INTERPRETATIONS, ACCORDING TO TARSKI
Harvey M. Friedman*
Nineteenth Annual Tarski Lectures
Interpretations of Set Theory in Discrete Mathematics and
Informal Thinking
Lecture 1
Delivered, April 9, 2007
Expanded May 24, 2007

Special thanks to Albert Visser for considerable input.

1. Interpretations.

The notion of interpretation was first carefully defined
and developed in the book [TMR53].

The notion of interpretation is absolutely fundamental to
mathematical logic and the foundations of mathematics. It
is also crucial for the foundations and philosophy of
science - although here some crucial conditions generally
need to be imposed; e.g., "the interpretation leaves the
mathematical concepts unchanged".

The most obvious and direct use of interpretations is for
relative consistency.

PROPOSTION 1.1. Suppose S is interpretable in T. If T is
consistent then S is consistent.

But the notion was mostly used by Tarski to show that
various formal systems are undecidable in the sense that
there is no algorithm for determining whether a sentence in
its language is provable.

PROPOSITION 1.2. Suppose S is interpretable in T. If T has
a consistent decidable extension in its own language, then
S has a consistent decidable extension in its own language.
I.e., if S is essentially undecidable then T is essentially
undecidable.

Rafael M. Robinson in [Ro52] set up a crucial very weak system of arithmetic, Q, and showed that Q is essentially undecidable.

Proposition 1.2 was successively applied, starting with Q, invoking transitivity: S interpretable in T, T interpretable in W, therefore S is interpretable in W.

We now define interpretations in the setting of ordinary first order predicate calculus with equality.

There are some choices along the road - the notion is still not entirely standard. We will work with the most liberal natural and convenient notion.

A theory T is a set of sentences together with a relational type, L(T). Here we won't demand that we close T under logical consequence.

We begin semiformally and semantically, with several details postponed.

π is an interpretation of S in T iff

i. π specifies a domain of objects from viewpoint of T. Given by formula(s) in L(T).

ii. π specifies an interpretation of all constant, relation, and function symbols of L(S), on the domain in i, by means of formulas in L(T).

iii. This data induces a map π from formulas of L(S) into formulas of L(T). We require that for all $\varphi \in S$, T $|-$ $\pi\varphi$.

We have hidden details. We take the most liberal natural view of what is allowed; i.e., "liberal" definability.

The specified domain D consists of tuples of any lengths ≥ 1. Mixed lengths are allowed. There must be a bound on the lengths used. D is to be given by a finite set of formulas in L(T), one for each length allowed. In every model of T, D is nonempty.

We have a binary relation E on D, given by a finite set of formulas in L(T), one for each pair of lengths allowed.

In every model of T, E is an equivalence relation on D. We have formulas in L(T) that interpret the constant, relation, function symbols of S, on D/E. We don't perform factoring, but instead treat the interpreted constants, relations, functions as relations on D that respect E, in all models of T.

We have lied a little: we will allow parameters. I.e., extra variables in the above, where we merely say that in any model of T, there exist choices for these variables (parameters) so that all of the above hold.

We are asserting that there is a way of defining a model of S inside any model of T, uniformly, multidimensionally, via equivalence, and with parameters. The definition is required to be uniform but the choice of parameters has no uniformity requirement. There may be many choices of parameters that work. We call this liberal definability.

This definition is particularly well behaved in the case where S is finite. To begin with, since S is finite, the data used and the requirements made in the definition are finite *before* we existentially quantify over choices of parameters. Therefore we are asserting that T logically implies a finite set of sentences. Hence we can replace these logical implications by provability in T. This gives us the *provability* version of (liberal) interpretability.

The following result shows a great deal of robustness. On the other hand, we can remove any requirements of uniformity.

THEOREM 1.3. A finite S is interpretable in T if and only if for all M |= T, some M' |= S is liberally defined in M.

Proof: The forward direction is obvious. For the reverse direction, we use compactness and completeness.

We claim that M' can be liberally defined in M, uniformly, up to a choice of parameters. I.e., there is a fixed presentation of L(S) data in L(T) such that for all M |= T, there is some choice of parameters from dom(M) such that the data constitutes a model M' |= S.

Suppose this is false. Then for any such presentation, there exists M |= T such that for all choices of parameters from dom(M), the presentation fails to define a model of S.

We then conclude that for any finite set of such presentations, there exists M |= T such that for at least one of these presentations, for all choices of parameters from dom(M), the presentation fails to define a model of S. For if we had a finite set of such presentations without this property, then we could get a single such presentation without this property through definition by finitely many cases.

Let X be the infinite set of sentences asserting T, and that for each of the countably many presentations, and all choices of parameters, the presentation fails to define a model of S. By hypothesis, we know that X is not satisfiable. Hence some finite subset of X is not satisfiable. This finite subset of X provides a counterexample to the previous paragraph. QED

The case of finite S,T (in finite relational types) is fundamental. We can create unimpeachable certificates of interpretability: the required definitions in L(T) together with the required proofs in T.

We will identify these certificates with their Gödel numbers when convenient. Thus we speak of one certificate being smaller than another if and only if the Gödel number of the first certificate is less than the Gödel number of the second certificate.

Unimpeachable certification is important for relative consistency, where issues of impeachability are crucial.

After Tarski, many of the general investigations into interpretability, such as in Feferman, concentrated on the non finitely axiomatized cases (reflexive theories). We touch on this case in section 6, and infinite theories also are used in section 7.

In the infinite case (where S is infinite), it is natural to disallow the use of parameters, and consider only parameterless interpretations. Obviously the semantic version of parameterless interpretations is equivalent, by the completeness theorem, to the provability version of parameterless interpretations. However, we will lose Theorem 1.3.

We now focus on the finite case. By taking conjunctions, we can assume that we have single sentences A,B., We will have a limited discussion of the infinite case in section 6. Also infinite theories are present in section 7.

2. Tarski Degrees.

S ≤I T means that the theory S is interpretable in the theory T. Define

S =I T ↔ S ≤I T ∧ S ≤I T.
S <I T ↔ S ≤I T ∧ ¬T =I S.

For the purposes of this definition, we view theories as sets of sentences without mentioning any underlying relational type. Underlying relational types play no significant role here.

Obviously ≤I is reflexive and transitive. The equivalence classes of theories, under the equivalence relation =I, are called the Tarski degrees.

It is easy to show that there is a proper class of Tarski degrees, corresponding to the fact that there are a proper class of theories.

If we restrict to countable theories, then there are 2^ω countable theories, and 2^ω associated Tarski degrees. If we restrict to finite theories in finite relational types, there are ω associated Tarski degrees. Obviously, restricting to finite theories is the same as restricting to single sentences.

The unary Tarski degrees are obtained by restricting ≤I to single sentences. I.e., they are the equivalence relations of single sentences under =I. There are ω unary Tarski degrees.

The unary Tarski degrees bear a rough resemblance to two very well studied countable structures from recursion theory – the Turing degrees (of subsets of ω) and the r.e. sets (subsets of ω). We know from experience that the latter two structures are very complicated in various ways. It is not clear how complicated the unary Tarski degrees are, and how these three countable structures are related.

THEOREM 2.1. The unary Tarski degrees form a (reflexive) partial ordering with a minimum and maximum element. The maximum element 1 is the equivalence class of all sentences with no models. The minimum element 0 is the equivalence class of all sentences with a finite model.

Proof: Obviously every A is interpretable in any B with no models. We claim that in any M, we can define all models of all nonzero finite cardinalities using a single parameter x and the multidimensional apparatus. I.e., we can use x, (x,x), (x,x,x), ..., (x,x,...,x) as the domain elements. Thus any A with a finite model is interpretable in any B. QED

THEOREM 2.2. The unary Tarski degrees form a distributive lattice with a minimum and maximum element.

Proof: Let A,B be sentences. Define inf([A],[B]) = [A ∨ B]. To see that this is the inf, we have to verify that C ≤I A,B ↔ C ≤I A ∨ B. The reverse direction is obvious. Now assume C ≤I A,B. Let M |= A ∨ B. If M |= A then there is a model of C that is liberally defined in M. If M |= B then there is a model of C that is liberally defined in M.

sup([A],[B]) requires more care. Introduce a new monadic predicate symbol R. Take

sup([A],[B]) = [(∃x,y)(Rx ∧ ¬Ry) ∧ A^R ∧ $B^{¬R}$].

To see that this is the sup, we have to verify that

A,B ≤I C ↔ (∃x,y)(Rx ∧ ¬Ry) ∧ A^R ∧ $B^{¬R}$ ≤I C.

The reverse direction is obvious. Now assume A,B ≤I C. Let M |= C. Let M′,M′′ be liberally defined in M, where M′ |= A, M′′|= B. Using multidimensionality, we can disjointify the domains of A,B in order to liberally define a model of (∃x,y)(Rx ∧ ¬Ry) ∧ A^R ∧ $B^{¬R}$ in M.

For distributivity, because we are in a lattice, it suffices to check

sup(x,inf(y,z)) = inf(sup(x,y), sup(x,z)).

Choose representatives A,B,C of x,y,z. Choose unary R not in A,B,C. It suffices to show

($\exists$x,y)(Rx $\land$ ¬Ry)$\land$ A^R $\land$ (B $\lor$ C)$^{¬R}$ =I
(($\exists$x,y)(Rx $\land$ ¬Ry) $\land$ A^R $\land$ B$^{¬R}$) $\lor$ (($\exists$x,y)(Rx $\land$ ¬Ry) $\land$ A^R $\land$
C$^{¬R}$).

In fact, we have logical equivalence. QED

THEOREM 2.3. sup(x,y) = 1 $\leftrightarrow$ x = 1 $\land$ y = 1. inf(x,y) = 0 $\leftrightarrow$
x = 0 $\lor$ y = 0. I.e., in the Tarski degrees, 1 is not a sup,
and 0 is not an inf.

Proof: Let [A],[B] <I 1. Then ($\exists$x,y)(Rx $\land$ ¬Ry) $\land$ A^R $\land$ B$^{¬R}$
has a model, and so sup([A],[B]) <I 1. Let [A],[B] > 0.
Then A $\lor$ B has no finite model, and so inf([A],[B]) > 0.
QED

It will be convenient to write A $\land\land$ B for ($\exists$x,y)(Rx $\land$ ¬Ry)
$\land$ A^R $\land$ B$^{¬R}$. Here R is not in A,B. We can live with the
ambiguity of which R is chosen.

