DOES REDUCTIVE PROOF THEORY HAVE A VIABLE RATIONALE?

Solomon Feferman

Abstract

The goals of reduction and reductionism in the natural sciences are mainly explanatory in character, while those in mathematics are primarily foundational. In contrast to global reductionist programs which aim to reduce all of mathematics to one supposedly "universal" system or foundational scheme, reductive proof theory pursues local reductions of one formal system to another which is more justified in some sense. In this direction, two specific rationales have been proposed as aims for reductive proof theory, the constructive consistency-proof rationale and the foundational reduction rationale. However, recent advances in proof theory force one to consider the viability of these rationales. Despite the genuine problems of foundational significance raised by that work, the paper concludes with a defense of reductive proof theory at a minimum as one of the principal means to lay out *what rests on what* in mathematics. In an extensive appendix to the paper, various reduction relations between systems are explained and compared, and arguments against proof-theoretic reduction as a "good" reducibility relation are taken up and rebutted.

1 Reduction and reductionism in the natural sciences and in mathematics.

The purposes of reduction in the natural sciences and in mathematics are quite different. In the natural sciences, one main purpose is to *explain* certain phenomena in terms of more basic phenomena, such as the nature of the chemical bond in terms of quantum mechanics, and of macroscopic genetics in terms of molecular biology. In mathematics, the main purpose is *foundational*. This is not to be understood univocally; as I have argued in (Feferman 1984), there are a number of foundational ways that are pursued in practice. One such way is organizational; in that enterprise, reduction in the number of basic concepts and principles is valued, as is ease and naturalness of development. The purpose of other foundational ways is to deal with problematic concepts or principles by special kinds of reduction, such as (historically) the reduction of the complex numbers to the real numbers, or the reduction of the use of infinitesimals to the systematic use of limits. More recent examples from logic are the reduction of set theory with the axiom of choice to that without, or the reduction of classical arithmetic to intuitionistic arithmetic (insofar as the law of excluded middle is problematic for the constructivist). Foundational concerns are of course also important in the natural sciences, such as that of providing a philosophically satisfactory and physically adequate foundation for quantum mechanics; but one does not necessarily think of this as a reductive project. And, explanation is ubiquitous in mathematics, such as in the use of Galois theory to explain the unsolvability of the quintic, or of combinatorial topology to explain the Descartes-Euler formula for polyhedra; again, these are not usually thought of reductively.

Another contrast to be made is that between piecemeal or *local* projects of reduction in both the natural sciences and mathematics, and *global reductionist* programs in both. In the natural sciences, the philosophy of reductionism calls for a level-by-level theoretical reduction of the hierarchy of sciences to a basic material monism. This is envisioned, for example, by Oppenheim and Putnam (1958) as proceeding from social groups on down through multicellular organisms, living cells, molecules and atoms, all the way to elementary particles. (Nowadays, that is to be capped by the physicists' holy grail of the GUT, the "Grand Unified Theory", which is then to be the TOE, the "Theory of Everything".) I am personally very skeptical of this kind of reductionist program in science, for reasons that I have not tried to articulate, at least not in writing (and won't try to here).[1]

Global reductionist programs in the foundations of mathematics share the monistic view with scientific reductionism, but there is no real analogy with the hierarchical account. The most prominent examples to consider are the logicist program, set-theoretical foundations, functional foundations[2], categorical foundations and—in a sense—Hilbert's program (in its original conception). There are well-known problems with each of these that I shall not repeat here, though advocates of one or another of these programs persist in pushing them. My own view is again skeptical, and leads me to pursue local projects of reduction instead. But I would hope that even those who don't share my general skepticism as to global reductionist programs see the interest of such local projects as an illumination of what rests on what in mathematics. To the extent that various parts of mathematics are represented by formal systems, that comes down to considering relations of reduction between such systems.

2 Plan of the paper.

There are three main kinds of local reduction relations S$\leq$T between formal systems S and T that have been dealt with in the metamathematical literature: *relative interpretation, translation*, and *proof-theoretic reduction*. It is this last relation which is basic to reductive proof theory. Informally, the idea for it is that one has an effective (in practice, primitive recursive) method for passing from proofs in S of formulas ϕ in a distinguished class of formulas Φ to proofs in T of the same ϕ, and that this is established in a third system U which is considered to be privileged in some sense or other. In words, we say that S *reduces to*T *preserving* Φ, *provably in* U; in symbols, the relation is written here as S$\leq$T [Φ] (in U). When this holds, S *is a conservative extension of*T *for the formulas in*Φ; with false statements such as $0 = 1$ taken to belong to Φ, this insures that S *is relatively consistent to*T (provably in U). In practice, U can be taken to be a quite weak system such as Primitive Recursive Arithmetic PRA, or its conservative extension $I\Sigma_1$, the subsystem of Peano Arithmetic PA based on the Σ_1 Induction Axiom. The Appendix to this paper provides comparative examples, definitions, and basic properties of the three reduction relations in question. It concludes with a defense, against the challenges raised in (Niebergall 2000), of proof-theoretic reduction as a "good" reducibility relation. The details of the Appendix are not needed for appreciation of the main question of this paper, as to the viability of rationales for reductive proof theory. Before getting into that, we begin in sec. 3 with a sketch of how reductive proof theory, with its essential use of ordinals, works in practice. Two rationales for reductive proof theory are then taken up, the constructive consistency-proof rationale in sec. 4 and the foundational reduction rationale in sec. 5. It is the recent advances in proof theory that force one to consider the viability of these rationales; these are taken up in sec. 6. Despite the genuine problems raised for current rationales by that work, the body of the paper concludes in sec. 7 with a defense of reductive proof theory as one of the principal means to lay out *what rests on what* in mathematics, at least in its "everyday" parts.

