square y|R)(B) \square (\square y|R)(B')$$

where B* is the full relativization of B.

case 4. A is $(\square y)(B)$. Repeat case 4 with $\square$ replaced by $\square$. QED

LEMMA 5.8. TTCR + RC proves TR.

Proof: We will prove a stronger statement by induction on formulas A of L(TT).

For variables x,y of the same type, we write $x \sim_R y$ for $(\square z|R)(z \square x \square z \square y)$.

Let A be a formula of L(TT) in which the highest type of any variable is n. Assume that the free variables of A are among the variables $x_1,\ldots,x_k,y_1,\ldots,y_r$, where $x_1,\ldots,x_k$ have type n and $y_1,\ldots,y_r$ have type < n. Let $z_1,\ldots,z_k$ be distinct variables not in A, where each $t(z_i) = t(x_i)$. Let A' be the result of relativizing at least all quantifiers of type < n. Then TTCR + RC proves

$$(x_1 \sim_R z_1 \square \ldots \square x_k \sim_R z_k \square R(y_1) \square \ldots \square R(y_r)) \square (A \square A'[x_1,\ldots,x_k/z_1,\ldots,z_k]).$$

QED

THEOREM 5.9. The systems TTCR + IR + TR, TTCR + IR' + TR, TTCR + ES + RC are logically equivalent. They prove 14, 15.

Proof: By Lemmas 5.4, 5.5, we see that TTCR + IR' + TR logically implies ES + RC. By Lemmas 5.7, 5.8, TTCR + ES + RC logically implies IR + TR. Hence TTCR + IR' + TR logically implies IR + TR, and so TTCR + IR' + TR and TTCR + IR + TR are logically equivalent. They prove 14, 15 by

Lemma 5.6. QED

6. Interpreting extensionality.

It is very useful to have

16. Extensionality. EXT. All sentences of the form $(\square z)(z \square x \square z \square y) \square x \equiv y$.

This is not provable in TTCR + IR + TR. However, we show that TTCR + IR + TR + EXT is interpretable in TTCR + IR + TR.

For this purpose, we define the hereditarily extensional objects as follows. The HE objects of type 0 are just the objects of type 0.

The HE objects of type k+1 are just the classes xk+1 such that

i) every element of x^{k+1} is an HE object of type k;
ii) any object of type k that is extensionally equal to an element of x^{k+1} is an element of xk+1.

The HE interpretation of any formula of L(TT,R) is obtained by relativizing each quantifier to the HE objects of its type, and expanding the formula out to a formula of L(TT,R). Of course, if the original formula does not mention R then its HE interpretation does not mention R.

We write A^{HE} for the HE interpretation of A.

LEMMA 6.1. The HE interpretation of the closure of each instance of EXT^{HE} is provable in TTCR.

Proof: We show that TTCR proves

$$(\square x,y|HE)((\square z|HE)(z \square x \square z \square y) \square$$
$$(\square z|HE)(x \square z \square y \square z)).$$

Assume HE(x), HE(y), $(\square z|HE)(z \square x \square z \square y)$, HE(z). Since x,y consist entirely of z's with HE(z), clearly x,y are extensionally equal. Hence $x \square z \square y \square z$. QED

LEMMA 6.2. TTCR proves every sentence $(\square x_1,\ldots,x_n|HE)(\square y_1,\ldots,y_n|HE)(A \square A[x_1,\ldots,x_n/y_1,\ldots,y_n])$, where the free variables of A are among the variables

$x_1,\ldots,x_n$, A does not mention $y_1,\ldots,y_n$, for all $1 \leq i,j \leq n$, x_i is $x_j \leftrightarrow y_i$ is y_j, and $t(x_i) = t(y_i)$.

LEMMA 6.3. The HE interpretation of the closure of each instance of Comprehension is provable in TTCR.

Proof: We show that TTCR proves each sentence

$$(\forall x_1,\ldots,x_n|HE)(\exists y|HE)(\forall z|HE)(z \in y \leftrightarrow A^{HE})$$

where y is not free in A.

Let $x_1,\ldots,x_n$ have HE. Let y be such that $(\forall z)(z \in y \leftrightarrow (HE(z) \wedge A^{HE}))$. By Lemma 6.2, HE(y). Clearly $(\forall z|HE)(z \in y \leftrightarrow A^{HE})$. QED

LEMMA 6.4. The HE interpretation of the closure of each instance of Choice is provable in TTCR.

Proof: It suffices to prove the following in TTCR:

Let S be an HE family of nonempty classes, where any two classes from the family with a common element are extensionally equal. Then there is an HE class containing exactly one element, up to extensional equality, from each of the nonempty classes.

Under the hypotheses, by Choice there is a class containing exactly one element, up to ≡, from each of the nonempty classes. This is even more than we need. QED

LEMMA 6.5. The HE interpretation of Restriction is provable in TTCR.

Proof: Immediate. QED

LEMMA 6.6. Let A be a formula of L(TT). Then TTCR + ES proves $(A^R)^{HE} \leftrightarrow (A^{HE})^R$.

Proof: We proceed by induction on A. This is obvious if A is atomic.

case 1. A is $\neg B$. Then $(A^R)^{HE} \leftrightarrow (A^{HE})^R$ is $\neg(B^R)^{HE} \leftrightarrow \neg(B^{HE})^R$.

case 2. A is B op C. Then $(A^R)^{HE} \leftrightarrow (A^{HE})^R$ is $((B^R)^{HE}$ op $(C^R)^{HE})) \leftrightarrow ((B^{HE})^R$ op $(C^{HE})^R)$.

case 3. A is $(\forall y)(B)$. Then $(A^R)^{HE} \leftrightarrow (A^{HE})^R$ is

$$(\forall y)((HE(y) \wedge R(y) \rightarrow (B^R)^{HE}) \leftrightarrow$$
$$(\forall y)((HE(y))^R \wedge R(y) \rightarrow (B^{HE})^R).$$

By ES, this is equivalent to

$$(\forall y)(HE(y) \wedge R(y) \rightarrow (B^R)^{HE}) \leftrightarrow$$
$$(\forall y)(HE(y) \wedge R(y) \rightarrow (B^{HE})^R)$$

which is provable in TTCR by the induction hypothesis.

case 4. A is $(\exists y)(B)$. By the induction hypothesis, the claim holds for B, and hence for $\neg B$. We can repeat case 3 and obtain the claim for $(\forall y)(\neg B)$. We conclude the claim for $(\exists y)(B)$. QED

LEMMA 6.7. The interpretation of the closure of each instance of ES is provable in TTCR + ES.

Proof: Let A be a formula of L(TT) whose free variables are among x1,...,xn. We need to show that TTCR proves

$$((\forall x_1,\ldots,x_n|R)(A \leftrightarrow A^R))^{HE}.$$

This sentence is

$$(\forall x_1,\ldots,x_n|HE)((R(x_1) \wedge \ldots \wedge R(x_n)) \rightarrow (A^{HE} \leftrightarrow (A^R)^{HE})).$$

By Lemma 6.6, this is provably equivalent over TTCR to

$$(\forall x_1,\ldots,x_n|HE)((R(x_1) \wedge \ldots \wedge R(x_n)) \rightarrow (A^{HE} \leftrightarrow (A^{HE})^R)).$$

This follows immediately from ES. QED

LEMAM 6.8. The interpretation of each instance of RC is provable in TTCR + ES + RC.

Proof: We have to prove

$$(\forall x|HE)(\exists y|R|HE)(\forall z|R|HE)(z \in x \leftrightarrow z \in y)$$

in TTCR + ES + RC. Let HE(x). Let R(y), where x,y have the same elements falling under R. Then

$$(\forall z|R)(z \in y \rightarrow HE(z)).$$

By ES,

$$(\forall z)(z \in y \rightarrow HE(z)).$$

Also

$$(\forall z,u|R)((z \in y \wedge z,u \text{ are extensionally equal}) \rightarrow u \in y).$$

This is because for z,u with R, if $z \in y$ and z,u are extensionally equal, then $u \in x$. Since R(u), we have $u \in y$.

By IR,

$$(\forall z,u)((z \in y \wedge z,u \text{ are extensionally equal}) \rightarrow u \in y).$$

We have verified that HE(y). QED

THEOREM 6.9. TTCR + ES + RC + EXT is interpretable in TTCR + ES + RC.

Proof: By Lemmas 6.1, 6.3, 6.4, 6.5, 6.7, 6.8. QED

7. Sketch of proof that TTCR + ES + RC is very strong.

We will show that ZFM = ZFC + "there exists a measurable cardinal" is interpretable in TTCR + ES + RC, or alternatively, in TTCR + IR + TR. By Theorems 5.9 and 6.9, it suffices to interpret ZFM in 1-16.

A linear ordering of type n objects consists of a class of type n+2 which is linearly ordered under inclusion, and where every object of type n lies in some element of the linear ordering. One type n object is less than another in the sense of the linear ordering if and only if the first lies in some element of the linear ordering that the second does not lie in.

A well ordering of type n objects is a linear ordering of type n objects where every nonempty set of type n objects has a least element.

LEMMA 7.1. 1-16 proves that there is a well ordering of the objects of any given type.

Proof: (to be polished later). Adapt the Zermelo well ordering theorem to this context. QED

THEOREM 7.2. It is provable in 1-16 that there is a countably additive nonprincipal ultrafilter on all subsets of R[0].

Proof: Let W be a well ordering of the class 0 objects. By IR, we can assume that R(W).

Suppose W does not have a limit point. Then it is easy to prove that every point falls under R. Hence by R, W has a limit point. Therefore it has a least limit point, and the initial segment up to there will serve as the natural numbers. One also sees that that initial segment also falls under R, and so it also serves as the natural numbers in the sense of R. Full induction and recursion is supported both in the usual sense and in the restricted universe represented by R.

Let R[0] = $\{x^0: R(x^0)\}$. Fix c = c^0 with $\neg$R(c). Let X $\subseteq$ R[0]. By RC, let R(X*), where X* and X have the same elements falling under R. By ES, X* is unique.

Let K = {X $\subseteq$ R[0]: c $\in$ X*}. Note that if X is a singleton then X* = X. Obviously K is an ultrafilter on the subsets of R[0], where singletons lie outside K.

We will show that K is a countably complete ultrafilter on the subsets of R[0].

Let $X_0,X_1,\ldots$ be an infinite sequence of elements of K indexed by the natural numbers in the sense of the second paragraph. So c lies in $X_0^*,X_1^*,\ldots$.

We claim that R($\{X_0^*,X_1^*,\ldots\}$). To see this, it suffices to prove that Y = $\{\{\{0\},X_0^*\},\{\{1\},X_1^*\},\ldots\}$ falls under R.

First note that each {i} falls under R. Since each X_i^* falls under R, we see that each $\{\{i\},X_i^*\}$ falls under R.

Let X be the intersection of the X_i's. We claim that X* is the intersection of the X_i^*'s. To see this, first note that X* and the intersection of the X_i^*'s fall under R. The former is by the definition of the * operator, and the latter is from the fact that Y falls under R, and IR.

Hence to see that X* is the intersection of the X_i^*'s, it suffices to show that the two sets have the same elements that fall under R. Now the elements of X* that fall under R are precisely the elements of X. So it suffices to show that the elements common to all X_i*'s, that fall under R, are exactly the elements of X.

Let x $\in$ X. Since each X_i includes X, we see that each X_i* includes X*, and so x lies in each X_i*.

On the other hand, suppose x lies in each X_i* and R(x). By the definition of the * operator, x lies in each X_i. Hence x $\in$ X.

We have now established that X* is the intersection of the X_i*'s. Hence c $\in$ X*. Therefore X $\in$ K, as required. QED

THEOREM 7.3. The following is provable in PA. If TTRC + ES + RC (or TTRC + IR + TR) is consistent then so is BZCM = BZC + "there exists a measurable cardinality".

Let ZC is Zermelo set theory with the axiom of choice, and BZC is the slightly weaker system where separation is restricted to formulas with bounded quantifiers. Also, we write "measurable cardinality" because without replacement in ZC, we cannot get a good theory of ordinals and cardinals. But nevertheless, we can well order any set in BZC, and so "measurable cardinality" carries a great deal of force. E.g., by standard arguments, ZFC + "there exists a Ramsey cardinal" is interpretable in BZC + "there exists a measurable cardinality".

However, we can go much further. But first we will go just a bit further to ZFM as follows.

THEOREM 7.4. ZFM is interpretable in TTRC + ES + RC (or TTRC + IR + TR).