THEOREM 2.4. [Fe60]. Every Tarski degree <I 1 lies strictly
below some Tarski degree <I 1.

Proof: Let A be consistent. Consider T $\land$ Con(A), where T is
the conjunction of a reasonably healthy finitely fragment
of PA = Peano Arithmetic. A $\leq$I T $\land$ Con(A) by formalizing
Gödel's completeness theorem in T. Since T $\land$ Con(A) is
true, T $\land$ Con(A) <I 1.

Suppose T $\land$ Con(A) is interpretable in A. Then

T |- Con(A) $\rightarrow$ Con(T $\land$ Con(A)).
T + Con(A) |- Con(T + Con(A)).
T + Con(A) is inconsistent (by Gödel's second
incompleteness theorem).

But T + Con(A) is true. QED

A particularly appropriate choice of T $\subseteq$ PA is EFA =
exponential function arithmetic = IΣ_0(exp). See [HP98], p.
37-44, p. 495.

EFA is in 0,S,+,•,exp,< with the proper axioms

1. Sx $\neq$ 0.
2. Sx = Sy $\rightarrow$ x = y.
3. x+0 = x.
4. x+Sy = S(x+y).

5. x•0 = 0.
6. x•Sy = x•y + x.
7. exp(x,0) = S0.
8. exp(x,Sy) = exp(x,y)•x.
9. x < y ↔ (∃z)(y = x+Sz).
10. Induction for all bounded formulas.

It is known that EFA is finitely axiomatizable. E.g., this can be obtained from [HP98], Theorem 5.8, p. 366.

A Π^0_1 sentence is a sentence in 0,S,+,•,<, beginning with universal quantifiers, followed by a formula with only bounded quantifiers.

LEMMA 2.5. There is a polynomial time function α from Π^0_1 sentences to sentences in 0,S,+,•,exp,<, such that

if EFA |- φ then α(φ) has a finite model.
if EFA |- ¬φ then α(φ) has no model (A is inconsistent).

Proof: Let φ be Π^0_1. Let α(φ) be the conjunction of:

1. < is a linear ordering with minimum 0, and at least three elements.
2. x is greatest → Sx = x+y = y+x = x•y = y•x = exp(x) = x.
3. x < y → Sx is the immediate successor of x in <.
4. x ≠ 0 → (∃y)(x = Sy).
5. (x,y < z ∧ Sx = Sy) → x = y.
6. x < y → x + 0 = x.
7. x < y → x • 0 = 0.
8. S(x+y) < z → x+Sy = S(x+y).
9. x•y + x < z → x•Sy = x•y + x.
10. x < y → exp(x,0) = S0.
11. exp(x,y)•x < z → exp(x,Sy) = exp(x,y)•x.
12. x is greatest → there is a proof with Gödel number < x of EFA → φ, and there is no proof with Gödel number < x of EFA → ¬φ.
13. There is no greatest element → EFA ∧ Con(EFA + φ).

In the above, we assume a fixed finite axiomatization of EFA, and also that "x is the Gödel number of a proof of the formula with Gödel number y" is formalized by a bounded formula in the language of EFA, where y is formally required to bound all of the values of the relevant terms.

Suppose EFA |- φ. Let n be the Gödel number of a proof of EFA → φ. We can build a model of α(φ) on 0,1,...,n+1 as

follows. Take < to be as usual. Define S,+,•,exp on [0,n] as usual provided the result is ≤ n; n+1 otherwise. Define S,+,•,exp to be n+1 if at least one argument is n+1. Axioms 1-7,10 are immediate.

For 8, let S(x+y) < z in this model. Then val(x+y) < n. Hence val(x+y) = val(x)+val(y) < n. Hence val(S(x+y)) = val(x)+val(y)+1. Also val(x),val(y) < n. Hence val(Sy) = val(y)+1 ≤ n. And val(x)+val(Sy) = val(x)+val(y)+1 ≤ n. Hence val(x+Sy) = val(x)+val(y)+1 = val(S(x+y)).

For 9, let val(x•y+x) < z in this model. Then x,y ≤ n. Also val(x•y+x) = val(x•y)+x ≤ n. Hence val(x•y) = x•y ≤ n. Hence val(x•y+x) = x•y+x = x(y+1) ≤ n. Also val(Sy) = y+1. Hence val(x•Sy) = x(y+1) since x(y+1) ≤ n.

For 11, assume exp(x,y)•x < z in this model. Then val(exp(x,y)) ≤ n, and so y ≤ n, S(y) ≤ n+1, val(exp(x,y)) = x^y ≤ n, x^y•x ≤ n, val(exp(x,y)•x) = x^y•x = x^{y+1}. Hence val(exp(x,Sy)) = x^{y+1} since x^{y+1} ≤ n.

For 12, let x be greatest. Then x = n+1, and n is the Gödel number of a proof of EFA → φ. Hence this holds in the model. Also, there cannot be a proof with Gödel number ≤ n of EFA → ¬φ, according to the model, since otherwise, there would actually be a proof of EFA → ¬φ. But this contradicts that there is a proof of EFA → φ.

Obviously 13 is vacuously true in this model.

Now suppose EFA |- ¬φ. Let M be a model of α(φ). Suppose there is a greatest element x in M. In M, < is a linear ordering with least element 0, and < must be discrete by axioms 3,4. Hence we can generate standard integers 0,1,2,... which either goes on forever, or stops when it reaches x.

If it goes on forever, then this standard part is a copy of the standard model with ordinary +,•,exp. Hence the second part of axiom 12 is violated.

If it does not go on forever, then x is standard. We again contradict the second part of axiom 12.

Now suppose there is no greatest element x in M. By axiom 13, M |= EFA + Con(EFA + φ). But this is again impossible by EFA |- ¬φ. QED

THEOREM 2.6. The unary Tarski degrees form a dense distributive lattice with 0 < 1. I.e., a <I b → (∃c)(a <I c <I b).

Proof: Let A,B be sentences, where A <I B. By self reference, let φ be a Π^0_1 sentence such that EFA proves φ if and only if

Every interpretation certificate of (A ∧∧ α(φ)) ∨ B in A has a smaller interpretation certificate of B in (A ∧∧ α(φ)) ∨ B.

Suppose (A ∧∧ α(φ)) ∨ B ≤I A, and let n be the least interpretation certificate.

case 1. There is an interpretation certificate of B in (A ∧∧ α(φ)) ∨ B that is smaller than n. Then EFA |- φ. By Lemma 2.5, α(φ) has a finite model. Now there is an interpretation of B in A ∧∧ α(φ). Since A ∧∧ α(φ) ≤I A, we have B ≤I A, which is a contradiction.

case 2. There is no interpretation certificate of B in (A ∧∧ α(φ)) ∨ B that is smaller than n. Then EFA |- ¬φ. By Lemma 2.5, α(φ) has no model. Hence (A ∧∧ α(φ)) ∨ B is logically equivalent to B. Therefore B ≤I A, which is a contradiction.

We have refuted (A ∧∧ α(φ)) ∨ B ≤I A.

Now suppose B ≤I (A ∧∧ α(φ)) ∨ B. Let n be the least interpretation certificate.

case 3. There is no interpretation certificate of (A ∧∧ α(φ)) ∨ B in A smaller than or equaled to n. Then EFA |- φ. By Lemma 2.5, α(φ) has a finite model. Clearly B ≤I A ∧∧ α(φ). Hence B ≤I A, which is a contradiction.

case 4. There is an interpretation certificate of (A ∧∧ α(φ)) ∨ B in A smaller than or equaled to n. Then EFA |- ¬φ. By Lemam 2.5, α(φ) has no model. Hence (A ∧∧ α(φ)) ∨ B is logically equivalent to B. Therefore B ≤I A, which is a contradiction.

We have refuted B ≤I (A ∧∧ α(φ)) ∨ B.

Since $A \leq_I (A \wedge\wedge \alpha(\varphi)) \vee B \leq_I B$, we have $A <_I (A \wedge\wedge \alpha(\varphi)) \vee B <_I B$, as required. QED

We can sharpen Theorem 2.6 as follows.

In any lattice, let $a_1, a_2, \ldots$ be a finite or finite sequence of points. We say that this sequence is independent if and only if in the lattice, for all $n, m \geq 1$,

if $\inf(a_{i1}, \ldots, a_{in}) \leq \sup(a_{j1}, \ldots, a_{jm})$ then $\{i_1, \ldots, i_n\} \subseteq \{j_1, \ldots, j_m\}$.

THEOREM 2.7. In the unary Tarski degrees, let $a <_I b$. There exists an independent infinite sequence $c_1, c_2, \ldots$ lying in (a, b).

Proof: Let $A \in a$, $B \in b$. Then $A <_I B$. By self reference, let $\varphi(n)$ be a Π^0_1 formula with only the free variable n, such that EFA proves $\varphi(n)$ if and only if

for every interpretation certificate of $(A \wedge\wedge \alpha(\varphi(n^*))) \vee B$ in A, or one which ruins the independence of the finite sequence

$$(A \wedge\wedge \alpha(\varphi(1^*))) \vee B, \ldots, (A \wedge\wedge \alpha(\varphi(n^*))) \vee B$$

using $(A \wedge\wedge \alpha(\varphi(n^*))) \vee B$ on the left (inside the inf), there is a smaller interpretation certificate of B in $(A \wedge\wedge \alpha(\varphi(n^*))) \vee B$, or a smaller interpretation certificate which ruins the independence of the finite sequence

$$(A \wedge\wedge \alpha(\varphi(1^*))) \vee B, \ldots, (A \wedge\wedge \alpha(\varphi((n)^*)) \vee B$$

using $(A \wedge\wedge \alpha(\varphi(n^*))) \vee B$ on the right (inside the sup).

Here we use n^* for the closed term $S \ldots S0$, with n S's.

We now show that the infinite sequence

$$(A \wedge\wedge \alpha(\varphi(1^*))) \vee B, (A \wedge\wedge \alpha(\varphi(2^*))) \vee B, \ldots$$

is independent and lies in (a, b). Obviously, this infinite list lies in $[a, b]$.