3 The role of ordinals in reductive proof theory.

Gerhard Gentzen's work in the 1930s (see Gentzen 1969) has been the most influential for the development of modern reductive proof theory in practice. That made its major strides beginning in the 1950s, especially through the efforts of Kurt Schütte and Gaisi Takeuti, whose work is summarized in their treatises (Schütte 1960, 1977) and (Takeuti 1975, 1987). The aim of the Gentzen-Schütte-Takeuti line of development is what I call the *constructive consistency-proof rationale for reductive proof theory*, which I will take up in the next section. Since Gentzen's demonstration of the consistency of arithmetic by transfinite induction on an ordering of order type Cantor's ordinal ε_0, that has led to an enormous amount of proof-theoretical work of a prima-facie reductive character in which ordinals (or, more generally, well-founded orderings) play a central role. On the other hand, it is not immediately obvious how that work in practice relates to the notions of proof-theoretical reduction treated in Appendix A.4; the purpose here is to outline how these fit together. As far as the mechanics of the work go, the emphasis will be on the approach that grew out of the Schütte school; for a recent survey of that, see (Pohlers 1998). Due to recent work of Buchholz (1997, 1999) the approaches of the Schütte and Takeuti schools are much more closely related than had been thought for a long time, so this restriction is not so one-sided.[3]

The general pattern is as follows. Given a formal system S which one wishes to reduce proof-theoretically to more justified or privileged principles in some sense or other, one first embeds S in a system S^* which is more amenable to proof-theoretical transformations and analysis than S . Since Gentzen, this means that S^* is formalized in a sequent calculus, and since Schütte, this means that S^* may involve infinitary rules of inference. The paradigm is given by the system PA of Peano Arithmetic for S, which is embedded in an extension S^* of Gentzen's calculus LK for the classical first-order predicate calculus by a form of the ω-rule. This first step provides an effective map from proofs in S to proofs (in general infinite) in S^*,

$$(1) \qquad (\ )^* : \mathrm{Proof}_{\mathrm{S}} \to \mathrm{Proof}_{\mathrm{S}^*}.$$

The cut-rule in sequent calculi,

$$(2) \qquad \text{from } \Gamma \vdash \triangle, \phi \text{ and } \phi,\ \Gamma' \vdash \triangle' \text{ infer } \Gamma, \Gamma' \vdash \triangle, \triangle',$$

like the rule of modus ponens in Frege-Hilbert style calculi, involves potential "detours" through formulas ϕ of greater complexity than may appear in the concluding formula(s) of a proof. Gentzen showed how cuts could be eliminated completely in LK, but only partially in his sequential formulation of arithmetic with the usual logical rules and a rule of induction; Schütte showed how complete cut-elimination could be restored in his infinitary extension of LK. In general, one has an operation

$$(3) \qquad \rho : \mathrm{Proof}_{\mathrm{S}^*} \to \mathrm{Proof}_{\mathrm{S}^*},$$

from proofs in S^* to new proofs in S^* which reduces their over-all logical complexity by eliminating cuts as far as possible. Then for proofs p in S of certain formulas ϕ of a particularly simple form, the proof $\rho(p^*)$ can be analyzed to show that ϕ can be proved by more direct means than may be apparent in S. Taking Φ to be such a class of formulas, we appear in this way to be moving in the direction of establishing the proof-theoretic reduction of S to a system T, conservatively for Φ. But the choice of T is not yet clear from this.

Here is where ordinals come into the picture. What Gentzen had done in his consistency proof of PA was assign to each proof p an ordinal $|p| < \varepsilon_0$

such that if p is a proof of $0 = 1$ then $|\rho(p)| < |p|$; hence the consistency of PA follows by transfinite induction up to ε_0, because otherwise its ordering would not be well-founded. In Schütte's version, one assigns ordinals simply as lengths to the infinite proof figures in PA^*, and it is shown that if p is a proof in PA then $|\rho(p*)| < \varepsilon_0$. Moreover, if p ends in a closed equation $s = t$ then transfinite induction up to $|\rho(p*)|$ shows the equation to be true. There is a natural primitive recursive ordering $\prec_{\varepsilon_0}$ of order type ε_0 and, except for transfinite induction on that ordering, Gentzen's argument can be carried out in the system PRA of primitive recursive arithmetic. That can also be achieved via the primitive recursive representation of the infinite proof-figures and operation ρ on them in Schütte's argument.[4] But the latter tells us a bit more (though this is also implicit in Gentzen's work): it gives us a proof-theoretic reduction of PA to PRA supplemented by the scheme of transfinite induction for each initial segment of the natural ordering of type ε_0. Using $\mathrm{TI}(\alpha)$ to indicate the scheme of transfinite induction up to the initial segment of type α in the given ordering (applied to suitable formulas depending on the context)[5], we symbolize this by the reduction in the sense of sec. 4,

$$\text{(4)} \qquad \mathrm{PA} \le \mathrm{PRA} + \{\mathrm{TI}(\alpha)\}_{\alpha<\varepsilon_0},$$

preserving equations, provably in $\mathrm{I}\Sigma_1$. Since $\mathrm{TI}(\varepsilon_0)$ proves the consistency of the system to the right, this proof theoretic reduction also gives

$$\text{(5)} \qquad \mathrm{I}\Sigma_1 \vdash \mathrm{TI}(\varepsilon_0) \to \mathrm{Con}_{\mathrm{PA}}.$$

Then, as Gentzen showed, that is best possible in ordinal terms, since PA proves transfinite induction up to each ordinal $\alpha < \varepsilon_0$.

All of this is a paradigm for the general pattern in subsequent reductive proof-theory, using infinite proof figures in an extended calculus S^* of sequents depending on the choice of S. One associates ordinals as lengths to the proofs in S^*. Then proofs p in S lead to proofs p^* in S^* which can be reduced in complexity by (possibly partial) cut-elimination to $\rho(p^*)$. This leads to a computation of the least upper bound α_{S} of the ordinals $|\rho(p*)|$, and an associated natural representation by a primitive recursive well-ordering $\prec_{\alpha_{\mathrm{S}}}$ of order type α_{S}. This leads to a proof-theoretic reduction in the sense of sec. 4, of the form

$$ (6) \qquad \mathrm{S} \leq \mathrm{PRA} + \{\mathrm{TI}(\alpha)\}_{\alpha < \alpha_{\mathrm{S}}}, $$

preserving equations, provably in $\mathrm{I}\Sigma_1$, and then

$$ (7) \qquad \mathrm{I}\Sigma_1 \vdash \mathrm{TI}(\alpha_{\mathrm{S}}) \rightarrow \mathrm{Con}_{\mathrm{S}}, $$

where the transfinite induction is applied to a primitive recursive property. To show this is best possible, one establishes $\mathrm{TI}(\alpha)$ in S for each $\alpha < \alpha_{\mathrm{S}}$. Finally, when the passage from p to p^* preserves formulas in a class Φ, one may seek a familiar system T which is prima-facie weaker than S, in which $\mathrm{TI}(\alpha)$ can be established for suitable formulas (depending on Φ) and each $\alpha < \alpha_{\mathrm{S}}$, and for which

$$ (8) \qquad \mathrm{S} \leq \mathrm{T}[\Phi](\text{in } \mathrm{I}\Sigma_1). $$

In the current work on *ordinal analysis* of formal systems, as surveyed e.g. in (Pohlers 1998), the emphasis is on determining the least ordinal α_{S} satisfying (7) or related criteria, such as being the least (primitive recursive) ordinal not provably a well-ordering in S. In the practice of ordinal analysis, reductive concerns are relegated to the background, if not set aside altogether. This does, at least, lead to an ordering according to ordinal strength of those formal systems for which an ordinal analysis has been obtained. We shall return to the issues raised by ordinal analysis in sec. 6.