Proof: Look at the previous argument where c is chosen to be the least element not in R[0] in the well ordering W with R(W). One can go further and see that c is in fact a measurable cardinal, by checking < c additivity of the ultrafilter K.

Now let d (as a point in W) be the least measurable cardinal. Since d is definable from W, and W falls under R, we see that d falls under R. We have thus shown that there are at least two measurable cardinals. In fact, we get much more than ZFC + there exists a measurable cardinal out of this. QED

We now show how to obtain an interpretation of ZFC + {there is an elementary embedding from some V(□+n) into some V(□+n), □ < □}$_n$ in TTRC + ES + RC (or TTRC + IR + TR).

NOTE: TO BE COMPLETED LATER.

8. Interpreting 1-16.

Let j:V(□+□) □ V(□+□) be a nontrivial elementary embedding with critical point □ < □. We take the set of type 0 objects of U to be V(□), and build up from there using V(□+□). We take the objects of R to be the values of j.

The only delicate point is the verification of RC. This amounts to verifying the following. Let S be a subset of j[V(□+n)]. Then there exists x □ rng(j) such that x,S have the same elements from rng(j).

Fix S □ j[V(□+n]) and write S = j[S'], S' □ V(□+n). Note that for all y □ V(□+n),

j(y) □ j(S') □ y □ S' □ j(y) □ S.

Hence j(S') and S have the same elements from j[V(□+n)], and hence the same elements from rng(j). QED

9. The general theory of relativization.

We use a language LTQ (language of typed quantifiers) for discussing the quantifiers in arbitrary sentences of L(TT).

We write p(Q,A) for the position of the quantifier Q in the sentence A.

We write t(Q,A) for the position of the type of the quantifier Q in the sentence A in the list of types of quantifiers in A in decreasing order. Thus the quantifiers Q of highest type in A have t(Q,A) = 1.

The atomic formulas in LTQ are

$$p(Q) = i$$
$$t(Q) = i$$

where i ≥ 1. Here Q is the unique variable of LTQ and each i

≥ 1 is treated as a constant rather than a variable. Formulas in LTQ are built up from the atomic formulas of LTQ using the usual connectives □,□, ,□ ,□ .

Note that if A has exactly k quantifiers and i > k, then p(Q) = i will be satisfied by no quantifier. Also if A has exactly k sorts and i > k, then t(Q) = i will be satisfied by no quantifier.

A relativization rule is just a formula □ in LTQ.

Relativization rules □ are intended to apply to sentences A of L(TT), resulting in a sentence □(A) of L(TT,R). The idea is that □(A) is the result of relativizing exactly those quantifiers Q in A that obey □.

A relativization principle is again given by a relativization rule □, and is identified with the scheme of all equivalences

A □ □(A)

where A is a sentence of L(TT).

Note that axiom IR' can be construed as three separate relativization principles given by the formulas

p(Q) = 1 □ □p(Q) = 1 (relativize no quantifiers)
p(Q) = 1 (relativize only the first quantifier)
p(Q) = 1 p(Q) = 2 (relativize only the first two quantifiers)
p(Q) = 1 p(Q) = 2 p(Q) = 3 (relativize only the first three quantifiers).

Note that axiom IR can be construed as the infinitely many relativization principles given by the formulas

p(Q) = 1 ... p(Q) = k

where k ≥ 0. The empty disjunction corresponds to truth, which means that no quantifiers are relativized.

Note that axiom TR can be construed as the single relativization principle given by the formula

t(Q) = 1.

THEOREM 8.1. Let be a relativization principle (given by a formula in LTQ). Then TTCR + PR + is consistent if and only if 1-16 proves .

THEOREM 8.2. 1-16 is logically equivalent to TTCR together with the union of all relativization principles (given by a formula in LTQ) which are consistent with TTCR + PR.

10. Removing choice.

We are working on a version without choice. The axiom of restriction must be strengthened. The basic idea is that any "large" class must have elements that do not fall under R. The interpretation power will be roughly the same as above, or at least as large as measurable cardinals of high order.

11. Towards very general contexts.

Suppose we start with a theory (with no constant or function symbols) which is axiomatized by schemes involving all formulas in its language. We can then add a unary predicate symbol R, and restate the axioms where the schemes are now extended to all formulas in the expanded language. We then add a principle of restriction - either that there is at least one object that does not fall under R, or perhaps that the "large" extensions of formulas have at least one object that does not fall under R. Then we make a study of the relativization of sentences. The idea is that in many contexts, perhaps one gets enormous logical strength out of this.

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 6
11/6/2002
11/7/2002

NOTE: We talked about a simplifying approach via the cumulative theory of types, but I don't think this works. Also we talked about subjective isomorphisms, and this is a big topic with big potential which I am currently working out. So that should appear in some supplementary notes later.

****

We discuss a new approach, related to Lecture 5, for getting at what's philosophically behind set theory with large cardinals. We comment on the relationship between this and the previous work.

The previous work had two aspects. A first is the relationship between a type structure and a type substructure, which leads to reasonable axioms concerning their interaction. The intention is to see the philosophical plausibility and/or clarity of what is being asserted.

The second is that after writing down the obvious axioms for a type structure and a type substructure, not including the two axiom schemes that drive the first aspect, we can make a general study of relativization of quantifiers in sentences about the type structure to the type substructure. Partial results support the conjecture that one can completely analyze all syntactic relativization rules and show that this leads to an array of formal systems that line up linearly with a highest level logically equivalent to what we obtain in the previous paragraph, which corresponds to roughly extendible cardinals. One also has support for the conjecture that if a set of relativization rules are applied, either they lead to "obvious" inconsistencies, or they are consistent.

We think that the new approach below seems arguably more compelling and fundamental than the first aspect above. However, it does not seem to be any kind of direct competitor for the second aspect above.

1. Type theory.

We begin with the theory of types, TT, with variables over each sort 0,1,2,..., and where the atomic formulas are of the form

$x^n \in y^{n+1}$.

The sole axioms are the full comprehension axioms, which are formulas in the language of the form

$(\exists x)(\forall y)(y \in x \leftrightarrow \varphi)$

where φ is a formula in the language in which x is not free.

This system has an obvious model with exactly one object of type 0, and finitely many objects at each type.

We now consider the following principle. To formalize it in TT we rely on well known methods of formalization in TT.

Call a set at any nonzero type large if and only if it cannot be mapped one-one into type 0 objects. The formalization of this concept in TT is well known using Leibniz equality

$$x \equiv y \leftrightarrow (\forall z)(x \in z \leftrightarrow y \in z)$$

and ordered pairing across types. Ordered pairing across types is achieved by leveling out the types using iterated singletons.

In TT, binary relations are treated as sets of ordered pairs of objects of a given type. The ordered pair <x,y> is taken to be {{x},{x,y}}, where there are possibly many such ordered pairs. In fact there are possibly many {x}. All of them have exactly one element, namely x, and this is formulated using Leibniz equality.

A relation on type n is a set of ordered pairs of objects of type n. Its field may be any set of objects of type n. A relation on type n is itself of type n+3.

Here is the statement of the first principle.

T1. Every large relation on type n is second order (higher order) equivalent to a proper subset.

For each $n \geq 1$ this is a statement or scheme. For second order, or any specific level of order, we can either formalize T1 for n as a single sentence using a truth definition for second order logic, or any specific level of order, available in TT, or formalize it as a scheme.

The most straighforward formalization of T1 is as a double scheme, indexed by the choice of n and by the choice of finitely many sentences in higher order logic that are to be transferred. This does not require the construction of any truth predicates within T1.

There are really 3 different kinds of formalization of T1.

i) use truth definitions as much as possible. Still, we cannot consolidate the types n nor can we consolidate the levels of order (in higher order);

ii) use schemes entirely, avoiding any truth definitions, but only for choices of single higher order sentences to be transferred.

iii) use schemes entirely, avoiding any truth definitions, and use arbitrary finite sets of higher order sentences to be transferred.

THEOREM 1.1. All three ways of formalizing T1 are logically equivalent. T1 is mutually interpretable and equiconsistent with ZFC + $\{$there exists an n-extendible cardinal$\}_n$.

The calculation of specific n's and levels of order that give high strength is unclear. We are confident that n = 3 and second order, using schemes throughout without truth definitions, gives more than a measurable cardinal in strength, and at least something close to a 1-extendible cardinal in strength.

As we shall see in section 5, the axiomatization is simplified considerably in the cumulative theory of types because the notion of largeness can be avoided completely.

2. Type theory with choice.

Relations whose field is the set of all type 0 objects are of particular importance, and we wish to accomplish what we did in section 1 with such relations. We call such relations

full relations on type 0.

In section 3 we consider full relations on type n.

We will be using the axiom of choice in the following form, in every type.

> Let x be a set of nonempty sets, any two of which have the same elements or no elements in common. There exists a set which has exactly one element in common with each element of x.

Here "exactly one" is formulated in terms of Leibniz equality. Also x is of any particular type ≥ 2.

This results in the system TTC (type theory with choice). We consider the following second order principles. The remarks about 3 basic choices of formalization made in section 1 apply here as well.

It seems clear, but we have not checked the details, that TT plus any or all of the principles T2 - T7 below should be no stronger than TT plus the axiom of infinity. So choice is really needed.

T2. Every full relation on type 0 is second order (higher order) equivalent to a proper subset.

THEOREM 2.1. TTC + T2 corresponds roughly to indescribable cardinals. At higher types, we still do not go beyond such cardinals.

T3. Every full relation on type 0 is second order (higher order) equivalent to two proper subsets, neither of which is a subset of the other.

THEOREM 2.2. TTC + T3 corresponds roughly to a 1-extendible cardinal.

T4. Every full relation on type 0 is first order equivalent to a proper subset of the same cardinality.

THEOREM 2.3. TTC + T4 interprets sharps and is interpretable in a Ramsey cardinal.

This is because T4 is intimately related to Jonsson cardinals in set theory.

T5. Every full relation on type 0 is second order (higher order) equivalent to a proper subset of the same cardinality.

THEOREM 2.4. TTC + T5 corresponds roughly to a nontrivial elementary embedding from V into M, where M is a transitive class containing the power set of the first fixed point above the critical point, in the case of second order. For higher order, roughly L(V(□)) into itself, which is below V(□+1) into itself.

A nontrivial embedding from some $V(\lambda+1)$ into itself is the strongest large cardinal axiom normally considered that is believed to be consistent with the axiom of choice. A bit stronger and more technical is $L(V(\lambda+1))$ into itself.

T6. Every full relation on type 0 is isomorphic to a proper subset.

THEOREM 2.5. T6 is inconsistent.

3. Type theory with extensionality and choice.

T7. Every full relation on type n is second order (higher order) equivalent to a proper subset.

THEOREM 3.1. TTEC + T7 for $n = 0$ corresponds roughly to an indescribable cardinal. For $n \geq 1$ it corresponds roughly to an n-extendible cardinal.

4. Type theory with indistinguishability.

We now discuss principles of indistinguishability in type theory. The indistinguishability of relations is stronger than their higher order equivalence. We generally obtain greater logical strength using indistinguishability rather than higher order logical equivalence.

Let x,y be of the same type. We say that they are indistinguishable if and only if they obey the same properties with no parameters. This is formulated as a scheme.

Recall the notion of large in TT. A full relation on A is a binary relation whose field is A.

In TT, we consider

T8. In each large cardinality and type there is a set A such that every full relation on A is indistinguishable from a proper subset (superset).

This can be formalized with finite conjunctions of transfer, or with single property transfer. They are provably equivalent. Also it is provably equivalent to formalizations that use truth predicates available in TT.

THEOREM 4.1. TT + T8 interprets extendible cardinals and is interpretable in Vopenka's principle.

In TTC, we consider

T9. There is a set A of type 0 objects such that every relation with field A is indistinguishable from a proper subset (proper superset).

THEOREM 4.2. TTC + T9 corresponds roughly to a subtle cardinal. These are stronger than indescribable cardinals.

T10. There is a set A of type 0 objects such that every relation with field A is indistinguishable from two proper subsets (proper supersets), neither of which is a subset of the other.

THEOREM 4.3. TTC + T10 has interpretation power between an extendible cardinal and Vopenka's principle.

T11. There is a set A of type 0 objects such that every relation with field A is indistinguishable from two isomorphic proper subsets (proper supersets), one of which is a proper subset of the other.

THEOREM 4.4. TTC + T11 corresponds to roughly a nontrivial elementary embedding from a rank into itself.

PHILOSOPY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 7
11/13/02
11/14/02

We now take up our second topic: logic in the universal domain.