Let n be least such that it is not the case that

$$(A \wedge\wedge \alpha(\varphi(1^*))) \vee B, (A \wedge\wedge \alpha(\varphi(2^*))) \vee B, \ldots,$$

$$(A \wedge\wedge \alpha(\varphi(n*))) \vee B$$

is independent. We will obtain a contradiction.

Consider the following two cases.

case 1. $(A \wedge\wedge \alpha(\varphi(n*))) \vee B \leq I A$, or there is a ruining of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B, ..., (A \wedge\wedge \alpha(\varphi(n*))) \vee B$, using $(A \wedge\wedge \alpha(\varphi(n*))) \vee B$ on the left.

Suppose EFA |- ¬$\varphi(n*)$. Then by Lemma 2.5, $\alpha(\varphi(n*))$ has no model, and so $(A \wedge\wedge \alpha(\varphi(n*))) \vee B$ is logically equivalent to B. Hence $B \leq I A$ or there is a ruining of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B, ..., (A \wedge\wedge \alpha(\varphi(n-1*))) \vee B,B$, using B on the left. Since $(A \wedge\wedge \alpha(\varphi(1*))) \vee B, ..., (A \wedge\wedge \alpha(\varphi(n-1*))) \vee B \leq I B$, the B, being on the left side, which is an inf, gets absorbed. Hence $B \leq I A$ or there is a violation of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B, ..., (A \wedge\wedge \alpha(\varphi(n-1*))) \vee B$. The former is impossible, and the latter contradicts the choice of n.

Since EFA does not refute $\varphi(n*)$, we see that EFA |- $\varphi(n*)$. Hence by Lemma 2.5, $\alpha(\varphi(n*))$ has a finite model, and so $(A \wedge\wedge \alpha(\varphi(n*))) \vee B =I A \vee B =I A$.

Also since EFA does not refute $\varphi(n*)$, we see that $B \leq I (A \wedge\wedge \alpha(\varphi(n*))) \vee B$, or there is a ruining of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B,...,(A \wedge\wedge \alpha(\varphi((n)*))) \vee B$ using $(A \wedge\wedge \alpha(\varphi(n*))) \vee B$ on the right. Hence $B \leq I A$, or there is a ruining of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B,...,(A \wedge\wedge \alpha(\varphi((n-1)*)) \vee B,A$, using A on the right. Since $A \leq I (A \wedge\wedge \alpha(\varphi(1*))) \vee B,...,(A \wedge\wedge \alpha(\varphi((n-1)*)) \vee B$, the A, being on the right side, which is a sup, gets absorbed. Hence $B \leq I A$, or there is a violation of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B,...,(A \wedge\wedge \alpha(\varphi((n-1)*))) \vee B$. The former is impossible, and the latter contradicts the choice of n.

case 2. $B \leq I (A \wedge\wedge \alpha(\varphi(n*))) \vee B$, or there is a ruining of the independence of the finite sequence $(A \wedge\wedge \alpha(\varphi(1*))) \vee B, ..., (A \wedge\wedge \alpha(\varphi(n*))) \vee B$, using $(A \wedge\wedge \alpha(\varphi(n*))) \vee B$ on the right. Because case 1 is impossible, we have EFA |- $\varphi(n*)$. Hence by Lemma 2.5, $\alpha(\varphi(n*))$ has a finite model. Hence $(A \wedge\wedge \alpha(\varphi(n*))) \vee B =I A \vee B =I A$.

Hence B ≤I A, or there is a ruining of the independence of the finite sequence (A ∧∧ α(φ(1*))) ∨ B, ...,(A ∧∧ α(φ((n-1)*))) ∨ B,A, using A on the right. The former is impossible. For the latter, since A ≤I (A ∧∧ α(φ(1*))) ∨ B,...,(A ∧∧ α(φ((n-1)*)) ∨ B, the A, being on the right side, which is a sup, gets absorbed. Hence there is a violation of the independence of the finite sequence (A ∧∧ α(φ(1*))) ∨ B,...,(A ∧∧ α(φ((n-1)*))) ∨ B. This contradicts the choice of n.

Since both cases lead to a contradiction, we see that there is no violation of the independence of the finite sequence (A ∧∧ α(φ(1*))) ∨ B, ...,(A ∧∧ α(φ(n*))) ∨ B using (A ∧∧ α(φ(n*))) ∨ B. Hence there is a violation of the independence of the finite sequence (A ∧∧ α(φ(1*))) ∨ B, ...,(A ∧∧ α(φ((n-1)*))) ∨ B. This contradicts the choice of n.

Thus we have established the independence of the infinite sequence

    (A ∧∧ α(φ(1*))) ∨ B,(A ∧∧ α(φ(2*))) ∨ B,...

It remains to show that these sentences lie in (a,b). Let m be least such that (A ∧∧ α(φ(m*))) ∨ B ≤I A. Since there is no interpretation of B in (A ∧∧ α(φ(m*))) ∨ B, we see that EFA |- ¬φ(m*). Hence α(φ(m*)) has no model, and so B ≤I A, which is a contradiction. Hence for all m, not (A ∧∧ α(φ(m*))) ∨ B ≤I A.

Let m be least such that B ≤I (A ∧∧ α(φ(m*))) ∨ B. Since there is no interpretation of (A ∧∧ α(φ(m*))) ∨ B in A, we see that EFA |- φ(m*). Hence α(φ(m*)) has a finite model, and so B ≤I A ∨ B and B ≤I A. This is impossible. Hence for all m, not B ≤I (A ∧∧ α(φ(m*))) ∨ B. QED

Do all unary Tarski degrees other than 0,1 look alike? The answer is strongly no.

We have already seen that 1 is not a sup. Therefore the sup of any two incomparable Tarski degrees is a sup ≠ 0,1. However, other unary Tarski degrees ≠ 0,1 are not sups.

Let M,M′ be two structures with disjoint domains. We write M + M′ for the structure with domain dom(M) ∪ dom(M′), where the relations and constants of M and M′ remain

unchanged, the functions of M are extended to dom(M') by returning the first argument, and a new unary relation is introduced caring out dom(M).

LEMMA 2.8. Let M,M' be two structures with disjoint domains. Suppose M + M' liberally defines a linear ordering with no greatest element. Then M or M' liberally defines a linear ordering with no greatest element.

LEMMA 2.9. Let M,M' be two structures with disjoint domains. Suppose M + M' liberally defines a pairing function with at least two inequivalent elements. Then M or M' alone liberally defines a pairing function with at least two inequivalent elements.

THEOREM 2.10. In the unary Tarski degrees, "Linear ordering with no greatest element" is not a sup. "Pairing function with at least two elements" is not a sup.

Proof: Suppose [linear ordering with no greatest element] = sup([A],[B]). Let M |= A ∧∧ B. Let M' be the part of M for A and M'' be the part of M for B. Since M liberally defines a linear ordering with no greatest element, M' or M'' liberally defines a linear ordering with no greatest element. Hence [A] or [B] is [linear ordering with no greatest element]. The analogous argument proves the second claim. QED

A sentence A is said to be complete if and only if for all sentences B in L(A), A |- B or A |- ¬B.

THEOREM 2.11. Let A be a complete sentence. Then [A] is not an inf in the unary Tarski degrees.

Proof: Let A be a complete sentence. Let [A] = inf([B],[C]) = [B ∨ C]. Let M |= A. Then M has a liberally defined model of B ∨ C. Hence M has a liberally defined model of B or M has a liberally defined of C. We now use the completeness of A. In the former case, every M |= A has a liberally defined model of B. In the latter case, every M |= A has a liberally defined model of C. QED

3. Adequate Sentences and Relative Consistency.

A very important condition is adequacy of a theory T. It is convenient to first define adequacy for a model.

We say that M is adequate if and only if there is an M definable binary relation whose cross sections are closed under single point additions and single point deletions.

Another, more common, definition is that there is an M definable binary relation with an empty cross section, whose cross sections are closed under single point additions.

THEOREM 3.1. These two definitions of adequacy are equivalent.

Proof: Let M be a model with an M definable binary relation R whose cross sections are closed under single point additions and single point deletions. Fix $c \in dom(M)$. Define $S(x,y) \leftrightarrow (R(x,y) \leftrightarrow \neg R(c,y))$. Clearly the cross section of S at c is empty. Now let x,y be given. If $\neg R(c,y)$ then choose a cross section of R obtained by inserting y in the cross section of R at x. If $R(c,y)$ then choose a cross section of R obtained by deleting y from the cross section of R at x.

Let M be a model with an M definable binary relation R with an empty cross section, whose cross sections are closed under single point additions.

We first use the following construction for getting: an empty cross section, and closure under single point additions and pairwise intersection. Call a cross section good if and only if its intersection with every cross section is a cross section. Obviously, empty cross sections are good. Let X,Y be good cross sections and $x \in X$. To check that $X \cup \{x\}$ is good, clearly $X \cup \{x\} \cup \{y\}$ and $(X \cup \{x\}) \cap Z$ are cross sections since $X \cap Z$ is a cross section. To check that $X \cap Y$ is good, clearly $(X \cap Y) \cup \{x\}$ and $(X \cap Y) \cap Z$ are cross sections, since $X \cap Y$, $Y \cap Z$ are cross sections.

Call a cross section X of R very good if and only if it is good, and any single point deletions of any good cross section that is a subset of X, is a good cross sections.

We claim that the very good cross sections of R are closed under single point additions and single point deletions. To see this, let X be a very good cross section and $x \in dom(M)$. We must show that $X \cup \{x\}$ and $X\setminus\{x\}$ are both very good cross sections. Clearly there are both good cross

sections. We can assume without loss of generality that $x \notin X$.

Let $Y \subseteq X \cup \{x\}$ be a good cross section of R. Any $Y\setminus\{y\}$ is $Y\setminus\{x\}\setminus\{y\}$ or $(Y\setminus\{x\}\setminus\{y\}) \cup \{x\}$. Since X is good, $X \cap Y$ is good, and $X \cap Y = Y|\{x\} \subseteq X$. Since X is very good, $Y\setminus\{x\}$ is a good subset of X, and so $Y\setminus\{x\}\setminus\{y\}$ and $(Y\setminus\{x\}\setminus\{y\}) \cup \{x\}$ are good. Therefore $X \cup \{x\}$ is very good.

Let $Y \subseteq X\setminus\{x\}$ be good. Since $Y \subseteq X$, single point deletions to Y remain good. QED

We give four equivalent forms of adequacy for a theory T. We begin with two syntactic definitions.