4 The constructive consistency-proof rationale for reductive proof theory.

Hilbert's program was motivated by the view that the "actual infinite" in mathematics is problematic, leading in some cases to contradictions. The program aimed to justify various parts of mathematics that make implicit or explicit use of principles based on the actual infinite, by representing them in formal systems which would be shown to be consistent by purely finitistic arguments. Patently, the use of set theory and even of impredicative principles (such as that of the least upper bound) in analysis would require such justification on this view. But, according to Hilbert, already arithmetic (as represented in the system PA) makes use of the actual infinite in its application of classical logic to statements involving quantification over the natural numbers. That is already seen in the assumption of the law of excluded middle for statements of the form $\forall x R(x)$ with R quantifier-free; such cannot in general be decided in a finite number of steps—one must "run through" the totality of natural numbers to determine their truth or falsity. The Hilbert school did not delimit finitist proofs by defining finitist mathematics in terms of a formal system. In practice, at least in the early stages, it did not go beyond PRA, and that has been argued by (Tait 1981) to be the upper limit of finitism, a thesis which is largely accepted these days. In any case, whatever formal system S_0 would be determined to represent finitism, Gödel's second incompleteness theorem showed that one would not be able to establish the consistency of systems stronger than S_0 by the means available in S_0. In other words, to continue the consistency program, the idea of a privileged basis for that kind of justification of all of mathematics would have to be abandoned, and would have to be replaced by a shifting basis on some other sort of constructive principles.

In Gerhard Gentzen's groundbreaking 1936 article "Die Widerspruchsfreiheit der reine Zahlentheorie" (referred to here through its English translation in Gentzen 1969, pp. 132-201), it was shown how, in the case of arithmetic, the consistency program might be extended while hewing to finitist principles as closely as possible. Gentzen's paper contains several sections discussing the aims and significance of his consistency proof—besides its extensive technical work whose general character was indicated in sec. 3 above. Under the heading, "How are consistency proofs possible?", Gentzen says: "*There can be no 'absolute consistency proof'.* A consistency proof can merely *reduce* the correctness of certain forms of inference to the correctness of other forms of inference. ... in a consistency proof we can use only forms of inference that count as considerably *more secure* than the forms of inference of the theory whose consistency is to be proved."(op. cit., p. 138) Gentzen then goes on to say that because of Gödel's incompleteness theorem, it is not possible to establish the consistency of arithmetic using a part or all of the methods used in that system, but: "[it] remains quite conceivable that the consistency of elementary number theory can in fact be verified by means of techniques which, in part, no longer belong to elementary number theory, but which can nevertheless be considered to be *more reliable* than the doubtful components of elementary number theory itself."(op. cit., p. 139) He later argues (op. cit., pp. 193ff) that his entire proof is finitistic except possibly for the application of transfinite induction up to ε_0, and that inspection of the argument for that principle shows it to be "*indisputable*", in contrast to the "transfinite" principles of the formal system of number theory.

In the further pursuit of the consistency program for analysis and various of its subsystems, Schütte and Takeuti provided similar rationales for their work. Schütte's explanation of the rationale for his proof-theoretical work is that "Gödel's investigations (1931) have shown that the strictest finitist methods are basically inadequate for carrying out the consistency proof required by Hilbert's programme. So proof theory needs not only the very strict finitist methods of a combinatorial nature but also higher level proof procedures. Thus we arrive at methods, first used by Gentzen (1936), using induction which in fact goes beyond the usual complete (mathematical) induction but still has a constructive character... We use inductive methods for the consistency proofs but do not admit *Tertium non datur* as a proof procedure." (Schütte 1977, p. 3) The consistency-proof rationale was explained as follows by Takeuti (1987), p.101: "We believe that our standpoint is a natural extension of Hilbert's finitist standpoint, similar to that introduced by Gentzen, and so we call it the Hilbert-Gentzen finitist standpoint. Now a Gentzen-style consistency proof is carried out as follows: (1) Construct a suitable standard ordering, in the strictly finitist standpoint. (2) Convince oneself, in the Hilbert-Gentzen standpoint, that it is indeed a well-ordering. (3) Otherwise use only strict finitist means in the consistency proof." Takeuti then goes on (loc. cit.) to explain what is supposed to be admitted under (2): these are "concrete" constructive methods, in contrast to those admitted to intuitionism, which calls on abstract notions of proof and construction.

In summary, the Gentzen-Schütte-Takeuti modified form of the consistency program, that I shall call the *extended Gentzen program*, comes down to carrying out the following three things:

(1) Describe *finitistically* the ordering relation of a notation system for ordinals up to an ordinal α_{S}.

(2) Give a *finitist* proof that the principle of transfinite induction up to α_{S}, $\mathrm{TI}(\alpha_{\mathrm{S}})$, implies the consistency, $\mathrm{Con}_{\mathrm{S}}$, of S.

(3) Give a *constructive proof* of (the instances used in (2) of) $\mathrm{TI}(\alpha_{\mathrm{S}})$.

The first obvious criticism to be made of the extended Gentzen program is that the notions of finitist and constructive proof required for it are vague. In particular, there are many varieties of constructivism, which on the one hand do not always square with each other, and on the other hand are not simply characterized by saying that one uses only inductive methods without the law of excluded middle. However, I think we can be charitable to an extent here, for two reasons. First of all, one can inspect specific executions of the program and, for the most part, see that the arguments employed in (1) and (2) are prima-facie finitist, while those employed in (3) are prima-facie constructive. Secondly, we now have a great deal of knowledge of formal systems which explicate finitism and constructivism in their various forms, with respect to which we can locate more precisely what a given execution of the program succeeds in doing. Still, one can anticipate that questions will have to be raised in borderline cases, as we shall see in sec. 6.

A second obvious criticism is that there is no reason given for the asymmetry of methods in (1), (2) as against those in (3). If one is to admit constructive proofs in (3), why not allow constructive definitions and proofs in (1) and (2), not just those that are finitist? And, if one takes that step, why not consider quite different constructive foundations, such as the simple reduction of classical to intuitionistic arithmetic (i.e. PA to HA) by the Gödel translation? (Interestingly, Gentzen describes the translation (1969, pp. 169-170), but doesn't make an argument why that is insufficient for his aims.)