1. General remarks.

Frege intended his quantifiers to range over absolutely everything. An obvious question is:

> if the quantifiers are to range over absolutely everything in first order predicate calculus then what sentences are satisfiable?

The interpretation of predicate calculus in any domain involves constants, multivariate relations, and multivariate functions on that domain. When the domain is a set, this is a matter of set theory, and is well understood. Even here, there are some interesting delicate issues, as we shall see.

But when the domain is something outside the realm of set theory, and outside the realm of mathematics, then we need to think carefully about the nature of multivariate relations and functions on that domain.

It could be the case that the answer to our question is greatly dependent on the precise nature of the multivariate relations and functions on the universal domain. However, we shall see that only relatively modest - and plausible - principles about the relations and functions on the universal domain are sufficient to determine the answer to our question. For relatively modest fragments of predicate calculus, the principles needed are relatively modest. As we bite off more and more of predicate calculus, we will need sharper principles.

In this regard, the situation is not all that different from the usual set theoretic interpretation of predicate calculus (in set theoretic domains). Here, very little is needed about the nature of the multivariate relations and functions on the domain in order to determine the sentences satisfiable in that domain.

The relevant principles about the multivariate relations and functions on the universal domain generally assert that every multivariate relation acts very symmetrically on some (usually finite) list of distinct objects. Functions will throughout be treated as univalent relations. Thus the relevant principles are principles of indiscernibility.

2. Official presentation of predicate calculus.

We care about full predicate calculus and also its fragments. In full predicate calculus we have

i) variables x_n, $n \geq 1$;
ii) constant symbols c_n, $n \geq 1$;
iii) relation symbols $R^n{}_m$, $n,m \geq 1$. The arity is n;
iv) function symbols $F^n{}_m$, $n,m \geq 1$. The arity is n;
v) connectives □,□, ,□ ,□ ;
vi) quantifiers □,□;

vii) equality =.

The terms are given by the following inductive clauses:

i) every variable and constant is a term;
ii) if F is a function symbol of arity n and $t_1,...,t_n$ are terms, then $F(t_1,...,t_n)$ is a term.

The atomic formulas are the expressions

$$s = t$$
$$R(t_1,...,t_n)$$

where $s,t,t_1,...,t_n$ are terms, and R is an n-ary relation symbol.

The formulas are given by the following inductive clauses:

i) every atomic formula is a formula;
ii) if □,□ are formulas then so are (□□), (□ □), (□ □ □), (□ □ □), (□ □ □).
iii) if □ is a formula then (□x_n)(□), (□x_n)(□) are formulas.

We call the above language, PC(=). The semantics of PC(=) are well known, using any nonempty domain of objects. Of course, this involves the concepts of multivariate relation/function on that domain. The semantics of PC(=) was first given formally by Tarski.

We are interested in fragments of PC(=). A particularly important way of specifying fragments is in terms of giving a set of the constant, relation, and function symbols, and also asserting whether or not = is allowed. The constant, relation, and function symbols are called the nonlogical symbols.

Let □ be a set of nonlogical symbols. We write PC(□) for predicate calculus using only nonlogical symbols from □. We write PC(□,=) for predicate calculus using only nonlogical symbols from □ and =.

Sometimes constant and func-tion symbols are awkward, and we use the important fragments PC(rel) and PC(rel,=). Here "rel" is the set of all relation symbols.

Another independent way of designating fragments of PC(=) is to specify a quantifier prefix for all formulas to be considered. E.g., such expressions as PC(□,□...□), PC(□,=,□...□□...□) are self explanatory.

3. Set domains.

We now take up the case where the domain is a set. This is definitely not the case of the universal domain:

THEOREM 3.1. No set has everything in it. In particular, no set has all sets in it.

Proof: Let S be a set. Consider the set $x = \{y \in S: y \notin y\}$. Since S is a set, x is a set. If $x \in S$ then $x \in x \leftrightarrow x \notin x$ which is a contradiction. QED

THEOREM 3.2. A set of sentences of PC(=) is satisfiable (in some nonempty set) iff it is consistent within any one of the many standard systems of axioms and rules of inference for PC(=).

How can this be established conveniently with a minimum of commitment? Since the domains are sets, it is natural for us to formulate these Theorems in set theory. Let BST' (basic set theory prime) be the following system in $\in$,=:

1. Extensionality.
2. Pairing.
3. Union.
4. The set of all finite subsets of any set exists.
5. Separation for bounded formulas.
6. Infinity. There is a least set x such that $\emptyset \in x$ and $(\forall y \in x)(y \cup \{y\} \in x)$.
7. Induction for all formulas.

THEOREM 3.3. Theorem 3.2 is conveniently formalizable and is provable in BST'.

Proof: This is enough to conveniently handle the syntax of PC(=) and the axioms and rules of inference. We can also conveniently handle the finite tuples from x of any indefinite length. This supports multivariate relations and functions. Thus we have structures of the kind that are the interpretations of PC(=).

We have to formalize the notion of truth of a sentence of PC(=) in a structure with a given nonempty set domain. We use the variant of Tarksi's truth definition, which involves partial satisfaction relations. These partial satisfaction relations exist of any indefinite level by a basic induction argument.

The forward direction is by a well known induction. We show that in every proof in PC(=) from a set of sentences of PC(=), every line is true in any given model of that set of sentences. Hence a satisfiable set of sentences of PC(=) must be consistent.

We now come to the reverse direction. Here is where matters could get quite delicate but they do not. The usual proofs of completeness start with a consistent set S of sentences in PC(=) and find a model of S with domain N = the set of all nonnegative integers, but with = interpreted as an equivalence relation. In fact, one actually constructs a bit more than is needed: the entire satisfaction relation for the structure. This construction can be easily carried out in BST'. Then the factoring out by the equivalence relation is readily carried out in BST'. This results in a model whose domain is a nonempty initial segment of N (possibly all of N). We will also have the entire satisfaction relation for the structure. QED

Note we skirted all philosophical issues in establishing Theorem 3.2. This is provided one accepts BST', with its minimum set theoretic commitments. This is because the domains that are the proper initial segments of N are sufficient, and the multivariate relations/functions on them needed for satisfying consistent theories are so directly definable from the theories.

We now consider a more delicate matter. Let us be given a nonempty set D, which will serve as the domain of our interpretations of PC(=). What can we say about the sets of sentences of PC(=) that are D-satisfiable; i.e., have a model with domain D?

If D is finite, then of course there are no philosophical conundrums involved. However, what if D is infinite?

INF is the set of sentences

$$(\exists x_1)\ldots(\exists x_n)(x_1 \neq \ldots \neq x_n).$$

DIGRESSION. This sentence has length quadratic in n. However, the logically equivalent reaxiomatization

$$(\forall x_1)\ldots(\forall x_{n-1})(\exists x_n)(x_n \neq x_1 \wedge \ldots \wedge x_n \neq x_{n-1})$$

has length linear in n.

THEOREM 3.4. The following is provable in BST'. A set of sentences in PC(=) is N-satisfiable iff it is consistent with INF. If a set of sentences in PC(=) is E-satisfiable for finite E of infinitely many different sizes, then it is N-satisfiable.

Let BST* be

1. Extensionality.
2. Pairing.
3. Union.
4. The set of all finite subsets from any set exists.
5*. Separation for all formulas.
6. Infinity.

A well known generalization of the completeness theorem, nor-mally proved using the compactness theorem for arbitrary sets of sentences, will establish the following.

THEOREM 3.5. The following is provable in BST*. Let D be an infinite well ordered set. A set of sentences in PC(=) is D-satisfiable if and only if it is N-satisfiable if and only if it is consistent with INF.

From the modern set theoretic point of view, every set is well ordered. Hence Theorem 3.5 disposes of the question of characterizing the D-satisfiable sets of sentences in PC(=). If D is finite then there is nothing much to say (although interesting from the point of view of computational complexity), and the D-satisfiable sets of sentences in PC(=) are all incomparable under inclusion as the size varies. On the other hand, if D is infinite, then the D-satisfiable sets of sentences in PC(=) are all the same and different from any of the finite domain cases.

But we are interested in non set theoretic domains, and in particular the universal domain. We maintain that there is no good reason to think that all domains are set theoretic in character, at least in the sense that they have a well ordering.

In fact, let us recall just why set theorists believe that any set is well ordered. This is the Zermelo well ordering theorem. To prove it, we basically need a choice function that picks an element out of every nonempty subset.

But which choice function? In set theory, we just take its existence as part of the setup of set theory. It is now standard to do this, armed with the knowledge that it facilitates a lot of elegant abstract mathematics. Another rationale is to say that it is inherent in the very concept of set that any imaginable arrangement for a set is realized. And a choice function - as a set theoretic object - is just such an imaginable arrangement.

However, we can either take the concept "domain" to be quite different, where we require explicitness, or at least some semblance of explicitness for carving out relations and functions on the domain. E.g., this may have to do with some inherent indiscernibility of the elements of the domain that one is compel-led to respect. Or, alternatively, we can say that we are interested in interpreting PC(=) by means of explicit predicates and functions.

In any case, it would seem that no matter how we talk about this, we must adopt such a view about the universal domain if we are going to adopt such a view about any domain. E.g., if we can well order the universal domain then we can clearly well order any domain. Even if we can linearly order the universal domain then we can linearly order any domain. So it seems reasonable to take as inherent in the universal domain that it cannot be well ordered, or even linearly ordered.

If the universal domain does not have a linear ordering, then of course we know that the sentence

R is a linear ordering

is not satisfiable, and so we are in for some work to figure out what the satisfiable (sets of) sentences are. Clearly the universal domain is not behaving like any ordinary infinite set like N.

Is there an interesting threshold for richness of the multivariate relations on D so that D-satisfiability is the same as N-satisfiability?

A beautiful one. Let us call D logically complete iff for all sets of sentences in PC(=), D-satisfiability is the same as N-satisfiability.

Let FS(D) be the set of all finite sequences from D. Finite sequences are defined as functions from proper initial segments of ω, which are defined in terms of ordered pairs, which are in turn defined in terms of unordered pairs. Note that FS(D) can be proved to exist in BST'.

LEMMA 3.6. The following is provable in BST'. Let $F:DxD \rightarrow D$ be one-one, where D has at least two elements. Then F has a countably infinite subset. Furthermore, there is a one-one $G:FS(D) \rightarrow D$.

Proof: Let $F:DxD \rightarrow D$ be one-one. Fix x_1,x_2 to be two distinct elements of D. Since F is one-one, there are four values of F at arguments x_1,x_2. Set x_3 to be the first one that is not x_1,x_2. Now there are nine values of F at arguments x_1,x_2,x_3. Set x_4 to be the first one that is not x_1,x_2,x_3. Continue in this way. This generates a countably infinite subset of D.

Let $x \in FS(D)$. Write $x = \langle x_1,\ldots,x_n\rangle$ and prove the existence of the finite sequence $y_1,\ldots,y_n$ where $y_1 = x_1$, $y_2 = x_2$, and for $i \geq 3$, $y_i = F(y_{i-2},y_{i-1})$. Let $z = \langle z_1,\ldots,z_n\rangle$ and $w_1,\ldots,w_n$ be the corresponding finite sequence. It is clear that for each i, $y_i = w_i$ if and only if $x_1 = z_1 \wedge \ldots \wedge x_i = z_i$. Hence $x_n = z_n \leftrightarrow x = y$. Thus we define $G_n(x) = y_n$, where G_n maps sequences from D of length n into D. Then G_n is one-one. By the first claim, we have a one-one function $H:\omega xD \rightarrow D$. So define $G:FS(D) \rightarrow D$ by $G(x) = H(n,G_n(x))$, where x has length n. QED

We refer to a one-one function $F:DxD \rightarrow D$ as a pairing function.

LEMMA 3.7. The following is provable in BST'. Let D be a linearly ordered set with a pairing function and at least two elements. Let T be a set of sentences in PC(=) consistent with INF. Then T has a weak model whose domain is the set of closed terms in the extension of PC(=) by Skolem function symbols and the constants c_x, $x \in D$. The equality relation between closed terms s,t depends only on i) the symbols in s = t from left to right with the c_x all replaced by a common marker, and ii) the order type of the subscripts in the c_x's that appear in s = t from left to right.

Proof: Let < be a linear ordering of D and $F:DxD \to D$ be one-one. It will be convenient to make (D,<) into a dense linear ordering (D,<') without endpoints, by surrounding each point with a copy of the rationals. To make this construction in BST', first consider the lexicographic order on DxQ using the < on D. By Lemma 3.6, we can convert the domain to D.