We say that a theory T is adequate if and only if there exists a formula $\varphi(x,y,z_1,\ldots,z_n)$ in L(T) such that the following is provable in T. There exists $1 \le i \le k$ (as a disjunction) and $z_1,\ldots,z_n$ (normal universal quantification) such that

i. $(\exists x)(\forall y)(\neg\varphi_i(x,y,z_1,\ldots,z_n))$.
ii. $(\forall u,w)(\exists x)(\forall y)(\varphi_i(x,y,z_1,\ldots,z_n) \leftrightarrow \varphi_i(u,y,z_1,\ldots,z_n) \vee y = w)$.

i'. $(\forall u,w)(\exists x)(\forall y)(\varphi_i(x,y,z_1,\ldots,z_n) \leftrightarrow \varphi_i(u,y,z_1,\ldots,z_n) \vee y = w)$.
ii'. $(\forall u,w)(\exists x)(\forall y)(\varphi_i(x,y,z_1,\ldots,z_n) \leftrightarrow \varphi_i(u,y,z_1,\ldots,z_n) \wedge y \neq w)$.

In the case of finite T, it is easily seen that adequacy of T is equivalent to the model theoretic condition that every model of T is adequate (in any of the two senses shown equivalent by Theorem 3.1). This uses a compactness argument.

It will also be convenient to rephrase the adequacy of T in terms of interpretations.

We call an interpretation regular if the domain used for the interpretation is the full domain, and the equality relation used is equality.

It is clear that the adequacy of T is equivalent to the regular interpretability of the following weak system NW of set theory, formulated in the usual language $\in, =$.

N (Null axiom). $(\exists x)(\forall y)(y \notin x)$.
W (With axiom). $(\exists y)(\forall z)(z \in y \leftrightarrow (z \in x \vee z = x))$.

Or we can equivalently use WD.

W (With axiom). $(\exists y)(\forall z)(z \in y \leftrightarrow (z \in x \vee z = x))$.
D (Delete axiom). $(\exists y)(\forall z)(z \in y \leftrightarrow (z \in x \wedge z \neq x))$.

The system NW is credited to Ed Nelson in [MM94], where it is shown that Q is interpretable in NW. (They also credit Jan Krajicek for an earlier and unpublished proof of this result, not known to them at the time of their publication).

The interpretability of NW in Q was already mentioned in [TMR53], p. 34, where the authors report that the interpretability of even NWE had been established by W. Szmielew and A. Tarski in 1950. Here E is

E. Extensionality axiom. $(\forall z)(z \in x \leftrightarrow z \in y) \rightarrow x = y$.

The argument in Theorem 3.1 establishes the mutual regular interpretability of WD and NW. However, WD, NW, WDE, NWE are only mutually interpretable, not mutually regularly interpretable. E.g., they have the same Tarski degree.

In fact, WD represents a very important Tarski degree with many different kinds of representatives. A particularly important representative of this Tarski degree is the well known weak system of arithmetic called Q, or Robinson's arithmetic after R. Robinson [Ro52]. Also see [HP98], p. 28. The language is $0, S, +, \cdot, \leq, =$, and the nonlogical axioms are

1. $S(x) \neq 0$.
2. $S(x) = S(y) \rightarrow x = y$.
3. $x \neq 0 \rightarrow (\exists y)(S(y) = x)$.
4. $x + 0 = x$.
5. $x + S(y) = S(x + y)$.
6. $x \cdot 0 = 0$.
7. $x \cdot S(y) = (x \cdot y) + x$.
8. $x \leq y \leftrightarrow (\exists z)(z + x = y)$.

Obviously the explicit definitional axiom 8 is redundant. However, since for many purposes, Q is normally extended by additional axioms that do involve $\leq$, it is convenient to incorporate 8.

One such important extension of Q is the system IΣ_0. This
also goes under the name "bounded arithmetic". I prefer the
name PFA = polynomial function arithmetic. In this system,
we add
the induction scheme for bounded formulas in L(Q) to Q.

A bounded formula in L(Q) is a formula all of whose
quantifiers are bounded to terms that do not contain the
variable; i.e., ($\forall x \le t$), ($\exists x \le t$), where t is a term of
L(Q) in which x does not appear. Thus the nonlogical axioms
of PFA are 1-8 and

9. Bounded induction scheme. ($\varphi[x/0] \wedge (\forall x)(\varphi \to \varphi[x/Sx])$)
$\to \varphi$, where φ is a bounded formula of L(Q).

PFA is believed to be not finitely axiomatizable, although
this is not known. See [HP98], p. 350.

It is well known that PFA is interpretable in Q. See
[HP98], p. 366.

THEOREM 3.2. WD, NW, WDE, NWE, Q, PFA are mutually
interpretable.

Theories of a tame character such as Presburger arithmetic,
real closed fields, and algebraically closed fields, are
interpretable in Q, but cannot interpret Q.

Let T be adequate. An adequacy mechanism for T consists of
formulas of L(T) that

i. define the internal nonnegative integers by a unary
predicate, together with an equivalence relation
representing equality. T must prove that this is an
equivalence relation E.
ii. define <,0,S,+,• on the natural numbers. T must prove
everywhere definedness and univalence (with respect to E).
iii. define the internal finite sequences by a unary
predicate.
iv. define the length function and value function for
finite sequences (i.e., the i-th term).
v. T proves that there is a finite sequence of length 0,
and that the finite sequences are closed under appending
any object on the right - and that appending raises the
length
by 1.

vi. T proves the axioms of PFA formulated using ii above, with bounded induction extended to incorporate all of the symbols of L(T).

THEOREM 3.3. Every adequate theory has an adequacy mechanism.

We need a notion of restricted proof in first order predicate calculus with equality. For definiteness, we use the following logical axioms and rules in $\forall, \exists, \wedge, \vee, \rightarrow, \leftrightarrow, \neg, =$.

i. All tautologies.
ii. $(\forall x)(\varphi) \rightarrow \varphi[x/t]$, t substitutable for x in φ.
iii. $\varphi[x/t] \rightarrow (\exists x)(\varphi)$, t substitutable for x in φ.
iv. x = x.
v. $x = y \rightarrow (\varphi \rightarrow \varphi')$, where φ is atomic and φ' is obtained from φ by replacing an occurrences of x in φ by y.
vi. From $\varphi \rightarrow \psi$ derive $\varphi \rightarrow (\forall x)(\psi)$, where x is not free in φ.
vii. From $\varphi \rightarrow \psi$ derive $(\exists x)(\varphi) \rightarrow \psi$, where x is not free in ψ.
viii. From φ, $\varphi \rightarrow \psi$, derive ψ.

A proof consists of a finite sequence of formulas, each of which is either an axiom above, or follows from previous entries in the sequence by one or more rules of inference above. A proof of φ is a proof whose last entry is φ. We write PROV(φ).

For any formula φ, we define lth(φ) to be the total number of symbols in φ, including parentheses, connectives, quantifiers, constants, relations, functions, and variables. Officially, the constants, relations, functions, and variables are given with a single symbol adorned with subscripts in binary. Each occurrence of these symbols is charged one plus the length of the subscript.

Note that there is a small fixed constant b such that the number of formulas φ with lth(φ) $\leq$ n is at most b^n. In fact, we take b to be the length of

$$\forall \; \exists \; \wedge \; \vee \; \rightarrow \; \leftrightarrow \; \neg \; = \; R \; F \; c \; 0 \; 1 \; (\;) \; ,$$

which is 16.

A restricted proof is a proof with the additional requirements that

i. All nonlogical symbols used are present in the last entry, φ.
ii. Every entry is a propositional combination of formulas, each of which have at most lth(φ) quantifiers.

We write RPROV(φ) for "φ has a restricted proof". The size of a proof is the total number of symbols that it contains. This is the same as the sum of the lengths of its entries. We write RPROV(φ,n) for "φ has a restricted proof of size ≤ n".

We write RCON(A) for ¬RPROV(¬A).

We write $2^{[p]}(n)$ for an exponential stack of p 2's with n on top, where if p = 0, this is n.

THEOREM 3.4. Let (∀n)(∃m)(φ(n,m)) be a Π^0_2 sentence, where φ is bounded in L(Q). The following are equivalent.
i. EFA |- (∀n)(∃m)(φ(n,m)).
ii. There exists p ∈ N such that PFA |- (∀n,r)(r = $2^{[p]}(n)$ → (∃m ≤ r)(φ(n,m))).

Proof: In ii, we are using an adequate formalization of the exponentiation relation in PFA, which is well known to exist. Let ψ = (∀n)(∃m)(φ(n,m)) be as given. Assume i. Suppose ii is false. Let T be the theory in L(Q) with constants d,e, consisting of

i. PFA.
ii. The scheme d = $2^{[p]}(e)$ ∧ (∀r ≤ d)(¬φ(d,r)), p ≥ 0.

Since ii is false, every finite fragment of T is consistent. Hence by compactness, T is consistent. Let M |= T. Let M' be the initial segment of M determined by the iterated exponentials of d. Then M' |= EFA ∧ ¬(∀n)(∃m)(φ(n,m)). Hence ii holds.

The converse ii → i is trivial. QED

Let M be adequate (see Theorem 3.1), and let an adequacy mechanism be given (see Theorem 3.3). A cut in M is taken to be an initial segment of (the numerical part of) M, with no greatest element. A proper cut in M is the cut in M consisting of all nonnegative integers of M (under the given adequacy mechanism). We will work with M definable cuts in M, only.

The following result was essentially proved in [Fr76], [Fr80], and exposited in [Sm85], [Vi90].

THEOREM 3.5. (Interpretability = relative restricted consistency). Let A,B be sentences, where B is adequate. The following are equivalent.
i. A ≤I B.
ii. EFA |- RCON(B) → RCON(A).
i → ii does not require the adequacy of B. However, ii → i does require the adequacy of B.

Proof: i → ii. Let π be an interpretation of A in B. Now π specifies a parameterized family of structures M*. I.e., there are free variables for the unspecified parameters. We have a proof from B that parameters can be chosen so that π actually defines a specific model M* of A.

We now argue in EFA. Let α be a restricted proof of ¬A. So π converts α to a proof from B of

there exists a choice of parameters such that the specific model M* satisfies A as well as $\alpha_1,...,\alpha_k = \neg A$.