But the main point of criticism of the extended Gentzen program, as of the Hilbert program which it modifies, is the criterion of consistency itself as its be-all and end-all. This had its origin in Hilbert's early identification[6] of the "existence" of mathematical concepts with the consistency and completeness of axiom systems for them. Later, in the mature formulation of his program, Hilbert only emphasized the consistency criterion in service of an instrumentalist justification of formal systems, though he still presumed that completeness would also be established in the cases of interest. The idea was to eliminate the "ideal" statements of a system in favor of the "real" statements, which we can identify with Π^0_1 formulas (treated as open statements). Indeed, for systems S containing a modicum of arithmetic, if S is consistent, then every Π^0_1 statement provable in S is valid. It was Brouwer who first objected that consistency is insufficient to guarantee "correctness" in some intuitive interpretation. Then Gödel's incompleteness theorems bore out this criticism with the construction of a consistent system extending PA which is not valid in the natural numbers (namely $\mathrm{PA} + \neg\mathrm{Con}_{\mathrm{PA}}$).

The most vocal critic of the consistency criterion (in numerous essays) has been Georg Kreisel, saying for example in a late survey of his own work that "I was repelled by Hilbert's exaggerated claim for consistency as a sufficient condition for mathematical validity or some kind of existence" (Kreisel 1987, p. 395). Kreisel aimed instead to use proof theory to make "explicit the additional knowledge provided by those proofs." (loc. cit.) More explicitly, he sought to "unwind" mathematical proofs on the one hand and to provide "general formal criteria such as functional interpretations to replace the incomparable condition of consistency; 'incomparable' because the aim of functional interpretations is meaningful without restriction on metamathematical methods." (loc. cit.)[7]

Of course, consistency itself is meaningful without restriction on metamathematical methods, and one can point to systems of possible mathematical interest for which there may be a genuine question as to their consistency, e.g. Quine's system NF, or PA + ¬TP, where TP is the Twin Prime conjecture, or—more ambitiously—PA_2 + ¬TP, where PA_2 is full 2nd order arithmetic, or—still more ambitiously—ZF+¬TP. (One may substitute for TP here any currently open problem in number theory, such as Goldbach's Conjecture (GC) or the Riemann Hypothesis (RH), that is strongly suspected of being true but difficult to prove.[8]) But what about the consistency of PA and PA_2 and ZF? The most advanced current work in proof theory that may contribute to the extended Gentzen program hardly reaches beyond the subsystem Π^1_2-CA of $PA_2 (= \Pi^1_\infty-CA)$. I, for one, have absolutely no doubt that PA and even PA_2 are consistent, and no genuine doubt that ZF is consistent, and there seems to be hardly anyone who seriously entertains such doubts. Some may defend a belief in the consistency of these systems by simply pointing to the fact that no obvious inconsistencies are forthcoming in them, or that these systems have been used heavily for a long time without leading to an inconsistency. To an extent, those kinds of arguments apply to NF, which has been studied and worked on by a number of people. My own reason for believing in the consistency of these systems is quite different. Namely, in the case of PA, we have an absolutely clear intuitive model in the natural numbers, which in the case of PA_2 is expanded through the notion of arbitrary subset of the natural numbers. Finally, ZF has an intuitive model in the transfinite iteration of the power set operation taken cumulatively. This has nothing to do with a belief in a platonic reality whose members include the natural numbers and arbitrary sets of natural numbers, and so on. On the contrary, I disbelieve in such entities. But I have as good a conception of what arbitrary subsets of natural numbers are *supposed* to be like as I do of the basic notions of Euclidean geometry, where I am invited to conceive of points, lines and planes as being utterly fine, utterly straight, and utterly flat, resp. What is *not* evident on the latter conception without special work is the consistency of the system of Euclidean geometry with the parallel axiom replaced by its negation. Similarly, while the notion of arbitrary set and the cumulative hierarchy argues for believing straight off not only in the consistency of ZF but also of ZFC(=ZF+AC), much additional work had to be done to establish the consistency of ZFC+¬CH (namely by Paul Cohen's method of forcing). To return, for example, to NF, that has no intuitive model to support our direct belief in its consistency, and the problem of establishing such a result, if it is to be established at all, will no doubt require special metamathematical work, for which restriction in advance to constructive methods would be irrelevant. But if, say, we find out that $\mathrm{ZFC} \vdash \mathrm{Con}_{\mathrm{NF}}$ and we accept the consistency of ZF then we must accept the consistency of NF, since $\mathrm{Con}_{\mathrm{NF}}$ is a Π^0_1 statement.

5 The foundational reduction rationale for reductive proof theory.

Another modification of Hilbert's program was suggested by Kreisel (1958, 1968) and has been further expanded and pursued by me in (Feferman 1988, 1993). Since quite a bit of detail is supplied in the latter papers, I shall content myself here with as brief an explanation as possible for present purposes. In the 1993 paper, "What rests on what? The proof-theoretic analysis of mathematics", I distinguished four senses in which we can deal, from a logical point of view, with the question posed there, by reference to the following categories: an informally developed body of mathematics $\mathcal{M}$, a formal language L, formulas ϕ of L, a formal axiomatic system T in L, and a general foundational framework $\mathcal{F}$ (such as that of finitism, constructivism, set-theoretic platonism, etc.). The four senses are:

(i) $\mathcal{M}$ *rests on* T, in the sense that $\mathcal{M}$ *can be formalized in* T;

(ii) ϕ *rests on* T, in the sense that ϕ *can be proved in* T;

(iii) T *rests on* $\mathcal{F}$, in the sense that T *is justified by* $\mathcal{F}$; and

(iv) T_1 *rests on* T_2, in the sense that T_1 *is reducible to* T_2.

Any of the notions of reduction for formal systems, such as those discussed in the Appendix, could be taken in (iv), but here our main concern is with proof-theoretic reductions. All but (ii) (perhaps the most common of the four senses) are combined in the *foundational reduction rationale* for reductive proof theory, which was formulated in general terms as follows in my 1988 paper (p. 364):

> A body of mathematics $\mathcal{M}$ is represented in a formal system T_1 which is justified by a foundational or conceptual framework $\mathcal{F}_1$. T_1 is reduced proof-theoretically to a system T_2 which is justified by another, more elementary such framework $\mathcal{F}_2$.