Suppose T is a set of sentences consistent with INF. By standard model theory, doable in BST', T has a model M (with domain N) with countably infinitely many Skolem functions, where every element of the domain is generated by these Skolem functions over an infinite set of linearly ordered indiscernibles of order type Q.

Note that if a model satisfies every universal sentence that M satisfies, then that model satisfies T. This is a crucial fact about Skolemization.

We now construct a term model M* based on M, which will be seen to satisfy every universal sentence that M satisfies.

Introduce the new constants c_x for each $x \in D$. We define a structure M* whose domain S consists of the closed terms in these constants and the constant and function symbols of M. We will use the linear ordering of D, which linearly orders the subscripts of the new constants.

First let c be a constant in L(M). The interpretation of c in M* is the closed term c, which is an element of S.

Now let F be a k-ary function symbol of M. Let $t_1,\ldots,t_k \in S$. Take $F^*(t_1,\ldots,t_k)$ to be the closed term $F(t_1,\ldots,t_k)$.

Finally, let R be a k-ary relation symbol of M. We compute the truth value of $R^*(t_1,\ldots,t_k)$ by first making any order preserving assignment of indiscernibles in M to the new constants appearing in the $t_1,\ldots,t_k$. This results in terms $t_1',\ldots,t_k'$ with parameters from the indiscernibles in M. The truth value of $R^*(t_1,\ldots,t_k)$ is taken to be the truth value of $R(t_1',\ldots,t_k')$ in M. Because of indiscernibility in M, this calculation is independent of the choice of the order preserving assignment of indiscernibles to the new constants appearing in the $t_1,\ldots,t_k$.

The above paragraph is also done for the special 2-ary relation symbol =. Note that the interpretation in M* of = is a binary relation, but not necessarily equality.

Suppose that a sentence □($t_1,...,t_n$) holds in M*, where □ is a quantifier free formula and $t_1,...,t_n$ □ S. By making an order preserving assignment of indiscernibles in M to the new constants appearing in $t_1,...,t_n$, we see that □($t_1',...,t_n'$) holds in M*. This is because the truth values of the constituent atomic subformulas of □($t_1,...,t_n$) are preserved. This establishes that every universal sentence true in M remains true in M*. By the Skolemization, we now see that M and M* satisfy the same sentences of PC(=), and hence M* satisfies T. QED

We will refer to the equivalence relation given by Lemma 3.7 that is the interpretation of = in M* by ≡.

At this point, it is standard to simply take M* factored out by ≡ in Lemma 3.7 in order to get a model of T. However, we want to stay within BST' and be much more explicit. We also have an eye to dealing with the universal domain later, in which case also factoring out by the equivalence relation is not acceptable.

What we will use is a canonical presentation of definable relations in any dense linear ordering without endpoints. This is of some independent interest.

Let (D,<) be a dense linear ordering without endpoints.

Let □($x_1,...,x_k,b_1,...,b_p$) be a formula in <,=, where all free variables are shown, and there are parameters $b_1 < ... < b_p$ from D. We view □ as defining a k-ary relation on D carved out by the distinct variables $x_1,...,x_k$.

LEMMA 3.8. Suppose there exists $b_1' < b_2$ such that □($x_1,...,x_k,b_1,...,b_p$) and □($x_1,...,x_k,b_1',b_2,...,b_p$) define the same relation. Then for all $b_1' < b_2$, □($x_1,...,x_k,b_1,...,b_p$) and □($x_1,...,x_k,b_1',b_2,...,b_p$) define the same relation.

Proof: Assume hypotheses. This states a definable property of $b_1,...,b_p$ that is shared by $b_1',b_2,...,b_p$, and the definable property has no parameters. By quantifier elimination, falling under any given definable property without parameters is just a matter of order type. QED

LEMMA 3.9. Suppose that one of the parameters $b_1,\ldots,b_p$ can be moved, where the parameters still are strictly increasing, so that $\square(x_1,\ldots,x_k,b_1,\ldots,b_p)$ still defines the same relation, as in Lemma 1 for b_1. Then that parameter can be so moved to any alternative spot with the same result, as in Lemma 1.

Proof: Same as for Lemma 3.8. QED

LEMMA 3.10. Assume the hypotheses of Lemma 3.0. Then there is a formula $\square(x_1,\ldots,x_k,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_p)$ that defines the same relation as $\square(x_1,\ldots,x_k,b_1,\ldots,b_p)$.

Proof: Let $\square(x_1,\ldots,x_k,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_p)$ be $(\square b_i)(\square(x_1,\ldots,x_k,b_1,\ldots,b_p) \;\square\; b_{i-1} < b_i < b_{i+1})$. QED

LEMMA 3.11. Let $\square(x_1,\ldots,x_k,b_1,\ldots,b_p)$ and $\square(x_1,\ldots,x_k,c_1,\ldots,c_p)$ define the same relation, $b_1 < \ldots < b_p$, and $c_1 < \ldots < c_p$, and $\{b_1,\ldots,b_p\} \neq \{c_1,\ldots,c_p\}$. Then there exists $\square(x_1,\ldots,x_k,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_p)$, some i, or there exists $\square(x_1,\ldots,x_k,c_1,\ldots,c_{j-1},c_{j+1},\ldots,c_p)$, some i, that defines the same relation.

Proof: The hypothesis is a definable property of $(b_1,\ldots,b_p,c_1,\ldots,c_p)$, without parameters. Let i be least such that $b_i \neq c_i$. If $b_i < c_i$. Then we can move b_i around so that Lemma 3.10 holds. If $c_i < b_i$ then we can move c_i around so that Lemma 3.10 holds. QED

Here is the model theoretic fact of some independent interest.

LEMMA 3.12. Let R be a k-ary definable relation in (D,<) which can be defined by a formula with p parameters and no fewer. Then all formulas with p parameters that define R have exactly the same parameters.

Proof: This is immediate from Lemma 3.11. QED

We say that D is logically normal if and only if for any set T of sentences in PC(=), T is D-satisfiable if and only if T + INF is consistent.

THEOREM 3.13. The following is provable in BST'. D is logically normal if and only if i) D has at least two elements; ii) D has a linear ordering; and iii) there is a one-one function from DxD into D.

Proof: Since "R is a linear ordering", "binary F is one-one", and "there are at least two elements" are each consistent with INF, we see that if D is logically normal then i), ii), iii) hold. This establishes the forward direction.

For the reverse direction, assume i) - iii) and apply Lemma 3.7. We obtain a model M* of T whose domain is the set of closed terms in the extension of PC(=) by infinitely many Skolem function symbols and the constants c_x, $x \in D$. In addition, we have the condition expressed in the last sentence of Lemma 3.7.

Normally, we complete the proof by factoring out by this equivalence relation.

We would like to avoid doing this, as we will be interested in carrying out this argument for domains of unbounded extent, like the universal domain. The problem here is that the ensemble of equivalence classes moves us up a type.

We wish to develop a suitable canonical presentation of each equivalence class under the equality relation among closed terms of M*.

To be more concrete, this means that we wish to develop a suitable function J on the set of closed terms such that for all closed terms s,t, we have

$$s \equiv t \text{ if and only if } J(s) = J(t).$$

For any closed term s we write s# for the result of replacing all new constants in s by the common marker #.

Let s be a term. Let □ be the least t# such that s ≡ t, where we use some standard indexing of all finite sequences of the relevant symbols.

Let V be the set of all closed terms t such that s# = t#. Among $t \in V$, s ≡ t is determined solely by the order type of the subscripts of the new constants that appears in the expression s = t. Therefore the set □(s) of all sequences of subscripts of new constants in $t \in V$ with s ≡ t constitutes a definable set in (D,<) with parameters.

By Lemma 3.12, we know that there is a unique set □(s) □ D which serves as parameters for a definition in (D,<) of □(s), such that no proper subset so serves.

We then take J(s) to be the definition of □(s) in (D,<) which uses exactly the parameters □(s) and which is least in some standard indexing of definitions.

If s ≡ s' then it is easy to see that □(s') = □(s), □(s') = □(s), □(s') = □(s), and J(s') = J(s).

Note that the values of J are finite sequences from D □. □. Also note that the new constants lie in separate equivalence classes; i.e, $J(c_x) = J(c_y)$ □ x = y. Hence the values of J include D. So we can construe the resulting model as having domain D, as required. QED

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 8
11/20/02

Last time we gave some background about general domains, and showed that the following conditions on a domain D determine which sets of sentences in PC(=) are satisfiable with domain D (i.e., D-satisfiable):

D has at least two elements, has a linear ordering, and has a pairing function.

In fact, the sets of sentences of PC(=) that are D-satisfiable are exactly those that are consistent with INF.

1. Basic theory of predication.

We now want to get clear about the basic metatheory that we will be using in order to discuss predication on the universal domain W. We first present the theory BTP = basic theory of predicates.

BTP is a two sorted theory based on objects and unary predicates on the objects. The vocabulary of BTP consists of

1) object variables x_n, $n \geq 1$;
2) predicate variables $_{Pn,}$ $n \geq 1$, ranging over unary predicates on objects;

3) constant object 0;
4) binary function symbol < > from objects to objects;
5) equality = between objects.
6) the usual connectives and quantifiers, as well as commas and parentheses.

The object terms of BTP are given by

7) every object variable is an object term;
8) 0 is an object term;
9) if s,t are object terms then <s,t> is an object term.

The atomic formulas of BTP are
10) s = t, where s,t are object terms;
11) P(s), where P is a predicate variable and s is an object term.

The formulas of BTP are:

12) atomic formulas of BTP are formulas of BTP;
13) if φ,ψ are formulas of BTP then $(\neg\varphi)$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, $(\varphi \leftrightarrow \psi)$ are formulas of BTP;
14) if φ is a formula of BTP, x is an object variable and P is a predicate variable, then $(\forall x)(\varphi)$, $(\exists x)(\varphi)$, $(\forall P)(\varphi)$, $(\exists P)(\varphi)$ are formulas of BTP.

We refer to the language of BTP as L(BTP).

The axioms and rules of BTP are

15) the usual axioms and rules for predicate calculus based on the language of BTP;
16) Pairing. $\langle x,y\rangle = \langle z,w\rangle \rightarrow (x = z \wedge y = w)$;
17) Zero. $\neg\langle x,y\rangle = 0$;
18) Strict Comprehension. $(\exists P)(\forall x)(P(x) \leftrightarrow \varphi)$, where φ is a formula in L(BTP) with at most the free variable x.

We think of the object variables as ranging over absolutely everything. 0 can be taken to be the number zero, or if that is considered vague/meaningless, take it to be your favorite object, or even "the idea of the universal domain". Take <x,y> to be the ordered pair of x,y, or if you prefer, the idea of: x followed by y. Take the predicates to be predicates that sensibly applied to absolutely everything. BTP is based on the idea that there is a concept of predication that is being used which can be referred to when designating predicates. This is reflected in the use of quantifiers over predicates in Pure Comprehension. We use the word "Pure" to indicate that no free variables other than x are allowed in φ.

We shall see that BTP is already a rather strong system in that it is mutually interpretable with the formal system Z2 of f.o.m. In particular, one can faithfully develop arithmetic, and arithmetic mathematics. One can go quite a bit further, in some reasonable senses.

2. Identity of indiscernibles, singleton extension principle.

We now take up the issue of the status of the identity of indiscernibles. This is the principle

$$\text{IIS)}\ (\forall x,y)(x = y \leftrightarrow (\forall P)(P(x) \leftrightarrow P(y))).$$

THEOREM 2.1. BTP neither proves nor refutes IIS.

Proof: Let D be the set of all closed terms in the language 0,1,2,< >. There is a unique automorphism of (D,0,< >) that sends 1 to 2 and 2 to 1. Take the predicates to be all subsets of D that are fixed under this automorphism. I.e., the forward image under this automorphism is itself. QED

In fact, this proof will work for a rather innocent looking extension of BTP, which we write as BTP'. Here Strict Comprehension is extended as follows:

Pure Comprehension. $(\exists P)(\forall x)(P(x) \leftrightarrow \varphi)$, where φ is a formula in L(BTP) in which P is not free and where all free object variables in φ are x.

THEOREM 2.2. BTP' neither proves nor refutes IIS.

Now that we have introduced BTP', we can ask whether it follows from BTP.

THEOREM 2.3. BTP does not prove BTP'.