Obviously, this gives us a proof of a contradiction from B, and hence a proof of ¬B.

But how complicated are the formulas involved in the proof of ¬B obtained in this way? There are a number of sources of complexity here, including the use of function symbols in A that give rise to terms of uncontrolled complexity that have to be interpreted in M*.

So we only claim that the formulas involved in the refutation of B have a standard integer bound on their quantifier depth. Quantifier depth counts blocks of like quantifiers like a single quantifier.

Fortunately, cut elimination has been reworked in terms of quantifier depth. See, e.g., [Zh91], [Zh94], [Ge05]. Since we have only standard quantifier depth, we can reduce the cut complexity to zero in a finite number of steps within EFA. EFA will take care of the exponential blowup (no worse) at each stage.

The implication ii → i is considerably more involved. Here cuts in nonstandard models are used, combined with formalized completeness proofs.

Let A,B be given, where B is adequate. Let an adequacy mechanism for B be given (Theorem 3.3).

Assume EFA |- RCON(B) → RCON(A). Obviously

EFA |- (∀n)(∃m)(RPROV(¬A,n) → RPROV(¬B,m)).

Hence by Theorem 3.4, let p be such that

$I\Sigma_0$ |- (∀n,r)(r = $2^{[p]}$(n) ∧ RPROV(¬A,n) → RPROV(¬B,r)).

Let M |= B. We need to define a model M* |= A in M.

We use an attempted Henkin construction of a model of A in M. There are a number of problems that we encounter in order to make this work. All of these problems stem from the fact that we don't have much induction in M. However, we recover by making things work in M definable cuts.

Let L(A)' be L(A) together with the Henkin constants d0,d1,... . Let W be the set of all sentences φ of L(A)' which are propositional combinations of sentences whose number of quantifiers is at most #(¬A), and whose nonlogical symbols appear in A.

We begin with listing certain elements of W. If EFA holds in (the numerical part of) M, then there is no problem listing all elements of W. Specifically, for each k, there is a unique (coded) finite sequence consisting of the first k elements of W, ordered first by length and second lexicographically. These finite sequences are initial segments of each other. Thus we can form the M definable φ0,φ1,... which enumerates all elements of W in order.

If EFA fails in M, then there is an M definable cut. By cut shortening techniques, we can construct an M definable cut C which is closed under multiplication, and where there exists t ∉ C such that 2^2^t exists. Then we can easily enumerate the first t elements of W, lexicographically, in order, as

$\varphi_0, \varphi_1, \ldots, \varphi_t$.

We can restrict the indices to elements of C, obtaining $\{\varphi_i:$ i ∈ C} as a sequence.

It is easy to see that $\{\varphi_i:$ i ∈ C} enumerates exactly the elements of W whose length n has 2^n ∈ C. (Recall the discussion above that {φ: lth(φ) ≤ n} has at most 16^n elements, and that C is closed under addition).

Note that $\{\varphi_i:$ i ∈ C} is closed under propositional combinations since C propositional combinations roughly add lengths, and C is closed under multiplication. Also note that the terms appearing in the elements of $\{\varphi_i:$ i ∈ C} are closed under the term building operations, since the term building operations roughly add lengths, and C is closed under multiplication.

So far, we can consolidate the cases where M satisfies EFA and does not satisfy EFA, by allowing C to be the improper cut of all n ≥ 0, or C to be a proper cut. In either case, C is an M definable cut closed under multiplication.

We write log(C) for the cut $\{2^i:$ i ∈ C}. Then log(C) is an M definable cut closed under addition, and {φi: i ∈ C} = {φ ∈ W: lth(φ) ∈ log(C)} = W[log(C)].

Recall that we are working entirely within M.

case 1. RCON(A). A finite sequence of sentences $\psi_0,...,\psi_{2k+1}$ from W[log(C)] is called good for A if and only if it satisfies the following conditions for all 0 ≤ i ≤ k.

i. k ∈ log(C).
ii. ψ_{2i} is φ_i or $\neg\varphi_i$.
iii. If ψ_{2i} is ¬(∀x)(φ) and c is the first constant not appearing in $\psi_0,...,\psi_{2i}$, then ψ_{2i+1} is ¬φ[x/c].
iv. If ψ_{2i} is not of the form ¬(∀x)(φ), then $\psi_{2i+1} = \psi_{2i}$.
v. There is no proof of ¬(A ∧ ψ_0 ∧ ... ∧ ψ_{2k+1}) of any size, whose entries lie in W[log(C)].
vi. For all 0 ≤ i ≤ k, if $\psi_{2i} = \neg\varphi_i$, then there is a proof of ¬(A ∧ ψ_1 ∧ ... ∧ ψ_{2i-2} ∧ φ_i) of any size, whose entries lie in W[log(C)].

It is easy to see that there exists ψ0,ψ1 that is good for A. Also, suppose $\psi_0,...,\psi_{2k+1}$ is good for A, k ≥ 0. Then there is an extension $\psi_0,...,\psi_{2k+3}$ which is still good for A.

We now assume that there are good sequences for A of every even length ∈ log(C), and also for any two sequences good for A, one is an extension of the other. Of course, these two suppositions are obvious if we had enough induction in M - which we don't.

The previous paragraph defines a particular "path" if we think of this set of M sequences as forming an internal tree.

We can define a structure on the closed terms in the obvious way from the sequences good for A. Note that the atomic sentences in L(A)' among the φi are exactly the atomic sentences whose length is in log(C), which are the atomic sentences whose closed terms have length in log(C). These closed terms are closed under the term building operations since log(C) is closed under addition.

The domain of the prospective model of A is the set of closed terms of length in log(C). We make this into a structure interpreting L(A)' in the obvious way. The interpretation of equality is: $s \equiv t \leftrightarrow s = t$ lies on the path.

We have defined an interpretation M* of L(A)'. Obviously M* is M definable. We have to show that as a structure external to M, M* |= A.

Let q be the number of quantifiers in A. We show by external induction on j from 0 through q, that for all standard formulas ρ in L(A)' with at most j quantifiers,

ρ holds in M at any assignment comprised of internal closed terms of M of length ∈ log(C), in the external sense, if and only if the result of making the substitution, creating a sentence in W[log(C)] internally in M, lies on the internal path.

This follows from the usual basic facts about what is sitting on the path. Since A is a sentence on the path, we see that M* is in fact a model of A. Both M and the outside world agree on this.

case 2. ¬RCON(A). Since M |= PFA, we have

$$M \models (\forall n, r)(r = 2^{[p]}(n) \land RPROV(\neg A, n) \rightarrow RPROV(\neg B, 2^{[p]}(n)))$$

where p was chosen above in advance of the choice of M.

We now produce an M definable cut C such that ($\forall$n $\in$ C)(RCON(A,n)).

If there is no least n such that RPROV(¬A,n), then we immediately have our desired cut. Namely the cut of all n such that RCON(A,n).

Now let n be least such that RPROV(¬A,n).

case 2a. $2^{[p]}$(n) does not exist. By cut shortening techniques, we can explicitly construct an M definable cut C below n.

case 2b. $2^{[p]}$(n) exists. Then RPROV(¬B,$2^{[p]}$(n)). Let π be a witness to this statement. We now construct an M definable cut below $2^{[p]}$(n) as follows.

We would like to give an M definable valuation of all terms in L(B) at all assignments of objects. However, we may not have enough induction in M. in M, let C be the set of all m such that for every finite list t_1,...,t_n of terms in L(B), with a total of at most m symbols, and every assignment of objects to the variables in t_1,...,t_n, there is a unique assignment of objects to all subterms of t_1,...,t_n obeying the obvious valuation clauses. It is easy to see that C is an M definable cut (not necessarily proper).

Let C′ be the set of all m $\in$ C such that there is a satisfaction relation for all propositional combinations of formulas in L(B) with at most lth(¬B) quantifiers each, whose total number of symbols is ≤ m, based on the above definition of valuation. Again, C′ $\subseteq$ C is a cut.

Finally, let C′′ be the set of all m $\in$ C′ such that there is a satisfaction relation for all finite sequences of propositional combinations of formulas in L(B) with at most lth(¬B) quantifiers each, whose total number of symbols (in the whole finite sequence) is ≤ m, based on the above definition of satisfaction. Once again, C′′ $\subseteq$ C′ $\subseteq$ C is a cut.

We cannot have $2^{[p]}$(n) $\in$ C′′ because the satisfaction relation provides a sequence of truth values for the universal closures of the formulas in this Hilbert style proof, π. We can then do an induction on the resulting bit

sequence, taking the first entry which is false, and obtaining a contradiction.

In particular, C'' is a proper cut below $2^{[p]}(n)$. We can then do cut shortening to obtain an M definable cut below n.

The cut shortening procedure can be made to yield an M definable cut C* below n in which PFA holds. We can also obtain an M definable cut C** $\subseteq$ C* several exponentials lower, where C** is closed under multiplication.

We are now in the same position that we were at the beginning of case 1. Here the full numerical part of M corresponds to the cut C* here, and the cut C there corresponds to the cut C** here. The identical argument will produce an M definable model of A. QED

The following result is used crucially in section 5.

THEOREM 3.6. Let A be a Π^0_1 sentence. There exists an adequate sentence B in L(Q) such that EFA |- A $\leftrightarrow$ RCON(B). B can be taken to be the conjunction of any sufficiently large finite fragment of PFA with some Σ^0_1 sentence in L(Q).

Proof: Firstly, we may assume A is in L(Q), since every Π^0_1 sentence is provably equivalent, in EFA, to a Π^0_1 sentence in L(Q).

Choose a sufficiently large finite fragment PFA' of PFA. Then PFA' is adequate. By self reference, let C be a Σ^0_1 sentence in L(Q) provably equivalent over PFA' to

C*. there exists a restricted proof α of ¬(PFA' $\wedge$ C) such that A is true (with outermost universal quantifiers restricted to natural numbers) $\leq$ lth(α).

The adequate sentence B that we want is the sentence B = PFA' $\wedge$ C*. Note that C* is Σ^0_1. To see that this works, we argue in EFA.