Note that in each case it is only expected that T_i, and thence $\mathcal{M}$, represents a *part* of what can be justified by $\mathcal{F}_i$, so what a given proof-theoretic reduction achieves under this rationale is only a *partial foundational reduction.* Hilbert's program called for the reduction of formal systems T_1 justified by an infinitary framework $\mathcal{F}_1$ to systems T_2 justified by the finitary framework $\mathcal{F}_2$; the general foundational pair in this case is briefly indicated here by $\langle$infinitary, finitary$\rangle$. Because of the limitations due to Gödel's incompleteness theorems as to how far Hilbert's program can be carried, other pairs of foundational frameworks are considered under the above rationale, such as $\langle$impredicative, predicative$\rangle$ and $\langle$non-constructive, constructive$\rangle$. But one can also in some cases reduce systems which require for their direct justification the uncountable infinite in some form or other (e.g. if they have quantified variables ranging over "arbitrary" sets of natural numbers) to those that do not require such, but are still infinitary in their justification (e.g. systems of arithmetic in quantificational logic); this distinction leads to the pair $\langle$uncountable infinitary, countable infinitary$\rangle$. Since these various categorizations are vague to a certain extent, when judging whether a given proof-theoretic reduction accomplishes a foundational reduction under the above rationale, one seizes on the most obvious features of the foundational frameworks involved, so that "insofar as possible, what the work achieves will speak for itself" (op. cit., p. 367).

One could construe the extended Gentzen program as falling under the rationale for the pair $\langle$infinitary non-constructive, constructive$\rangle$. But the requirements of that program as described in sec. 4 (1)–(3) above are rather specialized, involving as they do a mix of reductions to finitism and constructivism as well as the essential role of ordinals. In the papers (Feferman 1988, 1993), I have referred to the foundational reduction rationale for reductive proof theory as a *relativized form of Hilbert's program*, and that seems to me to be a more apt description than for the extended Gentzen program with its specialized character. Note that no requirements are made as to how the proof-theoretic reductions are to be carried out under this rationale. They could be achieved by cut-elimination techniques accompanied by ordinal analysis as described in sec. 3, or by functional interpretations, or by other proof-theoretical methods.

Here, briefly, for purposes of illustration, are some examples of known proof-theoretic reductions which provide foundational reductions under the above rationale; details, many more examples and references to the reductive work involved are given in the aforementioned papers.

5.1 Reductions of the infinitary to the finitary

Here the familiar example is that provided by Parsons' reduction

(1) $$\mathrm{I}\Sigma_1 \leq \mathrm{PRA}.$$

Later this was improved (following model-theoretic conservation results by Harvey Friedman) by Wilfried Sieg to the proof-theoretic reduction

(2) $$\mathrm{RCA} + \mathrm{WKL} + \mathrm{I}\Sigma_1 \leq \mathrm{PRA},$$

where RCA is the recursive comprehension axiom and WKL is the so-called Weak König's Lemma given by restriction of KL to binary branching trees. Note that (2) accomplishes a foundational reduction $\langle$uncountable infinitary, finitary$\rangle$.

5.2 Reductions of the uncountable infinitary to the countable infinitary.

As explained above, systems whose language permits quantification over variables for sets of natural numbers require, on the face of it, the uncountable infinitary framework, while arithmetic only requires the countable infinitary framework. The classical example here is

$$(3) \qquad \mathrm{ACA}_0 \leq \mathrm{PA},$$

where ACA_0 is the system based on the arithmetical comprehension axiom with restricted induction (indicated by the sub '0'). Following model-theoretic conservation results by Barwise and Schlipf and, independently, Friedman, (3) was strengthened by Feferman and Sieg to the proof-theoretic reduction

$$(4) \qquad (\triangle^1_1 - \mathrm{CA})_0 \leq \mathrm{PA},$$

where the subsystem of analysis on the left is based on the $\triangle^1_1$ (or "Hyperarithmetic") Comprehension Axiom.

5.3 Reductions of the impredicative to the predicative.

Systems of (unramified) analysis employing instances of the comprehension axiom in which set variables occur bound are prima-facie impredicative. The characterization of predicativity via autonomous transfinite progressions of ramified systems obtained independently by Feferman and Schütte in 1964, arrived at Γ_0 as the least impredicative ordinal, where Γ_0 is the least fixed point α of $\phi_\alpha(0) = \alpha$, and where $\langle \phi_\xi \rangle$ is the Veblen hierarchy of critical functions. Since $\phi_0(\beta) = \omega^\beta$, we have $\phi_1(\beta) = \varepsilon_\beta$; in particular, $\varepsilon_0 = \phi_1(0)$, hence $\varepsilon_0 < \Gamma_0$. The ramified analytic progression up to ordinal α is equivalent to ACA iterated α times, so the following proof-theoretic reduction obtained by Feferman and Sieg (again after a related model-theoretic conservation result due to Friedman), illustrates the reduction ⟨impredicative, predicative⟩:

$$(5) \qquad \triangle^1_1 - \mathrm{CA} \leq (\mathrm{ACA})_{<\epsilon_0}.$$

In both systems here, induction is unrestricted, and in the system on the right, ACA is iterated α times for each $\alpha < \epsilon_0$.

5.4 Reductions of the non-constructive to the constructive.

Here the paradigm is given by Gödel's translation of classical Peano Arithmetic into Heyting's intuitionistic arithmetic:

(6) $$\mathrm{PA} \leq \mathrm{HA}.$$

That translation in general works to translate classical systems T into systems $\mathrm{T}^{(i)}$ based on intuitionistic logic, but the translations of the axioms of T need not be constructively valid. Thus other proof-theoretic reductions, often heavy-duty, become necessary. An example is provided by:

(7) $$\triangle^1_2 - \mathrm{CA} \leq \mathrm{ID}^{(i)}_{<\epsilon_0},$$

where the intuitionistic system on the right provides for the iteration of accessibility inductive definitions any number $\alpha < \epsilon_0$ times, and is thus constructive. The result (7) is due to a long chain of work by Friedman, Feferman, Pohlers and Buchholz, and is described in the introduction to (Buchholz et al 1981).

5.5 Discussion.

Note first that we have not said anything about what part of mathematics $\mathcal{M}$ may be formalized in the system T_1 which is being reduced in each of the preceding examples. This requires case studies in each case, some indication of which can be found in the final section of (Feferman 1993); much relevant information has been obtained through the so-called Reverse Mathematics (R.M.) program, which is exposited in (Simpson 1998). Among the claims of my 1993 paper as well as earlier papers reproduced in (Feferman 1998) is that all scientifically applicable mathematics can be directly formalized in systems which are proof-theoretically reducible to PA; this is justified by a number of case studies. In fact, more refined research in the R.M. program as well as by Feng Ye (1999) has shown that substantial portions of that can already be carried out in systems proof-theoretically reducible to PRA. The philosophical significance of this is that the part of mathematics that is so far indispensible to science rests on completely arithmetical grounds, and in fact to a large extent on finitary grounds. To that extent at least, Hilbert's original aims are realized (cf., in this respect, Simpson 1988). In general, investigating what part of mathematics $\mathcal{M}$ can be formalized in a given reducible system T_1 carries the foundational reduction achieved over to mathematical practice, so that one can say that if $\mathrm{T}_1 \leq \mathrm{T}_2$ and T_2 rests on the foundational framework $\mathcal{F}_2$, then so also does $\mathcal{M}$. Thus we can say, that $\mathcal{M}$ ultimately rests on finitary, countably infinitary, predicative, or constructive grounds, as the case may be.