Proof: Let D be the closed terms in 0,1,2,3,< >. Use as predicates the $P \subseteq D$ such that for some $1 \leq i \leq 4$, P is fixed under the six automorphisms of D that fix i and permute all the other three elements of {1,2,3,4}. This family of predicates is fixed under all 24 automorphisms of D in the appropriate sense. Hence this construction is a model of BTP. If the model satisfies BTP', then {1,2} exists. However, {1,2} does not have the property that for some $1 \leq i \leq 4$, {1,2} is fixed under the automorphisms of D that fix i and permute all of the other elements of {1,2,3,4}. QED

The obvious axiom to add to BTP that will allow us to prove IIS is the following principle of Singleton Extensions:

$$\text{SEP)}\ (\forall x)(\exists P)(\forall y)(P(y) \leftrightarrow y = x).$$

THEOREM 2.4. BTP + SEP proves IIS.

Proof: Let x,y be such that $(\forall P)(P(x) \rightarrow P(y))$. By SEP let P be such that $(\forall z)(P(z) \leftrightarrow z = x)$. Then P(x). Hence P(y), and so y = x. QED

Is BTP' + SEP consistent? Of course. Let D be the set of all closed terms in 0,< >, and take the predicates to be all subsets of D.

But does SEP follow from IIS? No, not even over BTP'.

THEOREM 2.5. BTP' + IIS does not prove SEP.

Proof: Probably this can be done without heavy duty machinery, but here goes. Let M be a countable transitive model of ZFC + V = L. Let M[x] be a generic extension of M by adding a Cohen subset $x \subseteq \omega$. Now let D be $S(\omega)$ in the sense of M[x] and the predicates on D be the ordinal definable subsets of $S(\omega)$ in the sense of M[x]. Let 0 be ∅ and let < > be defined in the obvious way use a standard pairing function on ω. Then this defines a model of BTP' since any set given by pure comprehension must be an ordinal definable subset of $S(\omega)$ in the sense of M[x]. However SEP is clearly false in this model since {x} is not ordinal definable in the sense of M[x]. However IIS holds since for any $x,y \in D$, there exists n such that $n \in x \leftrightarrow n \notin y$. Let $n \in x \backslash y$. Then $x \in \{z: n \in z\}$ and $y \notin \{y: n \in y\}$. The other case is handled symmetrically. QED

THEOREM 2.6. BTP + SEP does not prove BTP'.

Proof: Again, probably this can be done without heavy duty machinery, but here goes. We start with a countable transitive model M of ZFC + V = L. We introduce in the usual way, the generic set $\{\{x_{11},x_{12},...\},\{x_{21},x_{22},...\},...\}$, where the x's are mutually Cohen generic subsets of □. We take D to be S(□) in the sense of M' = $M[\{\{x_{11},x_{12},...\},\{x_{21},x_{22},...\},...\}]$ = M[E], with 0,< > interpreted as before. The predicates are those subsets of D such that for some u □ E and x □ □, the subset is ordinal definable from E,u,x. Then BTP + SEP can be verified to hold in this model. However, $\{x_{11},x_{12},...,x_{21},x_{22},...\}$ does not lie in this model, and so BTP' fails. QED

There is a very natural interpretation of BTP + SEP in BTP. This is also an interpretation of BTP' + SEP in BTP'. This is called the cross sectional interpretation.

Under this interpretation, the object sort remains fixed, but the predicates are enlarged. The predicates are now pairs P,x, where P is a predicate and x is an object. We define (P,x)(y) if and only if P(<x,y>).

THEOREM 2.7. Under the cross sectional interpretation, BTP + SEP is interpreted in BTP. Also BTP' + SEP is interpreted in BTP'.

Proof: We first prove the cross sectional interpretation of SEP in BTP. Fix x. We need to find P,y such that for all z, (P,y)(z) □ z = x. We will use y = x. So we need to find P such that for all z, P(<x,z>) □ z = x. Take P such that for all w, P(w) □ (□u)(w = <u,u>). This is provided by BTP.

We next verify the cross sectional interpretation of BTP in BTP. We must verify (□P)(□x)(P(x) □ □), where □ has at most the free variable x.

Fix □ with at most the free variable x. Let the cross sectional interpretation of (□P)(□x)(P(x) □ □) be (□P,y)(□x)(P(<y,x> □ □'), where □' has at most the free variables x,y. Choose P such that (□z)(P(z) □ (□x)(z = <0,x> □ □')). This P is provided by BTP.

For the second claim, we need to verify the cross sectional interpretation of BTP' in BTP'. Let the interpretation of (□P)(□x)(P(x) □ □) be (□P,y)(□x)(P(<y,x>) □ □'), where □' has at most the free object variables x,y, and may have various free predicate variables other than P.

It clarifies matters to start with the sentence

$$(□R_1,...,R_k)(□P)(□x)(P(x) □ □)$$

whose interpretation is

$$(\forall R_1,x_1)\ldots(\forall R_k,x_k)(\exists P,y)(\forall x)(P(<y,x>) \leftrightarrow \varphi')$$

where φ' has at most the free variables $R_1,\ldots,R_k,x,x_1,\ldots,x_k$.

Let

$$\varphi'' = (\exists x_1,\ldots,x_k)(y = <x_1,\ldots,x_k> \wedge \varphi').$$

Using BTP', fix P such that

$$(\forall x,y)(P(<y,x>) \leftrightarrow \varphi'').$$

Now let $R_1,\ldots,R_k,x_1,\ldots,x_k$ be given. Let $y = <x_1,\ldots,x_k>$. Then

$$(\forall x,y)(P(<y,x>) \leftrightarrow (\exists x_1,\ldots,x_k)(y = <x_1,\ldots,x_k> \wedge \varphi'))$$

and so

$$(\forall x,y)(P(<y,x>) \leftrightarrow \varphi')$$

as required. QED

We now consider the strongest form of comprehension.

General Comprehension. $(\exists P)(\forall x)(P(x) \leftrightarrow \varphi)$, where φ is a formula of L(BTP) in which P is not free.

THEOREM 2.8. BTP' + SEP is equivalent to BTP with General Comprehension.

Proof: The free object quantifiers in φ can be replaced by predicate variables because of SEP. QED

3. Pure and general predication. BTPpg.

The upshot of the previous section is the emergence of two kinds of comprehension, one weak and one strong. The difference is what kinds of free variables are allowed. In the weakest, no free variables other than the comprehending object variable is allowed. In the strongest, any free variables are allowed except the beginning existential predicate variable.

The proper way of sorting this out is to distinguish between two kinds of predication.

We call the first, pure predication. Here we think of predication in some sort of language, where we are not entitled to simply refer to any object as we create, discover, or contemplate, a pure predicate. If we do want to refer to an object, we must define it: i.e., pick it out uniquely with a predicate, and this predicate is also subject to the same constraints.

We call the second, general predication. This is the one that is most directly relevant to set theory and mathematics. Here we can freely use any objects as parameters. Infinitely many objects can also be used as parameters provided that they together form an object. However, if they, together, form an object, then we don't need to consider infinitely many objects for this purpose.

Of course, this discussion is a bit murky, but it is readily backed up by very satisfying formalisms.

The pure predication concept is formalized by our previous BTP' with its Pure Comprehension. The general predication concept is formalized by General Comprehension. We are not asserting any kind of complete formalizations. Just that these formalisms represent the principal points and distinctions.

Both of these notions seem fundamental. It can be argued, however, that pure predication is most fundamental, and general predication is a derived notion. How derived? By cross sections. I.e., a general predicate P is always given by a pure predicate R together with an object x, where

$$(\forall y)(P(y) \leftrightarrow R(\langle x,y \rangle)).$$

At the other extreme might be a view that only general predication, steeped in great history from f.o.m., set theory, and mathematics, is coherent, and pure predication is not.

We have decided to incorporate both notions into a single theory, BTPpg. Here p is for pure and g is for general.

A number of questions arise when we consider BTPpg. Among them is the question of the relationship between the pure predicates and the general predicates. E.g., is it necessarily the case that every general predicate is a cross section of a pure predicate?

BTPpg is a three sorted theory based on objects, pure (unary) predicates on objects, and general (unary) predicates on objects. The vocabulary of BTPpg consists of

1) object variables x_n, $n \geq 1$;
2) pure predicate variables $P^p{}_n$, $n \geq 1$, ranging over pure predicates on objects;
3) general predicate variables $P^g{}_n$, $n \geq 1$, ranging over general predicates on objects;
4) constant object 0;
5) binary function symbol < > from objects to objects;
6) equality = between objects.
7) the usual connectives and quantifiers, as well as commas and parentheses.

The object terms of BTPpg are given by

8) every object variable is an object term;
9) 0 is an object term;
10) if s,t are object terms then <s,t> is an object term.

The atomic formulas of BTPpg are

11) s = t, where s,t are object terms;
12) P(s), where P is a pure or general predicate variable, and s is an object term.

The formulas of BTPpg are given by

13) atomic formulas of BTPpg are formulas of BTPpg;
14) if φ,ψ are formulas of BTPpg then $(\neg\varphi)$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, $(\varphi \leftrightarrow \psi)$ are formulas of BTPpg;
15) if φ is a formula of BTPpg and x is an object variable and P is a pure or general predicate variable, then $(\forall x)(\varphi)$, $(\exists x)(\varphi)$, $(\forall P)(\varphi)$, $(\exists P)(\varphi)$ are formulas of BTPpg.

We refer to the language of BTPpg as L(BTPpg).

The axioms and rules of BTPpg are

16) the usual axioms and rules for predicate calculus based on the language of BTPpg;

17) Pairing. <x,y> = <z,w> □ (x = z □ y = w);
18) Zero. □<x,y> = 0;
19) Pure Comprehension. (□P^p)(□x)(P^p(x) □ □), where □ is a formula in L(BTPpg) all of whose free variables are either pure predicate variables other than P^p or x;
20) General Comprehension. (□P^g)(□x)(P^g(x) □ □), where □ is a formula in L(BTPpg) in which P^g is not free.

We also have the two important fragments BTPp and BTPg. The first is the fragment in which no general predicate variables are allowed. The second is the fragment in which no pure predicate variables are allowed.

THEOREM 3.1. Every theorem of BTPpg without general predicate variables is a theorem of BTPp. Every theorem of BTPpg without pure predicate variables is a theorem of BTPg.

Proof: We have essentially proved this already. Suppose we have a model of BTPg. Then we can extend it to a model of BTPpg without changing the objects and general predicates by making the pure predicates the same as the general predicates. Alternatively, suppose we have a model of BTPp. Then we can extend it to a model of BTPpg without changing the objects and pure predicates by making the general predicates the cross sections of the pure predicates. QED

We define SEP and IIS for BTPpg in the same way as we did before, of course referring only to pure predicates. For general predicates, SEP and IIS are trivially true.

THEOREM 3.2. BTPpg proves SEP □ IIS, and does not prove IIS □ SEP, and does not prove IIS.

Proof: Immediate from Lemmas 2.2, 2.4, 2.5, and Theorem 3.1. QED

Note that BTPpg proves that every pure predicate has the same extension as some general predicate, using General Comprehension.

4. Mathematical development of BTPpg.

We can develop a pure predicate of natural numbers in BTPpg as follows. Consider the following condition on general predicates:

$$P^g(0) \;\square\; (\square x)(P^g(x) \;\square\; P^g(<x,0>)).$$

We let $\square^p$ be the intersection of all general predicates satisfying this condition. The general predicate that holds of everything satisfies this condition.

We claim that the induction principle holds:

$$(P^g(0) \;\square\; (\square x)(P^g(x) \;\square\; P^g(<x,0>))) \;\square\; (\square^p(x) \;\square\; Pg(x)).$$

To see this, let P^g obey the left side. By construction, $\square^p$ is contained in P^g.

As a consequence, we see that we have induction on $\square^p$ with respect to all formulas in L(BTPpg) since we have full comprehension for general predicates in BTPpg.

Similar ideas allow us to develop definition by recursion on $\square^p$. We have recursion resulting in general predicates, where the recursion data is in terms of general predicates. Alternatively, we have recursion resulting in pure predicates, where the recursion data is in terms of pure predicates.

In particular, we have $\square^p,0,1,+,\bullet,<,=$ as pure predicates, and so arithmetic is developed "purely". In particular, we have the successor axioms.

We can also develop finite sequences of objects purely. Consider the following condition on general predicates:

$$P^g(0) \;\square\; (\square x,y)(P^g(0) \;\square\; (P^g(x) \;\square\; P^g(<x,y>))).$$

The idea is that 0 counts as the empty sequence.

We let FS^p be the intersection of all P^g satisfying the above condition. This serves as the definition of finite sequence.

We can develop the length function from the extension of FSp to the extension of $\square^p$. This will be a pure function.