Suppose A. If ¬RCON(B) then C*. Hence C. Since C is a true Σ^0_1 sentence, we can prove PFA' $\rightarrow$ C using formulas in L(Q) which are propositional combinations of formulas each of which have at most a number of quantifiers that depends only on the choice of PFA' and C. Also, we have RCON(PFA'). Hence we conclude that RCON(PFA' $\wedge$ C), because we can

perform cut elimination several times (EFA is our
background theory). I.e., RCON(B).

Now suppose RCON(B). Let n be a counterexample to A. Note
that PFA' sees that there is no restricted proof α of $\neg$B
with lth(α) $\leq$ n, by RCON(B). Hence PFA' |- $\neg$C* using
formulas in L(Q) which are propositional combinations of
formulas each of which have at most a number of quantifiers
that depends only on the choice of PFA' and C. Hence PFA'
|- $\neg$C using formulas in L(Q) which are propositional
combinations of formulas each of which have at most a
number of quantifiers that depends only on the choice of
PFA' and C. The same must be true of a proof of $\neg$B = PFA' $\wedge$
C. By successive cut elimination, we obtain $\neg$RCON(B). Hence
the counterexample n to A must not exist. Therefore A. QED

Here is a sharpening of Theorem 3.6.

THEOREM 3.7. Let A be a Π^0_1 sentence and B be adequate, with
an adequacy mechanism. There exists a Σ^0_1 sentence C such
that EFA |- (A $\wedge$ RCON(B)) $\leftrightarrow$ RCON(B $\wedge$ C).

Proof: We again assume that A is in L(Q), and choose a
sufficiently large finite fragment PFA' of PFA. Then PFA'
is adequate. By self reference, let C be a Σ^0_1 sentence in
L(Q) provably equivalent over PFA' to

C*. there exists a restricted proof α of $\neg$(B $\wedge$ C) such that
A is true (with outermost universal quantifiers restricted
to natural numbers) $\leq$ lth(α).

We argue in EFA. Suppose A $\wedge$ RCON(B). If $\neg$RCON(B $\wedge$ C) then
C*. Therefore C. Hence B proves C with a restricted proof.
Therefore RCON(B $\wedge$ C), which is a contradiction. (For these
last two statements, we use cut elimination for a standard
number of steps in the background theory EFA). Hence RCON(B
$\wedge$ C).

Now suppose RCON(B $\wedge$ C). Let n be a counterexample to A.
Note that B sees that there is no restricted proof α of $\neg$(B
$\wedge$ C) with lth(α) $\leq$ n, by RCON(B $\wedge$ C). Hence B |- $\neg$C* with
restricted proof. Therefore B |- $\neg$C with restricted proof.
(We have again used cut elimination for a standard number
of stpes in the background theory EFA). This contradicts
RCON(B $\wedge$ C). QED

COROLLARY 3.8. Let A,B be consistent sentences, where B is adequate. There exists consistent adequate B ∧ C ≥I A. In fact, under any adequacy mechanism for B, C can be taken to be Σ^0_1.

Proof: Let A,B be as given. By Theorem 3.7, let C be such that EFA |- (RCON(A) ∧ RCON(B)) ↔ RCON(B ∧ C). In particular, EFA |- RCON(B ∧ C) → RCON(A), and B ∧ C is adequate. By Theorem 3.5, A ≤I B ∧ C. QED

An interpretation π from S to T is said to be faithful iff it does not use any parameters, and for all sentences φ ∈ L(S), S |- φ ↔ T |- πφ.

THEOREM 3.9. (Interpretability = faithful interpretability). Let A,B be sentences, where B is adequate without parameters. The following are equivalent.
i. A is interpretable in B.
ii. A is faithfully interpretable in B.

For a proof of Theorem 3.9, see [Vi05].

Theorem 3.9 is easier if we assume that there is an adequacy mechanism for B, without parameters, which satisfies all true Π^0_1 sentences. For this result, we need a Lemma.

LEMMA 3.10. Let B be consistent, with an adequacy mechanism. There is a formula φ(n) in L(B) such that for all n ≥ 0, B ∧ φ(0) ∧ ... ∧ φ(n-1) ∧ ¬φ(n) is conservative over B for all Σ^0_1 sentences. In fact, we can replace Σ^0_1 by Σ^0_k, for any k given in advance of φ.

Proof: By a well known Rosser construction. QED

Here is the weakened form of Theorem 3.9 that was promised.

THEOREM 3.11. (Interpretability = faithful interpretability). Let A,B be sentences, where B is given an adequate mechanism without parameters. Assume that B is consistent with the true Π^0_1 sentences. The following are equivalent.
i. A is interpretable in B.
ii. A is faithfully interpretable in B.

Proof: Let π be an interpretation from A into B, where B is adequate without parameters. Let φ(n) be as given by Lemma 3.10. We now define a faithful interpretation of A to B.

Let M |= B. We define a model of A within M.

Let n be least such that ¬φ(n). If this does not exist, use the default value n = 0.

Let E be the n-th sentence in L(A). If M satisfies the consistency of A ∧ E then return the model of A ∧ E obtained using the formalized completeness proof. Otherwise, return the model of A given by π.

We claim that for any E ∈ L(A) such that A ∧ E is consistent, some M |= B returns a model of A ∧ E.

Let E be as given, and let n be its Gödel number. Let B* = B ∧ φ(0) ∧ ... ∧ φ(n-1) ∧ φ(n). Then B* remains consistent with the true Π^0_1 sentences, according to Lemma 3.10.

Let M |= B*. Then the calculation in M of n, and therefore E, is correct. Hence M |= Con(A ∧ E) since Con(A ∧ E) is a true Π^0_1 sentence.

Now under this interpretation π' of A into B, which sentences ρ in L(A) are satisfied in all models arising in the interpretation? I.e., in all π'M, M |= B?

Such a sentence ρ must be consistent with every consistent sentence A ∧ E. So if A does not prove ρ, then A + ¬ρ is consistent, and ρ must be consistent with A + ¬ρ. This is patently absurd, and so ρ must be provable from A, as required. QED

4. Predicative Set Extensions and Relative Consistency.

The predicative set extension of a theory T is defined as follows. First we introduce a second sort for "sets", with the binary relation ∈ between objects of the original sort and objects of the new sort. Then we add the predicative comprehension axiom

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

where φ is a formula in the extended language without any quantifiers over the second sort, in which A is not free.

Since we are officially working in single sorted logic, convert this two sorted system to a single sorted system in the obvious way by introducing a unary predicate.

We write this extension as PSE(T). It is well known, by a simple model theoretic argument, that PSE(T) is a conservative extension of T.

THEOREM 4.1. Let A be an adequate sentence. PSE(A) is adequate, and finitely axiomatized. EFA proves RCON(PSE(A)) ↔ CON(A). EFA does not prove RCON(A) → CON(A). EFA does not prove CON(A) → CON(PSE(A)). PSE(A) is not interpretable in A.

Proof: Well known. See Theorems 5.5 and 7.1 of [Vi06]. QED

THEOREM 4.2. Let A,B be adequate sentences. The following are equivalent.
i. PSE(A) is interpretable in PSE(B).
ii. EFA proves CON(B) → CON(A).

Proof: Apply Theorem 3.5 to the adequate theories PSE(A), PSE(B). We obtain that i ↔ (EFA proves RCON(PSE(B)) → RCON(PSE(A))) → ii, using Theorem 4.1. QED

5. Adequate Tarski Degrees and P Degrees.

The adequate Tarski degrees are the Tarski degrees of adequate sentences. There is another kind of degree which is intimately connected with Tarski degrees.

It is most convenient to work in the language of EFA, $0, S$, $+, \bullet, \exp, <$. For Π^0_1 sentences A,B, we write

A ≤P B ↔ B logically implies A.

≤P is reflexive and transitive. Define =P, ≥P in the obvious way. The P degrees are the equivalence classes under =P.

EFA represents a particular P degree.

Recall Theorem 3.6.

THEOREM 3.6. Let A be a Π^0_1 sentence. There exists an adequate sentence B in L(Q) such that EFA |- A ↔ RCON(B). B can be taken to be the conjunction of any sufficiently large finite fragment of PFA with some Σ^0_1 sentence in L(Q).

THEOREM 5.1. The adequate Tarski degrees are isomorphic to the P degrees ≥ EFA.

Proof: Let A be adequate. Map it to A* = EFA ∧ RCON(A). By Theorem 3.5, for adequate A,B, we have

A ≤I B ↔ EFA |- RCON(B) → RCON(A).
A ≤I B ↔ EFA ∧ RCON(B) logically implies EFA ∧ RCON(A).
A ≤I B ↔ EFA ∧ RCON(A) ≤P EFA ∧ RCON(B).
A ≤I B ↔ A* ≤P B*.

This defines an embedding from the adequate Tarski degrees into the P degrees ≥P EFA.

To see that the embedding is onto, let EFA ∧ A be given, where A is a Π^0_1 sentence. By Theorem 3.6, Let B be an adequate sentence such that EFA |- A ↔ RCON(B). Then

B* = EFA ∧ RCON(B) =P EFA ∧ A.

QED

THEOREM 5.2. The adequate Tarski degrees, and the P degrees ≥ EFA, are isomorphic distributive lattices with 0,1. The P degrees are a distribute lattice with 0,1.

Proof: By Theorem 5.1, we need only deal with the P degrees. The P degrees ≥ EFA have EFA as 0 and 1 = 0 as 1. Obviously ∨ and ∧ are the inf and sup. The P degrees are also a distributive lattice 0 = 0 as the 0, and 1 = 0 as the 1. QED

THEOREM 5.3. In the adequate Tarski degrees, every degree < 1 is an inf. 1 is not an inf in the adequate Tarski degrees (obvious).

Proof: It suffices to work within the P degrees ≥ EFA, where it is basically well known. Let A be a consistent Π^0_1 sentence that logically implies EFA. By self reference, let φ be provably equivalent, over EFA, to

*) every proof of A → φ has a smaller proof of A → ¬φ.

Let ψ be

every proof of A → ¬φ has a smaller proof of A → φ.

Note that φ is the usual Rosser sentence construction over A, and ψ is its dual.

We claim that EFA |- φ ∨ ψ, so that A is the inf of A ∧ φ and A ∧ ψ. To see this, argue in EFA. Suppose ¬φ ∧ ¬ψ. Then ¬* ∧ ¬ψ. Then we get two distinct Gödel numbers, neither of which is > the other. This is impossible.