The main criticism to be made of the foundational reduction rationale described in this section is basically the same as the first criticism made of the extended Gentzen rationale made in sec. 4. Namely, it makes use of vague notions of being finitary, infinitary (countably and uncountably), predicative and constructive, among others, and the rationale can only be considered to be successfully met in specific cases when there is no question as to which of these notions applies. One can be charitable about this to an extent for the reasons given in sec. 4, but again there will be genuine problems about borderline cases, as we see next.

6 Are the rationales for reductive proof theory applicable to recent advances in ordinal analysis?

We are here entering an area which is unsettled in many respects. The most advanced work in proof theory involving ordinal analysis is that due to Michael Rathjen (1995). Related work in the style of the Takeuti school has been carried out by Toshiyasu Arai (1996, 1997). The discussion here is confined to the former, mainly because that is material with which I am more familiar.[9] One thing that is unsettled is that the article in question—(Rathjen 1995)—is a report, and full details of notions and proofs have not yet appeared, though work is in progress to provide a complete presentation of this very complicated material. For our purposes, though, the report suffices. The primary thing that is unsettled is the significance of that work vis-à-vis the rationales for reductive proof theory that have been discussed in this paper. Before getting into that, let me step back to relevant earlier work of Rathjen (1991). I shall not try to explain in full the various systems and principles involved; the reader not familiar with them will have to rely on the references to follow.

In Rathjen's 1991 paper, an ordinal analysis was given of the system KPM obtained by extending Kripke-Platek set theory KP (based on classical logic) by an axiom for a recursively Mahlo universe of sets. An admissible ordinal α is recursively Mahlo if for each α-recursive function f there exists an admissible ordinal $\beta < \alpha$ with β closed under f. For the corresponding universe L_α, the appropriate axiom is given by the Π_2 Reflection Principle (Π_2-Ref). The crucial question for the extended Gentzen rationale is whether a well-ordering proof for the (primitive) recursive ordinal representation system used in the ordinal analysis of KPM can be given constructively. The question for the foundational reduction rationale is whether KPM can be reduced proof-theoretically to a system T based on a more elementary framework than that required to justify KPM. It's not clear to me what framework justifies KPM that does not already justify ZF without the power axiom, but at any rate both require non-constructive infinitary principles, and so the first question here would be whether we can reduce KPM proof-theoretically to a constructive system T. Since the two rationales are related in practice, we consider only the question of reduction here. There are three candidates for such T, each obtained as an extension of a basic system in the literature which has been argued to be constructive, namely the constructive theory of types ML of (Martin-Löf 1984), the constructive theory of sets CZF introduced in (Aczel 1978), and the system of explicit mathematics T_0 in its intuitionistic version $T_0^{(i)}$ introduced in (Feferman 1975, 1979). For each of these systems S, an extension T by an axiom of Mahlo character has been proposed and the relation with KPM has been explored, and for each one there are then two questions to be asked:

1° Is KPM reducible to T?

2° Is T evidently constructive?

The situation concerning candidate extensions MLF and CZF + (a Mahlo rule) of ML and CZF, resp., is reported in (Rathjen 1998). It is shown there that each of these is proof-theoretically equivalent to $\mathrm{KPM}^{(r)}$, which is KPM taken with the foundation scheme restricted to $\triangle_0$ formulas. It seems to be generally agreed that MLF is constructive on the same basis as ML and hence that $\mathrm{KPM}^{(r)}$ is proof-theoretically reducible to a constructive system. As for *full* KPM, the situation is more delicate; in the final section of the just cited paper, Rathjen conjectures that it is proof-theoretically reducible to an extension of ML using "higher order universe operators" proposed by Erik Palmgren. His view is that this extension "is still in the vein of Martin-Löf's original papers," but that this is the limit of what can be justified by the kind of informal semantics with its pattern of introduction and elimination rules required by (Martin-Löf 1984). Anton Setzer (1998) has proposed an extension of ML denoted MLM, which is stronger than KPM, and both Martin-Löf and Setzer have argued for the constructivity of MLM. But at the end of his 1998 paper, Rathjen points out that "Setzer's Mahlo universe is generated by a non-monotonic inductive definition which is incompatible with elimination rules for the universe." He concludes that "[MLM] means a paradigm shift to a *new* Martin-Löf type theory. Therefore KPM is only a boundary for *old* Martin-Löf type theory." In other words, the arguments for constructivity of MLM do not make it *evidently* constructive in the sense that one had previously come to understand constructive type theory. Concerning this, Setzer has written me that "the Mahlo universe is constructively justified, as documented by meaning explanations... Since I have found them, I don't think any more about Mahlo and regard it as absolutely unproblematic – the [constructive] consistency of KPM is known." These "meaning explanations," which apparently do not require elimination rules of ML style, are as yet unpublished. On the same grounds, Setzer claims the constructive reducibility of theories far stronger than KPM, though still well short of KP + (Σ_1−Sep) to be discussed below.

The situation concerning extensions of the system of explicit mathematics $T_0^{(i)}$ in relation to KPM was up in the air until recently. In (Jäger and Studer 1999) Gerhard Jäger has proposed a Mahlo style axiom (M) which is very natural in the context of explicit mathematics. It provides for a uniform operation m for passing from a pair (f, a) in which f takes classes to classes and a represents a class, to $m(f, a)$ which represents a universe containing a and which is closed under f. It is shown, op. cit., that the classical system T_0 + (M) is interpretable in KPM (incidentally via a non-monotone inductive definition), and the same holds for corresponding restricted versions. Sergei Tupailo has very recently announced results (Tupailo 2000) giving the interpretation of a system CZFM (containing a full version of the Mahlo axiom in the language of CZF) in $T_0^{(i)}$ + (M). It may be argued that the latter system is constructively justified on the same direct basis as for $T_0^{(i)}$, which *is* constructively justified.[10] I understand that the system CZFM in Tupailo's result is conjectured by Rathjen to be proof-theoretically equivalent to KPM. Thus it would follow that T_0 + (M) in both its intuitionistic and its classical versions is equivalent in strength to KPM.