We can go on to develop the satisfaction relation for any structure whose domain is W. If the structure is a pure structure, then the satisfaction relation will be pure. If the structure is a general structure, then the satisfaction relation will be general. The underlying syntax used for this purpose is of course pure.

An important theorem of BTPpg is the Schroeder Bernstein theorem. We take it in the following forms:

THEOREM 4.1. The following is provable in BTPpg. Let $f:A \to B$ and $g:B \to A$ be pure one-one functions, where A,B are pure sets (predicates). Then there is a pure one-one onto function $h:A \to B$. The same holds with "pure" replaced by "general".

The idea here is to imitate the usual proof of the Schroeder Bernstein theorem with its analysis of the orbits of points.

5. Pure principle of symmetric arguments.

We now introduce the pure principle of symmetric arguments to BTPpg.

PPSA. $(\forall P^p)(\forall x_1 \neq \ldots \neq x_k)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i^k \leq k$ of $P^p(<x_1,\ldots,x_k>) \leftrightarrow P^p(<x_{i1},\ldots,x_{ik}>)$, where $k \geq 1$.

We break this up with two parameters. Let $k,r \geq 1$.

PPSA(k,r). $(\forall P^p)(\forall x_1 \neq \ldots \neq x_r)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq r$ of $P^p(<x_1,\ldots,x_k>) \leftrightarrow P^p(<x_{i1},\ldots,x_{ik}>)$.

In particular, PPSA(2,2) asserts

$$(\forall P^p)(\forall x \neq y)(P(<x,y>) \leftrightarrow P(<y,x>)).$$

We think of PPSA(k,r) as asserting that any pure predicate of k variables is symmetric on r distinct arguments.

THEOREM 4.1. For all $r \geq 1$, BTPp proves PPSA(1,r).

THEOREM 4.2. BTPp does not prove PPSA(2,2).

Proof: To see that BTPp does not prove PPSA(2,2), take D to be the closed terms in 0,< >, and the predicates to be all subsets of D. It suffices to see that in this model, there is a linear ordering of D. QED

THEOREM 4.3. BTPp + PPSA is consistent.

Proof: Let D be the set of closed terms in 0,< > and constants $c_1,c_2,\ldots$. Let the predicates be all subsets of D that are fixed under all automorphisms of D. These automorphisms are given by permutations of the c's. Obviously PPSA holds in this model. QED

Note that we can state PPSA(k,r) with parameters k,r, in BTPp.

THEOREM 4.4. The following is provable in BTPp. If $k \leq k'$ and $r \leq r'$ then $\mathrm{PPSA}(k',r') \rightarrow \mathrm{PPSA}(k,r)$.

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
LECTURE 9
11/27/02

We continue with a discussion of the principle of symmetric arguments.

1. Pure principles of symmetric arguments.

Recall the pure principle of symmetric arguments:

PPSA. $(\forall P^p)(\forall x_1 \neq \ldots \neq x_k)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq k$ of $P^p(<x_1,\ldots,x_k>) \leftrightarrow P^p(<x_{i1},\ldots,x_{ik}>)$, where $k \geq 1$.

PPSA(k,r). $(\forall P^p)(\forall x_1 \neq \ldots \neq x_r)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq r$ of $P^p(<x_1,\ldots,x_k>) \leftrightarrow P^p(<x_{i1},\ldots,x_{ik}>)$.

In particular, PPSA(2,2) asserts

$$(\forall P^p)(\forall x \neq y)(P^p(<x,y>) \leftrightarrow P^p(<y,x>)).$$

We think of PPSA(k,r) as asserting that any pure predicate of k variables is symmetric on r distinct arguments.

Recall the following from last time:

THEOREM 1.1. For all $r \geq 1$, BTPp proves PPSA(1,r).

THEOREM 1.2. BTPp does not prove PPSA(2,2).

THEOREM 1.3. BTPp + PPSA is consistent.

THEOREM 1.4. The following is provable in BTPp. If $k \leq k'$

and r □ r' then PPSA(k',r') □ PPSA(k,r).

We also consider the multiple forms of PPSA.

MPPSA(k,r). (□P_1^p),...,P_n^p)(□$x_1 \neq \ldots \neq x_r$)(the conjunction over all 1 □ $i_1 \neq \ldots \neq i_k$ □ r and 1 □ j □ n of $P_j^p(<x_1,\ldots,x_k)$ □ $P_j^p(<x_{i1},\ldots,x_{ik}>)$, where $k \geq 1$.

THEOREM 1.5. MPPSA(k,r) follows from PPSA(k,r).

Proof: Let $P_1,\ldots,P_n$ be given. We want to get r indiscernibles for $P_1,\ldots,P_n$ just using PPSA(k,r).

Let $P(x_1,\ldots,x_k)$ if and only if there exists 1 □ i □ n such that $x_1,\ldots,x_k$ are not indiscernibles for P_i, and for the least such i, $P_i(x_1,\ldots,x_k)$ holds.

Let $x_1,\ldots,x_k$ be indiscernibles for P. We assume that $x_1,\ldots,x_k$ are not indiscernibles for $P_1,\ldots,P_n$. I.e., there exists i such that $x_1,\ldots,x_k$ are not indiscernibles for $P_1,\ldots,P_n$.

case 1. $P(x_1,\ldots,x_n)$. Let i be least such that $x_1,\ldots,x_k$ are not indiscernibles for P_i. Then $P(x_1,\ldots,x_n)$. Let $y_1,\ldots,y_n$ be any permutation of $x_1,\ldots,x_n$. Then i is also least such that $y_1,\ldots.,y_k$ are not indiscernibles for P_i. Since $P(y_1,\ldots,y_n)$, we have $P_i(y_1,\ldots,y_n)$. This establishes that $x_1,\ldots,x_n$ are indiscernibles for P_i. This is a contradiction.

case 2. □$P(x_1,\ldots,x_n)$. We know that i exists, and so □$P_i(x_1,\ldots,x_n)$. Also for any permutation $y_1,\ldots,y_n$ of $x_1,\ldots,x_n$, by the same argument we have □$P_i(y_1,\ldots,y_n)$, since □$P(y_1,\ldots,y_n)$. This is also a contradiction. QED

We now consider the logical relationships between the PPSA(k,r).

THEOREM 1.6. Let $k \geq 2$. PPSA(k,k) does not imply PPSA(2,k+1) over BTPp.

Proof: The domain will be the closed terms in 0,< >,$c_1,\ldots,$ c_k. The pure predicates will be those subsets of the domain which are fixed under all automorphisms. The automorphisms are given by permutations of $c_1,\ldots,c_k$. We consider the following pure binary relations. Let s,t be terms in D. The syntax of s is defined to be the term with the subscripts removed from all c's. We assume an indexing of "syntaxes" by natural numbers.

Let s be given. #(s) is the number of occurrences of c's in s. A position is given a number $1 \leq i \leq \#(s)$. The occurrences of c's in s occupy positions 1,2,...,#(s).

A first position in s is a position whose constant differs from the constant at all earlier positions.

Let □(s) be the set of all first positions in s. Note that □(s) has at most k elements since there are only k constants $c_1,...,c_k$.

We assume an indexing of finite sets of natural numbers by natural numbers.

With these preliminaries, we are prepared to define some pure binary relations.

P(s,t) iff the index of the syntax of s is < the index of the syntax of t.

Q(s,t) iff the index of □(s) is < the index of □(t).

R(s,t) iff

i) $s \neq t$;
ii) s,t have the same syntax;
iii) let i be least such that the c at position i in s differs from the c at position i in t. Let c_p be at position i in s and c_q be at position i in t;
iv) the first position where c_p occurs in s is less than any position where c_q occurs in t.

Let $1 \leq i \leq k$. $S_i(s,t)$ iff the constant at position equaled to the i-th element of □(s) and the constant at position equaled to the i-th element of □(t) both exist and are unequal.

Now let $t_1 \neq ... \neq t_{k+1}$ be indiscernibles for $P,Q,R,S_1,...,S_k$.

Using P, we see that the t_i all have the same syntax. In particular, the sequence of c's appearing in the t_i all have the same length.

Using Q, we see that the □(t_i) are all the same.

Suppose $R(t_1,t_2)$. Then $R(t_2,t_1)$. Let i be least such that the c at position i in s differs from the c at position i in t. Let c_p be at position i in s and c_q be at position i in t. Then the first position where c_p occurs in s is < any position where c_q occurs in t. Also the first position where c_q occurs in t is < any position c_p occurs in t. This is a contradiction.

Hence not $R(t_1,t_2)$, not $R(t_2,t_1)$. Therefore the first position where c_p occurs in s is ≥ some position where c_q occurs in t, and the first position where c_q occurs in t is ≥ some position where c_p occurs in s. Hence the first position where c_p occurs in s is the same as the first position where c_q occurs in t.

In particular, there exists i such that the constant at position equaled to the i-th element of □(t_1) and the constant at position equaled to the i-th element of □(t_2) both exist and are unequal.

We now have $S_i(t_1,t_2)$, and there we have for all $1 \leq b \neq c \leq k+1$, $S_i(t_b,t_c)$. Hence the various constants that sit on the position equaled to the i-th element of the □(t)'s are all different. This is a contradiction since there are k c's and k+1 t's. So these k+1 indiscernibles cannot exist. QED

THEOREM 1.7. Let $k \geq 2$ and $n \geq k$. PPSA(k,n) does not imply PPSA(k+1,k+1) over BTPp.

Proof: Let the domain D be the set of all closed terms in 0,< >, $c_1,c_2,\ldots$. Construct a group G of permutations of $\{c_1,c_2,\ldots\}$ such that

i) every permutation of any k element subset of $\{c_1,c_2,\ldots\}$ can be extended to an element of G;
ii) if $f \in G$ maps E into E, where E is a k+1 element subset of $\{c_1,c_2,\ldots\}$, then f is the identity on E.

Now let the pure predicates be the subsets of D that are fixed under all elements of G. Then all k element sequences of distinct c's look alike.

Some work is needed to finish this proof by showing that PPSA(k+1,k+1) fails in this model. QED

Note that we have sort of encountered a yet stronger set of principles that makes sense in the pure case. Namely, what we call absolute indiscernibles.

UI(k,r). $(\forall x_1 \neq \ldots \neq x_r)(\forall P^p)$ (the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq k$ of $P^p(<x_1,\ldots,x_k>) \leftrightarrow P^p(<x_{i1},\ldots,x_{ik}>)$), where $k,r \geq 1$.

In Theorem 1.5, the model we construct obviously has UI(k,k), but not PPSA(2,k+1). In the partial proof of Theorem 1.6, we also obviously have UI(k,n) but presumably not PPSA(k+1,k+1).

THEOREM 1.8. Let $k \geq 2$. PPSA(k,k) does not imply UI(2,2) over BTPp.

Proof: Let D be the set of all closed terms in 0,< >,$c_1,c_2,\ldots$. Let the pure predicates be the sets such that for some n, that set is fixed under all automorphisms of D that fix $c_1,\ldots,c_n$. Then obviously MPPSA(k,k). However, UI(k,k) fails. To see this, obviously there is a pure predicate that holds of any given object and fails of any other given object. QED

2. Pure linear orderings.

Clearly PPSA(2,2) alone is enough to refute the existence of a linear ordering of W, over BTPp.

THEOREM 2.1. The nonexistence of a linear ordering is not sufficient to prove PPSA(2,2) over BTPp.

Proof: Let D be the set of all closed terms in 0,< >,$c_1,c_2,\ldots$. Construct a “generic” or “random” binary relation on the c’s such that for all $i \neq j$, $R(c_i,c_j) \leftrightarrow \neg R(c_j,c_i)$. This relation has plenty of automorphisms. Let G be the group of all of its automorphisms. Let the pure predicates be the sets that are fixed under all elements of G. To see that there is no linear ordering in this model, it suffices to show that there is no linear ordering on just the c’s that is fixed under all of the automorphisms in G. But there is an automorphism in G that fixes any finite set of c’s and moves any given remaining c into any other given remaining c. The automorphisms of a linear ordering cannot be that flexible. To see that PPSA(2,2) fails, consider the following relation R on terms s,t.

P(s,t) if and only if s,t have the same syntax and at the first place they differ and R holds at the constant in s at that position comma the constant in t at that position, or the syntax of s is less than the syntax of t.

Let s,t be indiscernibles for P.

case 1. s,t have the same syntax. Then clearly $P(s,t) \leftrightarrow \neg P(t,s)$, since the second disjunct does not have any effect.

case 2. s,t do not have the same syntax. Then clearly $P(s,t) \leftrightarrow \neg P(t,s)$, since the first disjunct does not have any effect.