It remains to refute A |- φ and A |- ψ.

Assume A |- φ. Let n be the least proof. If there is a smaller proof of A → ¬φ then EFA |- ¬A. This is impossible. Hence there is no smaller proof of A → ¬φ. So ¬φ is true. Since ¬φ is Σ^0_1, we have EFA |- ¬A. This is a contradiction.

Assume A |- ¬φ. Let n be the least proof. By the previous paragraph, there is no proof of A → φ. Hence EFA |- φ, and so EFA |- ¬A. This is a contradiction.

Note that the previous two paragraphs were conducted within EFA + Con(A). Thus we obtain

EFA + Con(A) |- Con(A + φ) ∧ Con(A + ¬φ).

Note that EFA + φ ∧ ψ |- Con(A).

Assume A |- ψ. Then A + φ |- Con(A). Hence A + φ |- Con(A + φ). By Gödel's second incompleteness theorem, A + φ is inconsistent. I.e., A |- ¬φ. This is impossible. QED

THEOREM 5.4. In the adequate Tarski degrees, every degree other than 0,1, is a sup. 0,1 are not sups in the adequate Tarski degrees (obvious for 0).

Proof: (repaired by Albert Visser, private communication). We work in the P degrees ≥ EFA. Let A be a consistent Π^0_1 sentence that proves EFA, but is not provable in EFA. We show that A is a sup in the P degrees ≥ EFA.

By self reference, let φ be provably equivalent, over EFA, to

φ*. Every proof of EFA ∧ ¬φ → A has a smaller proof of EFA → ¬A or of EFA ∧ φ → A.

Let ψ be

every proof of EFA → ¬A or of EFA ∧ φ → A has a smaller proof of EFA ∧ ¬φ → A.

We claim that A is the sup of EFA ∧ (A ∨ φ) and EFA ∧ (A ∨ ψ). Obviously A logically implies these two sentences.

Note that EFA ∧ φ ∧ ψ |- Con(A) |- A. To see this, argue in EFA, and assume φ,ψ. There cannot be a proof of EFA ∧ ¬φ → A, and so Con(A).

Assume EFA ∧ ¬φ |- A. There is no proof of EFA → ¬A and no proof of EFA ∧ φ → A. Hence EFA |- ¬φ, EFA |- A. This is a contradiction.

Assume EFA ∧ φ |- A. Let n be the least proof. If there is a smaller proof of EFA ∧ ¬φ → A then, since there is no proof of EFA → ¬A, we have EFA |- ¬φ, and hence EFA |-A. This is impossible. If there is no smaller proof of EFA ∧ ¬φ → A then EFA |- φ. Hence EFA |- A, which is again a contradiction.

Assume EFA ∧ ψ |- A. Since EFA |- φ ∨ ψ, we have EFA ∧ ¬φ |- A. But this is impossible. QED

Theorems 5.3 and 5.4 may suggest that in the adequate Tarski degrees, or alternatively, in the P degrees ≥ EFA, the degrees ≠ 0,1 look alike. This is not the case.

THEOREM 5.5. In the P degrees ≥ EFA, $(\exists d < 1)(\forall e > 0)(\inf(d,e) > 0)$. The d's with this property are closed upwards. This property holds of EFA ∧ RCON(EFA), but not of EFA ∧ φ, where φ is the Rosser sentence over EFA (see the proof of Theorem 5.3.).

Proof: Obviously the property is closed upwards. Let B be a Π^0_1 sentence. Assume EFA |- RCON(EFA) ∨ B. Then

EFA + ¬B |- RCON(EFA).
EFA + ¬B |- RCON(EFA + ¬B).
EFA + ¬B is inconsistent (Gödel's second incompleteness theorem for restricted consistency).

EFA |- B.

By the proof of Theorem 5.3, inf(φ,ψ) = 0, and ψ > 0. QED

COROLLARY 5.6. In the adequate Tarski degrees, (∃d < 1)(∀e > 0)(inf(d,e) > 0). The d's with this property are closed upwards. This property holds of EFA, but not of all adequate Tarski degrees.

Proof: Immediate from Theorem 5.5, using the isomorphism from the adequate Tarski degrees onto the P degrees ≥ EFA, defined in the proof of Theorem 5.1. The isomorphism maps EFA to EFA ∧ RCON(A). QED

6. Infinite Theories.

We scratch the surface concerning interpretability between infinite theories. The nature of this subject is quite different than the finite case, as is clear from Theorems 6.2 – 6.4. Here we will assume that we are dealing with parameterless interpretations only.

We begin with a well known sufficient condition for interpretability of infinite theories.

THEOREM 6.1. Every r.e. set of sentences S is interpretable in Q + {Con(S'): S' is a finite subset of S}. As a consequence, if S,T are theories, S is r.e., T is adequate with an adequacy mechanism, and T proves the consistency of each finite fragment of S, then S is interpretable in T.

Proof: Let M |= Q + {Con(S'): S' is a finite subset of S}. Pass to an M definable cut C in M satisfying PFA. Then C |= {Con(S'): S' is a finite subset of S}.

If C |= Con(S) then we can use formalized completeness to return a model of S. Suppose C |= ¬Con(S). In C, let n be largest such that in C, the first n axioms of S are consistent. Since in C, for each standard k, the first k axioms of S are consistent, this n must be a nonstandard integer. Again use formalized completeness to return a model M* which, according to C, is a model of the first n axioms of S. Hence M* really is a model of S. The stated consequence is immediate. QED

Theorem 6.1 applies to such pairs of theories as ZFC and ZFC + ¬CH.

THEOREM 6.2. (No certificates). There is a Σ^0_2 sentence A such that
i. PA + A is interpretable in PA.
ii. ZFC doesn't prove that PA + A is interpretable in PA, provided ZFC is consistent.
iii. Moreover, "PA + A interpretable in PA" is provably equivalent to Con(ZFC) over PA.
iv. ZFC does not prove (R)Con(PA) → (R)Con(PA+A), provided ZFC is consistent.

Proof: Let A = "the greatest n such that $I\Sigma_n$ is consistent is less than any inconsistency in ZFC".

We claim that PA proves the consistency of every $I\Sigma_n$ + A. To see this, we argue in PA, and fix n. We know that $I\Sigma_n$ + Con($I\Sigma_n$) + ¬Con($I\Sigma_{n+1}$) is consistent. By exhaustive search, we see that n is less than the Gödel number of any inconsistency in ZFC. Hence $I\Sigma_n$ + A is consistent.

By Theorem 6.1, PA + A is interpretable in PA. Note that this argument is verifiable in PA + Con(ZF).

Now suppose PA + A is interpretable in PA. Suppose ZFC is inconsistent, and let n be the least Gödel number of an inconsistency in ZFC. Since PA |- Con($I\Sigma_{n+1}$), we see that PA + A is inconsistent, and therefore not interpretable in PA. Hence ZFC is consistent. This establishes ii,iii.

In particular, we have shown that if PA + A is consistent then ZFC is consistent. Hence iv. QED

For additional results along these lines, see [Sh97].

THEOREM 6.3. Let A be a Π^0_1 sentence. PA + A is interpretable in PA if and only if PA proves A.

THEOREM 6.4. There is a Π^0_1 sentence A such that
i. PA + A is not interpretable in PA.
ii. EFA |- Con(PA) ↔ Con(PA + A) ↔ RCON(PA).

Proof: Set A to be a usual Rosser sentence over PA. Since PA does not prove A, we have i by Theorem 6.3. Also ii is clear as in the proof of Theorem 5.2. QED

On the other hand, interpretations from r.e. theories into single sentences behave like interpretations from single

sentences into single sentences, as the following result indicates.

THEOREM 6.5. Let T be r.e. and B be adequate. If EFA |- RCON(B) $\rightarrow$ RCON(T) then T is interpretable in B. T is interpretable in B if and only if T is faithfully interpretable in B.

Let T be r.e. and adequate, with finite L(T). In PSE(T) we can construct a truth definition for L(T) in the standard way. We can replace T by "all axioms of T are true" in PSE(T), resulting in the finitely axiomatized system PSE*(T), which is a conservative extension of T.

Note that PSE*(T) depends on the r.e. presentation of T. Depending on some details about how to formulate PSE*(T), one can prove the following result.

THEOREM 6.6. Let S,T be adequate r.e. theories with adequacy mechanisms. If EFA proves CON(T) $\rightarrow$ CON(S) then PSE*(S) is interpretable in PSE*(T).

[Sh97] tells us, for example, that the set of Σ^0_1 sentences φ such that ZF + φ is interpretable in NBG is complete Σ^0_3.

Section 7. Observed Linearity.

We restrict attention to the adequate theories from the literature which constitute a coherent system of axioms for mathematical reasoning. Here we don't require philosophically or foundationally coherent, as this would arguably be quite restrictive. We mean coherent only in that there is a modicum amount of naturalness. But we do require that there be no metamathematical ingredients present beyond a very standard base theory.

In particular, we include any theory of the form

EFA + $\{A_1,\ldots,A_k\}$
PA + $\{B_1,\ldots,B_n\}$
RCA_0 + $\{C_1,\ldots,C_m\}$
Z + $\{D_1,\ldots,D_r\}$
ZF + $\{E_1,\ldots,E_s\}$

and much more, where $A_1,\ldots,A_k$ are robust formulations of published mathematical theorems in the language of EFA, $B_1,\ldots,B_n$ are such in the language of PA, $C_1,\ldots,C_m$ are such

in the language of RCA_0, and $D_1,...,D_r,E_1,...,E_s$ are such in the language of set theory.

The striking observation is that one finds a remarkable linearity. This linearity is found not only with finitely axiomatized systems – which we have emphasized here – but with the non finitely axiomatized systems such as PA and ZFC.

This is perhaps the most intriguing, thought provoking, fundamental, and deep phenomenon in the whole of the foundations of mathematics.

In practice, the non finitely axiomatized systems encountered are almost finitely axiomatized, in the sense that the axiomatizations consist of finitely many axioms and finitely many axiom schemes. E.g., PA and ZFC meet this criteria.

Of course, for any adequate system T based on finitely many axioms and axiom schemes, there is a canonical version of PSE(T), which is finitely axiomatizable. Under interpretability, PSE(T) is a little bit higher than T. However, in the other orderings that we consider, with one exception, $\leq$S, PSE(T) and T are equivalent. Examples: PA and ACA_0. ZF and NBG.