The paper (Rathjen 1995) referred to at the beginning of this section concerns the system (Π^1_2−CA) of Π^1_2 comprehension in the language of 2nd order analysis, augmented by the scheme (BI) of Bar Induction; this is directly equivalent to the set-theoretic system KP augmented by the scheme (Σ_1-Sep) of Σ_1 Separation. Rathjen's ordinal analysis of KP + (Σ_1-Sep) is obtained by "slicing" (Σ_1-Sep) into transfinite degrees of reflection, thus going far beyond (Π_n-Ref) for each n. As a warm-up for this, Rathjen had earlier obtained an ordinal analysis of KP + (Π_3-Ref), which already required essentially new ideas beyond the treatment of KPM; see (Rathjen 1994). Whether *that* subsystem is constructively reducible is by no means clear, though again Setzer has argued for that on the same kind of grounds as for MLM above (personal communication). In any case, the status of the full system KP + (Σ_1-Sep) vis-à-vis our rationales for reductive proof theory is very much up in the air. Efforts are being made by Setzer (and perhaps others) to reduce it to some system of constructive character, but whether any such system will be *evidently* constructive remains to be seen; at the moment, there is little to encourage optimism in that respect. It is mainly for *this* reason that the question has been raised in the present paper whether reductive proof theory has a viable rationale.

One final remark, harking back to the discussion in sec. 4, is in order about all this work in relation to the extended Gentzen rationale. Even if one succeeds in reducing the system (Π^1_2-CA) $\pm$ BI to a constructive system (whether evidently so or not), one can hardly expect that doing so will appreciably increase one's belief in its consistency (if one has any doubts about that in the first place) in view of the difficulty of checking the extremely complicated technical work needed for its ordinal analysis.

7 Why reductive proof theory?

To conclude, I want to respond to questions as to the value of reductive proof theory which have been raised within the logical community. But first one must make a simple distinction, that could (and perhaps should) have been made much earlier. Recall the general explanation in sec. 2 of the relation of *proof-theoretic reduction* S $\leq$ T [Φ] (in U), where U = IΣ_1 or some other standard relatively weak base theory. Sections 3–6 have been entirely concerned with the use of this relation in *reductive proof theory*, which is the pursuit *via proof-theoretic reductions* of an extended Hilbert program in some sense or other, such as discussed in secs. 4 and 5 above. Now, *all* relations S $\leq$ T, including relative interpretations and translations, that have been established by logicians for specific systems in practice fall under the relation of proof-theoretic reduction for suitable Φ and U since, after all, the interpretation or translation has to be proved *somewhere*, and that somewhere is a prior accepted system U. With a little more fuss in each case, it may be seen that the relations $S \leq T$ in question are in fact subsumed under proof-theoretic reduction in the above strong sense. Thus, for example, all relative consistency results obtained in set theory by inner model constructions or forcing constructions can be seen to fall under this relation. Evidently, those cases have nothing to do with reductive proof theory as presently pursued, but the fact that the use of proof-theoretic reduction is common to both pursuits shows the clear value of this relation, as a general means of establishing intertheoretic reduction.

At any rate, with that distinction in mind, let us consider the kinds of questions that are often raised concerning the value of reductive proof theory:

1° Why reduce at all?

2° Why restrict to weak systems?

3° Why tie our hands?

In the background to these is the common view, sometimes referred to as *set-theoretic imperialism*, that set theory is the foundation of all of mathematics. Exactly what this means is not clear (and to the extent that it is clear is not, in my opinion, defensible[11]), but it is the general attitude that concerns me here, not any precise formulation. On this view, there is no reason to consider other foundational frameworks such as finitism, predicativism, constructivism, and so on, which—if accepted as the only source of legitimation—cast one out of "Cantor's Paradise" and apparently force one to jettison vast tracts of mathematics. However, if one can't swallow the kind of platonism required to justify infinitary, impredicative, non-constructive set theory[12], it is necessary to examine such alternative frameworks to see what philosophical arguments might be made for them and what sacrifices may be required if one is to live by them mathematically. The traditional way of seeing what can be achieved under such restrictions is to reconstruct mathematics from the ground up, while adhering to the informal principles of the preferred framework. The alternative offered by reductive proof theory is to formalize various parts of mathematics in subsystems of set theory and see which of these can be reduced to systems justified by one or another of these frameworks $\mathcal{F}$. This way of proceeding is non-committal to which such $\mathcal{F}$ is to be preferred, and leads more quickly to a survey of what parts of mathematics can be reconstructed, at least in principle, on the grounds of $\mathcal{F}$. (In doing so, the reductive proof-theorist may face, coming from the other side, criticism from the committed advocate of $\mathcal{F}$, who will say that it is only what can be explicitly worked out under the principles of $\mathcal{F}$ that is of interest to him or her. Well, you can't satisfy everybody.)

As has already been indicated in sec. 5.5, what has emerged, for example, from both the directly reconstructive and the reductive work is that substantially all of scientifically applicable mathematics can be formalized in systems which are predicatively reducible (in fact all the way down to PA) and considerable parts of that—e.g. the parts of functional analysis needed for quantum mechanics—can be grounded finitistically (as represented in PRA); see, for example, (Feferman 1993), (Simpson 1998, chs. III and IV) and (Ye 1999). On the other hand, interesting parts of pure mathematics which can be formalized in (Π^1_1-CA) definitely require impredicative methods; see, for example (Simpson 1998, ch. VI). In that respect then, one can see how one's hands would indeed be tied by restriction to the finitist or predicative frameworks, but not necessarily by the general constructive framework. In fact, as explained in (Buchholz, et al. 1981) one of the results of reductive proof theory is that (Π^1_1-CA) and many of its transfinite iterates are justified constructively by reduction to constructive theories of iterated accessibility inductive definitions. And evident constructive methods reach beyond these via the reduction obtained by (Jäger and Pohlers 1982) of ($\triangle^1_2$-CA) + BI to $\mathrm{T}_0^{(i)}$. As we have seen, it is currently an open question as to how much farther one can go constructively. From our present perspective, it is very likely that constructive methods taken in their widest possible sense do not reach far beyond (Π^1_2-CA) in strength if at all, and there are certainly many interesting results of descriptive set theory (for example) which are known to require substantially more. So, at that point the critic is certainly justified in raising question 3°. The devotee of finitism, predicativity, constructivity, or other such non-set-theoretical framework can simply answer, "So be it" and dismiss anything that cannot be justified by their preferred methods as being either meaningless or useless or both. A less ideological response to 3° might be that the bulk of "everyday"[13] pure and applied mathematics can be comfortably formalized in systems of strength well under that of (Π^1_2-CA), where the proof-theoretic reductions accomplish more or less clear foundational reductions.