In either case, we have a contradiction. QED

3. General principles of symmetric arguments; general linear orderings.

GPSA. $(\forall P^g)(\forall x_1 \neq \ldots \neq x_k)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq k$ of $P^g(<x_1,\ldots,x_k>) \leftrightarrow P^g(<x_{i1},\ldots,x_{ik}>)$, where $k \geq 1$.

GPSA(k,r). $(\forall P^g)(\forall x_1 \neq \ldots \neq x_r)$(the conjunction over all $1 \leq i_1 \neq \ldots \neq i_k \leq r$ of $P^g(<x_1,\ldots,x_k>) \leftrightarrow P^g(<x_{i1},\ldots,x_{ik}>)$.

In most cases, we never used anything that distinguishes the pure from the general case, and so all of the results go through virtually unmodified. The exception is the discussion of absolute indiscernibles. These are outright inconsistent if we use general predication.

THEOREM 3.1. GPSA does not follow from PPSA over BTPpg. In fact, it is consistent to assume each UI(k,k) and there is a general linear ordering, over BTPpg.

Proof: Let D be the set of all closed terms in 0,< >,$\{c_p: p \in Q\}$. Let the pure predicates be the sets that are fixed under all automorphisms. Let the general predicates be the sets such that for some finite subset of the c_p's, are fixed under all increasing order isomorphisms of Q into Q that fix the elements of that finite subset. The usual linear ordering of the c's will be a general predicate. However, the c_p's will form universal indiscernibles with respect to pure predicates. One has to verify BTPpg. QED

We say that a general (pure) predicate P is finite iff there is a finite sequence x (an object) such that

(□y)(P(y) □ y is a
term in x).

We say that a general (pure) predicate P is infinite iff it is not finite.

THEOREM 3.2. BTPpg proves that a general predicate is finite iff its extension is in general one-one correspondence with a proper initial segment of N. BTPpg does not prove all finite pure predicates are in pure one-one correspondence with a proper initial segment of N.

Proof: From a general one-one correspondence with a proper initial segment of N, one can get the required finite sequence by induction. And one can get the general one-one correspondence as a predicate of ordered pairs directly. In the pure case, let D be the set of closed terms in 0,< >,c_1,c_2. Let the pure predicates be the subsets of D that are fixed under both automorphisms. Let the general predicates be all subsets of D. Then {c_1,c_2} is pure but no two-tuple with both of them can be pure. QED

A minimally infinite general predicate is an infinite general predicate P such that no general predicate splits P. I.e., for any general predicate Q, either

i) there is a finite sequence x such that (□y)((P(y) □ Q(y)) y is a term in x); or
ii) there is a finite sequence x such that (□y)((P(y) □ □Q(y)) y is a term in x).

THEOREM 3.3. BTPg proves: if □ minimally infinite general predicate then GPSA. Converse not provable in BTPg.

Proof: The first claim is obvious. For the second claim, let D be the closed terms in 0,< >, c_1,c_2,... . Let the general predicates be the subsets of D that are fixed under all automorphisms that, for some i, fix all but the multiples of 2^i. Then GPSA holds, but there is no minimally infinite general predicate. QED

THEOREM 3.4. BTPg does not prove or refute □ a minimally infinite general predicate.

Proof: Let D be as usual. We can use all subsets of D, in which case it is false. Or we can use all subsets of D such that for some finite set of c’s, it is fixed under all automorphisms that fix each element of the finite set. Then the set of all c’s is a minimally infinite general predicate in this model. QED

An absolute POI (predicate of indisernibles) is a pure predicate P where any pure predicate holds or fails of any two equal length finite sequences of distinct objects from the extension of P.

THEOREM 3.5. BTPp does not prove or refute the existence of an infinite absolute POI.

Proof: Let D be the set of closed terms in 0,< >, $c_1,c_2,\ldots$. If we take the pure sets to be all subsets of D, then we don’t have an infinite absolute POI. On the other hand, if we take the pure predicates to be the sets fixed under all automorphisms, then $\{c_1,c_2,\ldots\}$ is a POI.

4. Satisfiability of universal sentences.

We now get back to logic. Recall the last substantial theorem about our main topic that we proved. In current terminology it is this.

We use Dp satisfiability for satisfiability in domain D with pure relations and functions. We use Dg satisfiability for satisfiability in domain D with general relations and functions. Recall that W is the universal domain.

THEOREM 4.1. BTPpg proves that the following are equivalent for any pure domain D.

i) the sets of sentences of PC(=) that are Dp satisfiable are exactly the sets of sentences of PC(=) that are Np satisfiable;
ii) the sets of sentences of PC(=) that are Dp satisfiable are exactly the sets of sentences of PC(=) that are consistent with INF(=);
iii) D has at least two elements, has a pure linear ordering, and has a pure pairing function.

The same result holds for general domains D and with “pure” replaced by “general” and with Dp replaced by Dg.

The extra trouble we went to in order to avoid factoring out by a congruence relation at the end of the proof allows us to conclude that the proof can be given in BTPpg.

By Wp satisfiable, we mean that we have a model with domain W and pure relations and functions. By Wg satisfiable, we mean that we have a model with domain W and general relations and functions.

SYM(=) consists of the following set of axioms. Let $k \geq 1$ and φ be a formula in PC(=) with free variables among $x_1,\dots,x_k$. We have the axiom

$$(\forall x_1 \neq \dots \neq x_k)(\text{conjunction of } (\varphi(x_1,\dots,x_k) \leftrightarrow \varphi(x_{\pi 1},\dots,x_{\pi k}))),$$

where the conjunction ranges over all permutations π of $1,\dots,k$.

We also consider the subsystem QFSYM(=) which is the same as SYM(=) except that φ is required to have no quantifiers.

In addition, we also consider the multiple forms of these two systems, MSYM(=) and MQFSYM(=). These are formulated in the obvious way using finitely many formulas (quantifier free formulas) φ at once.

LEMMA 4.2. SYM(=) and MSYM(=) are logically equivalent. QFSYM(=) and MQFSYM(=) are logically equivalent.

Proof: By adapting the proof of the equivalence of PPSA and MPPSA. QED

THEOREM 4.3. BTPp proves that every set of universal sentences of PC(=) consistent with QFSYM(=) is Wp satisfiable.

Proof: Let S be a set of universal sentences of PC(=) consistent with ATSYM(=). Let T be the following theory in PC(=) with new constant symbols $c_1,c_2,\dots$. The axioms of T are the sentences

$$\varphi(d_1,\dots,d_k) \leftrightarrow \varphi(e_1,\dots,e_k),$$

where $k \geq 1$, $d_1,\dots,d_k$ are distinct elements of $\{c_1,c_2,\dots\}$, $e_1,\dots,e_k$ are distinct elements of $\{c_1,c_2,\dots\}$, and φ is a quantifier free formula in PC(=) with the k free variables $x_1,\dots,x_k$.

We claim that S + T is consistent in PC(=) extended to accommodate the expanded language. To see this, it suffices to prove that every finite subset of S + T is consistent. Note that this follows immediately from the consistency of MQFSYM(=) + S by existentially quantifying out the relevant constants. The consistency of MQFSYM(=) + S is from Lemma 4.2.

Now let M be a model of S + T. Since the axioms of S + T are universal, we can assume that the domain D of M is generated from the new constants $c_1,c_2,...$ by the constant and function symbols of PC(=).

Working in BTPp, we now construct a model M* whose domain consists of the closed terms generated by the constants and functions of PC(=) and the additional constants c^*_x, for every object x. The interpretation of the constant and function symbols of PC(=) are obvious, and the interpretation of the relation symbols of PC(=) are given by reference to M. The relevant additional constants are mapped one-one back into the new constants $c_1,c_2,...$ and the truth value that results is imitated.

As in the proof of Theorem 4.1, we thus obtain a model whose equality relation is not identity. We avoid having to factor out by finding a canoncial "name" for each equivalence class just as in the proof of Theorem 4.1. In this case, we need only develop a canonical "name" for every definable relation in the set of additional constants c^*_x under equality. This is simpler, and follows from the development of canonical names in connection with Theorem 4.1, where we did this under a dense linear ordering without endpoints. As in Theorem 4.1, the proof is completed by intensive use of the Shroeder Bernstein theorem in BTPp. QED

Note that according to Theorem 4.3, we have a lower bound on the sets of universal sentences of PC(=) that are Wp satisfiable. If we accept PPSA, then our mission is accomplished for sets of universal sentences of PC(=). (At least under the pure interpretation).

THEOREM 4.4. The following are equivalent over BTPp.
i) The sets of universal sentences of PC(=) that are Wp satisfiable are exactly those that are consistent with SYM(=) (QFSYM(=), MSYM(=), MQFSYM(=));
ii) Every universal sentence of PC(=) that is Wp

satisfiable is consistent with SYM(=);

iii) The pure principle of symmetric arguments, PPSA, holds;

iv) The multiple pure principle of symmetric arguments, MPPSA, holds.

Furthermore, we can replace "purely" and "pure" throughout by "generally" and "general".

In i), we can use any choice of the four systems shown.

Proof: This follows from Theorem 4.3 together with the following observation. Under PPSA, every structure with the universal domain satisfies SYM(=). Thus we see that BTPp proves iii) implies i). To see that ii) implies iii) in BTPp, suppose ii) holds. Suppose PPSA fails. Then an instance of SYM(=) has a pure countermodel in the universal domain. Hence an instance of SYM(=) is not consistent with SYM(=). Hence SYM(=) is inconsistent. This is a contradiction, for we can take as a model of SYM(=), the natural numbers where the constants, relations, and functions are defined in a trivial way; e.g., the constants are all 0, the relations hold universally, and the functions are all first projection functions. We have already shown that iii) and iv) are provably equivalent in BTPp. QED

5. Provable Wp satisfiability.

This section concerns the determination of a modified notion of Wp satsifiability which does not require a principle like PPSA.

Let T be any extension of BTPp in its language.

We call a sentence □ of PC(=) provably Wp satisfiable over T if and only if T proves

□ is Wp satisfiable in the universal domain.

LEMMA 5.1. The following is provable in Peano arithmetic. Let □ be a universal sentence of PC(=). Then □ is consistent with SYM(=) if and only if □ has a model with exactly n elements other than the interpretation of constants in □, where □ is the relational type of □ and n is the highest arity of all relation symbols appearing in □, and any permutation of the domain that fixes the interpretation of the constants in □ is an automorphism of the model.

Proof: Suppose □ is consistent with SYM(=). In PA, we can build a model of SYM(=) + □ (with complete diagram), and cut down to the needed finite submodel with n elements other than the interpretation of constants in □. Symbols outside the relational type of □ are interpreted trivially.

On the other hand, suppose □ has a model as indicated. Then the model can be stretched to provide an infinite set of indiscernibles. Then SYM(□) is automatic. Again the symbols outside □ can be interpreted trivially.

THEOREM 5.2. Let □ be a universal sentence of PC(=). Then □ is provably Wp satisfiable over BTPp if and only if □ is consistent with SYM(=).

Proof: Let □ be as given. Suppose □ is provably Wp satisfiable over BTPp. If □ is not consistent with SYM(=), then □ can be refuted from SYM(=), in which case BTPp + PPSA proves that □ is not Wp satisfiable. In particular, BTPp + PPSA is inconsistent, which is a contradiction.

Suppose □ is consistent with SYM(=). By Lemma 5.1, BTPp proves that □ is consistent with SYM(=). By Theorem 4.3, □ is provably Wp satisfiable over BTPp. QED

THEOREM 5.3. Let □ be a universal sentence of PC(=). Then □ is provably Wp satisfiable over BTPp if and only if it is provable in Z_2 that □ is consistent with SYM(=). Also □ is provably Wp satisfiable over BTPp plus the true sentences of arithmetic if and only if □ is consistent with SYM(=).

PHILOSOPHY 532
PHILOSOPHICAL PROBLEMS IN LOGIC
SUPPLEMENARY NOTES
11/8/02
11/12/02

As indicated in Phil 532 Lecture 6, we will discuss subjective isomorphisms here. However, while developing this theory of subjective isomorphisms, we came across yet another approach that does not involve subjective concepts. In fact, this most recent approach seems perhaps the most promising of all. However, it may merge with the subjective isomorphisms approach at a later time.

Nevertheless, we think that the subjective isomorphisms idea has enough merit to warrant our discussing it in the first two sections of this supplement.

Then we move on to a newer topic: mereological foundations of set theory.