The formal systems from the literature range from very weak axioms of arithmetic, to the much stronger axioms of infinite set theory such as ZFC - and beyond, with the extensions of ZFC by the so called large cardinal axioms.

We consider the following relations between theories. For each of these notions, we allow systems with finitely many axioms and finitely many axiom schemes.

$\leq$I. T $\leq$I T' $\leftrightarrow$ T is interpretable in T'.

$\leq$S. Logical strength. T $\leq$S T' $\leftrightarrow$ EFA |- Con(T') $\rightarrow$ Con(T).

$\leq$S*. Liberal logical strength. T $\leq$S* T' $\leftrightarrow$ PA |- Con(T') $\rightarrow$ Con(T).

$\leq$R. Provably recursive functions. We say that f:N $\rightarrow$ N is a provably recursive function of T if and only if there is a Turing machine code e computing f such that T |- "e computes a function from N into N". We define T $\leq$R T' $\leftrightarrow$

every provably recursive function of T is a provably recursive function of T'.

≤1. T ≤1 T' ↔ every Π^0_1 sentence provable in T is provable in T'.

≤2. T ≤2 T' ↔ every Π^0_2 sentence provable in T is provable in T'.

≤∞. T ≤∞ T' ↔ every arithmetic sentence provable in T is provable in T'.

≤O. Provable ordinals. For all systems below discussing subsets of ω (or general sets), we say that α is a provable ordinal of T if and only if there is a Turing machine code e computing a well ordering of N of order type α, such that T |- "e computes a well ordering of N".

Here are some some observed linearity phenomena.

Note that comparability under ≤O is entirely automatic. This kicks in when we are extending RCA$_0$.

A. Any two natural systems interpreting Q are comparable under ≤I.

B. Any two natural systems extending EFA are comparable under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2. They agree under ≤S*, ≤R, ≤1, ≤2. If they are finitely axiomatized, then they agree under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2.

C. Any two natural systems extending PA are comparable under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2, ≤∞. They agree under ≤S*, ≤R, ≤1, ≤2, ≤∞. If they are finitely axiomatized, then they agree under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2, ≤∞.

D. Any two natural systems extending RCA$_0$ are comparable under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2, ≤∞, ≤O. They agree under ≤S*, ≤R, ≤1, ≤2, ≤∞, ≤O. If they are finitely axiomatized, then they agree under ≤I, ≤S, ≤S*, ≤R, ≤1, ≤2, ≤∞, ≤O.

E. If T and T' are natural systems extending PA, and T < T' under any of <S*, <R, <1, <2, <∞, <O, then T' proves the consistency of T.

Here is a table that lists the finite number of levels that have figured prominently in f.o.m. by means of preferred

representatives. We have linearity under ≤I (with axiom schemes allowed here). We also exemplify B – E above.

The relevant interpretations here can all be taken to be parameterless.

PFA.
EFA.
SEFA.
PRA.
RCA_0.
$I\Sigma_2$.
$I\Sigma_3$.
PA.
ACA_0.
$ACA_0 + (\forall n,x)(TJ(n,x)\downarrow)$.
ACA.
$RCA_0 + TJ(\omega)\downarrow$.
$ACA_0 + TJ(\omega)\downarrow$.
$ACA + TJ(\omega)\downarrow$.
$ACA_0 + (\forall x)(TJ(\omega,x)\downarrow)$.
$ACA_0 + \{(\forall\alpha,x)(TJ(\alpha,x)\downarrow):\alpha < \omega^\omega\}$.
$ACA_0 + \{(\forall\alpha < \omega^\omega)(\forall x)\ (TJ(\alpha,x)\downarrow)\}$.
$RCA_0 + TJ(\omega^\omega)\downarrow$.
$ACA_0 + TJ(\omega^\omega)\downarrow$.
$ACA_0 + \{(\forall x)(TJ(\omega^\omega,x)\downarrow)\}$.
$ACA_0 + \{(\forall x)(TJ(\alpha,x)\downarrow):\alpha<\in_0\}$.
$\Delta^1_1\text{-CA}$.
$RCA_0 + TJ(\in_0)$.
$ACA_0 + TJ(\in_0)$.
$ACA + TJ(\in_0)$.
$ACA_0 + (\forall x)(TJ(\in_0,x)\downarrow)$.
$\{ATI(\alpha):\alpha < \Gamma_0\}$.
ATR_0.
$ATI(<\Gamma_0)$.
ATR.
$\Pi^1_2\text{-TI}_0$.
$\Pi^1_2\text{-TI}$.
TI.
ID_2.
$ID_{<\omega}$.
$\Pi^1_1\text{-CA}_0$.
$\Pi^1_1\text{-CA}$.
$\Pi^1_1\text{-CA} + TI$.
$\Pi^1_1\text{-TR}_0$.
$\Pi^1_1\text{-TR}$.
$\Pi^1_2\text{-CA}_0$.

Π^1_2-CA.
Π^1_2-CA + TI.
Z_2.
Z_3.
Type Theory.
Weak Zermelo.
ZC.
ZC + $(\forall\alpha < \omega_1)(V(\alpha)\downarrow)$.
KP($\wp$).
ZFC.
ZFC + strongly inaccessible$\downarrow$.
ZFC + strongly Mahlo$\downarrow$.
ZFC + {strongly n-Mahlo$\downarrow$: n < ω}.
ZFC + $(\forall$n < $\omega)$(strongly n-Mahlo$\downarrow$).
ZFC + (weakly compact$\downarrow$).
ZFC + (indescribable$\downarrow$).
ZFC + (subtle$\downarrow$).
ZFC + (almost ineffable$\downarrow$).
ZFC + (ineffable$\downarrow$).
ZFC + {n-subtle$\downarrow$: n < ω}.
ZFC + $(\forall$n < $\omega)$(n-subtle$\downarrow$).
ZFC + $\kappa \rightarrow \omega$ $\downarrow$.
ZFC + $(\forall\alpha < \omega_1)(\kappa \rightarrow \alpha$ $\downarrow)$.
ZFC + 0# $\downarrow$.
ZFC + $(\forall$x $\subseteq \omega)$(x# $\downarrow$).
ZFC + $\kappa \rightarrow \omega_1$ $\downarrow$.
ZFC + Ramsey$\downarrow$.
ZFC + Measurable$\downarrow$.
ZFC + Concentrating Measurable$\downarrow$.
ZFC + Strong$\downarrow$.
ZFC + Woodin$\downarrow$.
ZFC + Superstrong$\downarrow$.
ZFC + Supercompact$\downarrow$.
ZFC + Extendible$\downarrow$.
ZFC + Vopenka$\downarrow$.
ZFC + Almost Huge$\downarrow$.
ZFC + Huge$\downarrow$.
ZFC + Superhuge$\downarrow$.
ZFC + $(\forall$n < $\omega)$(n-huge$\downarrow$).
ZFC + Rank into Itself$\downarrow$.
ZFC + Rank + 1 into Itself$\downarrow$.
VB + V into V$\downarrow$.

There are extremely natural theories for which linearity
fails, in which Q is not interpretable. For example, let

T = axioms for discrete linear orderings without endpoints (every point has an immediate predecessor and an immediate successor).

T' = axioms for dense linear orderings without endpoints (between any two elements there is a third, and no endpoints).

Then T ≤I T' and T' ≤I T both fail. Also note that T,T' are finitely axiomatized and complete.

Interpretability is very interesting in this kind of tame world, below Q – where it takes on an entirely different character than the theory ≥I Q that we have focused on.

In particular, it would be interesting to understand interpretability among subsystems of RCF (real closed fields) and ACF (algebraically closed fields) and Presburger arithmetic (the semigroup of nonnegative integers), finitely axiomatized or otherwise.

REFERENCES

[Fe60] S. Feferman, Arithmetization of metamathematics in a general setting, Fundamenta Mathematicae 49 (1960), 35-92.

[Fr76] Translatability and relative consistency, November, 1976, 7 p.

[Fr80] Translatability and relative consistency II, Ohio State University, unpublished, September, 1980, 6 p.

[Ge05] P. Gerhardy, The Role of Quantifier Alternations in Cut Elimination, Notre Dame Journal of Formal Logic, vol. 46, no. 2, pp. 165-171 (2005).

[HP98] P. Hajek, P. Pudlak, Metamathematics of First-Order Arithmetic, Perspectives in Mathematical Logic, Springer, 1998, 460 p.

[MM94] A. Mancinci, F. Montagna, A minimal predicative set theory, Notre Dame J. of Formal Logic 35 (1994) 186-203.

[Ro52] R.M. Robinson, An essentially undecidable axiom system, Proceedings of the 1950 International Congress of Mathematicians, Cambridge MA, 1952, pp. 729-730.

[Sh97] V.Y. Shavrukov, INterpreting Reflexive Theories in Finitely Many Axioms, Fundamenta Mathematice, vol. 152, p. 99-116, 1997.

[Sm85] C. Smorynski, Nonstandard models and related developments, in: Harvey Friedman's Research on the Foundations of Mathematics, North Holland: Amsterdam, 1985 pp. 179-229.

[Ta56] A. Tarski, Logic, Semantics, Metamathematics: Papers from 1923 to 1938 (translated by J.H. Woodger), Clarendon Press, Oxford 1956, 471 p.

[TMR53] A. Tarski, A. mostowski, R.M. Robinson, Undecidable Theories, North Holland: Amsterdam 1953, 98 p.

[Vi90] A. Visser, Interpretability logic, in: Mathematical logic, (Proceedings of the Heyting 1988 summer school in Vjarna, Bulgaria, Plenum Press, Boston, 1990, p. 175-209.

[Vi05] A. Visser, Faith and Falsity: a study of Faithful Interpretations and flase $\Sigma01$-sentences, Annals of Pure and APplied Logic, volume 131, p. 103-131, 2005.

[Vi06] A. Visser, The Predicative Frege Hierarchy, October, 2006, #246, http://www.phil.uu.nl/preprints/lgps/list.html.

[Zh91] W. Zhang, Cut elimination and automatic proof procedures, Theoretical Computer Science 91 (1991), 265-284.

[Zh94] W. Zhang, Depth of proofs, depth of cut-formulas and complexity of cut formulas, Theoretical Computer Science 129 (1994), 193-206.