1. Subjective automorphisms in arithmetic.

Recall the system $I\Delta_0$ of bounded arithmetic. The language is $0,S,+,\bullet,<,=$. We have some basic axioms (e.g., Robinson's Q) together with induction for all Δ_0 or bounded formulas. I.e., the quantifiers are all bounded to variables. (We could bound the quantifiers to terms as well).

We expand the language of $I\Delta_0$ by adding a new sort for subjective unary functions from N into N. We assume standard axioms and rules of inference of logic for this language.

We use the word "subjective" because we are definitely not going to allow use of these functions in the induction scheme. This would lead to an immediate inconsistency with the other axioms we use.

We define "F is an automorphism" as

$$F(0) = 0 \wedge (\forall x,y)(F(S(x)) = S(F(x)) \wedge F(x+y) = F(x)+F(y) \wedge F(x\bullet y) = F(x)\bullet F(y) \wedge (x < y \leftrightarrow F(x) < F(y)) \wedge (x = y \leftrightarrow F(x) = F(y))).$$

Obviously there are no automorphisms other than the identity, by the obvious argument by induction. However, there are, arguably, subjective ones; i.e., outside the realm of objective thought. Or, putting it differently, outside the realm of mathematical thought. Another way of looking at this is that their are subjective automorphisms with a shift in point of view. I.e., a change in the reference frame. We are suggesting that their is a coherent relativistic foundation for mathematics. What happens to f.o.m. if we make moves analogous to the moves made in physics by Einstein's special and general relativity theories?

We now present the theory BA# (bounded arithmetic sharp).

1. Bounded arithmetic (i.e., $I\Delta_0$).

2. $(\forall x)(\forall y)(\exists F)(F$ is an automorphism $\wedge$ $(\forall z < x)(F(z) = y)$ $\wedge$ $(\exists z > y)(F(z) \neq z))$.

THEOREM 1.1. BA# proves every axiom of PA (Peano arithmetic). In fact, the theorems of BA# in the language of BA are exactly the theorems of PA. BA# and PA are equiconsistent.

Of course, one can read this result as a result in the model theory of arithmetic as follows.

THEOREM 1.2. Every model of BA which has automorphisms fixing any given bounded set, but whose fixed points are bounded, is a model of PA. In fact, any consistent extension of PA has a countable model with automorphisms whose set of fixed points are arbitrarily high proper initial segments.

For technical reasons, we will need EFA rather than BA. EFA = $I\Sigma 0(\exp)$, which is in the language 0,S,+,•,2x,<,=.

Let the language of EFA* be the language of EFA together with the unary function symbols F,G. We will not need variables over subjective functions. F,G are to represent two specific subjective functions. The axioms of EFA* are:

1. EFA.
2. F,G are automorphisms (with respect to the language of EFA).
3. F,G each fix exactly two different proper initial segments.

Let EFA* be the resulting formal system.

THEOREM 1.3. Every finite fragment of PA is interpretable in EFA*. EFA* and PA are equiconsistent. In fact, using some standard tricks, PA is interpretable in EFA#. These statements are provable in EFA.

Proof: We first give a construction of a model of PA as a specific initial segment of any model M of EFA*. We then indicate how to turn this proof into the desired interpretation.

Let M have automorphisms f,g, and let the fixed points of f be the proper initial segment I, and let the fixed points of g be the proper initial segment J. Assume that I $\supseteq\neq$ J.

It is clear that I,J are each closed under exponentiation. Let x □ I\J and let y = 2^2^x. Then y □ I\J.

Let E be the set of elements of dom(M) generated by f,g,f-1,g-1,y. Then every element of E is a double power of 2.

We claim that for all z □ E, f(z) ≠ z or g(z) ≠ z. We prove this by induction on z.

case 1. z = y. Then g(y) ≠ y since y □ J.

case 2. z = f(w), f(w) ≠ w.

case 3. z = f(w), g(w) ≠ w. If z □ I then y □ I. This is because I is closed under f.

case 4. z = f-1(w), f(w) ≠ w.

case 5. z = f-1(w), g(w) ≠ w.

case 6. z = g(w), g(w) ≠ w.

case 7. z = g(w), f(w) ≠ w.

case 8 z = g-1(w), g(w) ≠ w.

case 9 z = g-1(w), f(w) ≠ w.

Let z □ E. If f moves x then f(x) > x or f-1(x) > x. If g moves x then g(x) > x or g-1(x) > x. Hence E has no greatest element. Since every element of E is a double power of 2, E defines an initial segment K which is closed under 0,1,+,•,<. We now show that K satisfies PA.

We can view E as the set of all values at y of the group G of automorphisms of M generated by composition and inverse from f,g,d. Note that all such automorphisms are also automorphisms of M|K, where the structure M|K uses only 0,S,+,•,<.

Let z □ E. Then z = h(y) for some h □ G. Now consider the automorphism hfh^{-1}, which is also in G.

Note that h|M is an isomorphism from (K,f|K) onto (K,hfh-1|K). Since every element of [0,y] is a fixed point of f|K, we see that every element of [0,z] is a fixed point of hfh-

1|K. Also, since the fixed points of f|K form a proper initial segment of K, the fixed points of hfh-1|K form a proper initial segment of K.

We have thus shown that K has the following property. There are automorphisms of K whose fixed points are arbitrarily long proper initial segments of K.

This is enough to ensure that K satisfies PA. To see this, we use that K satisfies $I\Delta_0$ = bounded arithmetic. We will not need exponentiation on K.

Suppose K satisfies $I\Sigma_k$, $k \geq 0$. In order to establish $I\Sigma_{k+1}$ in K, it suffice to establish in K that

$$(\exists b)(\forall i \leq x)((\exists y)(\varphi(i,y)) \rightarrow (\exists y \leq b)(\varphi(i,y)))$$

for any Σ_k formula φ with parameters in K.

Let f be an automorphism of K that is the identity on exactly the proper initial segment I of K, where $x \in I$. Suppose $i \leq x$ and $(\exists y)(\varphi(i,y))$ holds in K. Then by the induction hypothesis, let y be least such that $\varphi(i,y)$ holds in K. Then y is definable in K from i, and so y is a fixed point of f. Hence $y \in I$.

We have thus established in K that

$$(\forall i \leq x)((\exists y)(\varphi(i,y)) \rightarrow (\exists y \in I)(\varphi(i,y)))$$

and so, letting $I \subseteq [0,b]$, we have in K,

$$(\forall i \leq x)((\exists y)(\varphi(i,y)) \rightarrow (\exists y \leq b)(\varphi(i,y))).$$

We have thus established that K satisfies PA.

We now indicate how to turn this proof into the desired interpretation of any finite fragment T of PA into EFA*.

The specific place requiring modification is in the construction of E and K. Not only do we have access to very little induction (EFA), we have absolutely no access to any kind of induction involving the automorphisms f,g and their sets of fixed points I,J. .... QED

2. Subjective automorphisms in set theory.

We begin with BST = bounded set theory. We will be minimalistic about things and use only $\in$, and no extensionality. The axioms are

1. Bounded separation. $(\exists x)(\forall y)(y \in x \leftrightarrow (y \in a \wedge \varphi))$, where φ is a bounded formula (i.e., all quantifiers are bounded).

We now formulate BST#. The language will be that of BST together with variables over subjective functions. We write F is an automorphism for

$$(\forall x,y)((x \in y \leftrightarrow F(x) \in F(y)) \wedge (x = y \leftrightarrow F(x) = F(y))).$$

The axioms of BST# are

1. BST.
2. $(\forall x)(\exists y)(\exists F)((\forall z \in x)(F(z) = z) \wedge (\exists z \in y)(F(z) \neq z))$.

THEOREM 2.1. ZFC is interpretable in BST#. BST# is interpretable in NBG (von Neumann Bernays Gödel class theory). ZFC and BST# are equiconsistent.

3. Mereological foundations.

We now shift attention to a new approach, independent at the moment, of subjective isomorphisms.

In pure mereology, one considers only parts of the whole, where the null part is banned. Thus all objects are "parts". One has only the part/whole relation and no other concepts. The part/whole relation is assumed to be transitive and reflexive.

In mereology, one can take equality as primitive or defined. If primitive, then one relies on the usual classical first order predicate calculus with equality. Otherwise it is defined as

$$x \equiv y \leftrightarrow (x \leq y \wedge y \leq x).$$

It is easily seen that the usual axioms for equality can be derived for this interpretation of equality, so the two approaches are in an appropriate sense equivalent. We choose to take = as primitive.

There are two kinds of axiomatizations of pure mereology.

One is the axioms corresponding to a Boolean algebra (without 0). The other is the least upper bound principle for all formulas in the language. This scheme is called fusion. The least upper bound is called the mereological sum. The important classical theorem is that these two axiomatizations are logically equivalent.

It is important to recall that there are some important additional axioms that make the system complete. An atom is defined to be a part which is a part of any of its subparts. If we specify axiomatically an exact finite number of atoms, and whether there exists a part with no part that is an atom, then we get a complete system. Alternatively, if we specify axiomatically that there are infinitely many atoms (with infinitely many axioms), and whether there exists a part with no part that is an atom, then we also get a complete system. Furthermore, these are the only complete extensions.

Now let us be more formal. Let M (pure mereology) be the following system in the language L(□,=), which is the ordinary classical first order predicate calculus with □ and equality. We assume the usual axioms and rules for this language.

M1. □ is reflexive, transitive, and has no minimum element.
M2. (x □ y □ y □ x) □ x = y.
M3. (□x)((□y)(□ □ y □ x) □ (□z)((□y)(□ □ y □ z) □ x □ z)), where □ is a formula of L(□) in which x,z are not free.

We now introduce a naming system to mereology. All names are to be atoms, but we do not require that every atom be a name. A name can only name one part. (Everything is a part).

Formally, we let MN be mereology with naming. The language is L(□,NA,=), where NA is a binary relation. We assume the standard axioms and rules of inference for classical predicate calculus based on this language. The intended of NA is NA(x,y) iff x is a name of y.

The axioms of MN are

1. □ is reflexive, transitive, and has no minimum element.
2. (x □ y □ y □ x) □ x = y.
3. (□x)((□y)(□ □ y □ x) □ (□z)((□y)(□ □ y □ z) □ x □ z)), where □ is a formula of L(□,NA) in which x,z are not free.

4. NA(x,y) □ x is an atom.
5. (NA(x,y) □ NA(x,z)) □ y = z.

We say that y is named if and only if there exists x such that NA(x,y). We say that x is an atomic part of y if and only if x is an atom and x is a part of y.

We now sketch a proof in MN that something is not named.

LEMMA 3.1. The following is provable in M. Every atomic part of a fusion is a part of the parts fused.

Proof: Let x be an atom that is part of a fusion, but not part of the parts fused. Then all of the things fused are part of the complement of x. Hence the fusion is a part of the complement of x. This contradicts that x is part of the fusion. QED

THEOREM 3.2. The following is provable in MN.
(□x)(□y)(□NA(y,x)).

Proof: Suppose every part has a name. Consider the fusion S of all names which are not a part of the part that it names. Let x be the name of S. First suppose that x is a part of S. Since x is an atom, it must be a part of one of the names which are not a part of the part that it names. Hence x is not a part of the part that it names. Hence x is not a part of S.

Finally suppose that x is not a part of S. Then x is not a part of the part that it names. Hence x is a part of S.

Both suppositions have led to a contradiction. QED

Thus we see that the refutation of "every part has a name" corresponds closely to Russell's and Tarski's paradoxes.

We now define the bounded formulas of L(□,NA,=) by the following inductive clauses.

i) every atomic formula is bounded.
ii) if □,□ are bounded then so are □□, □ □, □ □ □, □ □ □, □ □ □.
iii) if □ is bounded and y,z are distinct variables then (□y □ z)(□) and (□y □ z)(□) are bounded.

We are now ready to give the axiom scheme of "local completeness". This says that the naming system is "locally complete".

6. Local completeness. □ □ (□z1 □ y1)...(□zk □ yk)(□[y1,...,yk/z1,...,zk] □ z1,...,zk are named), where □ is a bounded formula of L(□,NA,=), y1,...,yk is an enumeration of the free variables of □, the 2k variables y1,...,yk,z1,...,zk are distinct, and z1,...,zk do not appear in □.

The resulting system 1-6 is MLCN (mereology with locally complete naming).

THEOREM 3.3. MLCN corresponds roughly to a second order indescribable cardinal. In particular, ZFC is interpretable in MLCN. We also have equiconsistency. In the appropriate sense, MLCN and ZFC plus roughly an indescribable cardinal prove the same arithmetic sentences and more.

We are working on a very striking stronger principle to replace 6 which would put us far beyond measurable cardinals.