Finally, one might point out that question 3° is loaded, and that trying to respond to it is to fall into the trap of accepting some truth in the veiled accusation. The pursuit of reductive proof theory does not by itself imply an intention to tie anyone's hands. Indeed, as I have stated elsewhere, after having pointed out the mathematical strength of relatively weak subsystems of analysis: "I am not by any means arguing that everyday mathematical practice should be restricted to working in such subsystems. The instrumental value of 'higher' and less restricted set-theoretical concepts and principles is undeniable. The main concern here is rather to see: *what, fundamentally, is needed for what*?" (Feferman 1999, p. 109) This, then, yields a rather simple response to the above questions which is not hinged to what might be considered vague and subjective matters of philosophy or foundational frameworks. Namely, as ever in the history of mathematics, one wants to see, dispassionately: *what rests on what*? And among the different senses of that recalled in sec. 5, reductive proof-theory has had a pre-eminent role to play in answering the question in its fourth sense there, i.e. as to *what systems are reducible to what other systems*. That argument for the value of reductive proof theory as one tool among others in helping to lay out the landscape of logical dependencies in mathematics allows one to ignore the loaded question 3°. At the same time, it bypasses the questions raised in the body of this paper about the viability of current rationales for reductive proof theory growing out of Hilbert's program. This foundationally neutral way of looking at the present and no doubt continued value of our work is fine as far as it goes. But the foundational dimension should not, by any means, be dismissed. All along, reductive proof theory has rightly been called on to do more, and it is just that *more* which has been put up for examination here.

A Appendix

A.1 Notions of local reduction between formal systems.

There are three main kinds of relations $S \leq T$ dealt with in the metamathematical literature between formal systems S and T. Each is induced by a specific kind of reduction in a sense to be explained; since they reduce one formal system to another, rather than all formal systems to a single one, they are candidates for local rather than global reductions. The three are: *relative interpretation, translation* and *proof-theoretic reduction.* In section A.2 we give some basic examples of each. The relations are then defined and examined in the successive sections, with proof-theoretic reduction taken up at length in section A.5. Two theorems are proved there which show that under rather general conditions the non-uniform version of this relation is equivalent to the uniform version. The Appendix concludes in section A.6 with an examination of the results of Niebergall (2000) which he considers to raise problems about proof-theoretic reduction as a "good" reducibility relation.

A.2 Examples of the three kinds of reduction.

(a) *Relative interpretation.* The following are familiar:

- (a_1) PA $\leq$ ZF, i.e. the interpretation of Peano Arithmetic in Zermelo-Fraenkel set theory;
- (a_2) ZF + AC + GCH $\leq$ ZF, via Gödel's model of ZF + AC + GCH in the constructible sets;
- (a_3) S $\leq$ PA + Con_S, for recursive S, by the Bernays-Wang formalization of Gödel's completeness theorem.

(b) *Translation.* Some examples here are:

- (b_1) PA $\leq$ HA, the "negative" translation of PA into Heyting's intuitionistic arithmetic;
- (b_2) HA $\leq$ PR^ω, Gödel's "Dialectica" translation of HA into a quantifier free theory of primitive recursive functionals of finite type;
- (b_3) IPC $\leq$ S4, by Gödel's translation of the intuitionistic propositional calculus (IPC) into Lewis' modal system S4.

(c) *Proof-theoretic reduction.* The following are standard examples:

- (c_1) BG $\leq$ ZF, the reduction of Bernays-Gödel theory of sets and classes to ZF;
- (c_2) ACA_0 $\leq$ PA, the system based on Arithmetical Comprehension Axiom with restricted induction, reduced to PA;

(c$_3$) Σ_1–IA $\leq$ PRA, the subsystem of PA based on Σ_1–Induction Axiom, reduced to the quantifier free system of Primitive Recursive Arithmetic.

We now turn to a more detailed examination.

A.3 Relative interpretation.

One has a relative interpretation of S into T if with each basic relation, function, and constant symbol of the language L_{S} of S is associated as its interpretation a definition of it in the language L_{T} of T, and with each sort s of variable in L_{S} is associated a defined range of variation given by a formula $\delta_s(x)$ in L_{T}; it is usually assumed that the equality relation is interpreted by itself. Then with each formula ϕ of L_{S} is associated as its interpretation in L_{T} a formula $f(\phi)$ obtained by substituting the respective definitions for the basic symbols and relativizing quantified variables of sort s to δ_s. For simplicity, we assume, as is usual in applications, that the basic symbols of S and T are each effectively specified. It follows that the function f, which is defined recursively, is an effective map of the formulas of L_{S} into the formulas of L_{T}; we write this property in the following as:

$$(1) \qquad f : L_{\mathrm{S}} \to L_{\mathrm{T}} \quad \text{is effective.}$$

This interpretation function f is then defined to constitute a relative interpretation of S in T, if we have

$$(2) \qquad \mathrm{S} \vdash \phi \Rightarrow \mathrm{T} \vdash f(\phi).$$

S$\leq$ T is defined to hold in the sense of relative interpretability if there is such an interpretation f. In that case, various “good” properties of T transfer to S. In particular, since f preserves negation, i.e.

$$(3) \qquad f(\neg\phi) = \neg f(\phi) \quad \text{for any formula} \quad \phi,$$

consistency transfers in this way, i.e.,

$$(4) \qquad S \leq \mathrm{T} \text{ and T consistent} \Rightarrow \text{S consistent.}$$

Finally, if T$'$ is any extension of T and $\mathrm{S}' := \{\phi \mid \mathrm{T}' \vdash f(\phi)\}$ then S$'$ $\leq$T$'$ and so if T$'$ is consistent so also is S$'$.

Though long in actual use—as the examples A.2(a) show—the precise notion of relative interpretation seems first to have been defined by Tarski in (Tarski, Mostowski and Robinson 1953), where it was applied as a general tool in proofs of undecidability of various systems. The basis is that (with f effective) if S is relatively interpretable in T and T is decidable then so also is S. Hence if S is undecidable then so also is T and if S is essentially undecidable (i.e. no consistent extension S′ of S is decidable) then the same holds for T.

The relation of interpretability also serves as a measure of strength of formal systems. It is reasonable to say that T is essentially stronger than S under the relation $\leq$ of relative interpretability if S $\leq$T, but not T$\leq$S. It may well be that for $L_S \subseteq L_T$, T is stronger than S in the sense of proving more statements than S, but is not essentially stronger than S. For example, I showed in (Feferman 1960) that PA + ¬ $\mathrm{Con}_{PA} \leq$ PA, though ¬ Con_{PA} is not provable in PA by Gödel's first incompleteness theorem. On the other hand, I was able to show there that PA + Con_{PA} is essentially stronger than PA in the sense of relative interpretability, and the same holds for any consistent r.e. extension of PA. These kinds of results, together with a proper general formulation of the formalized completeness theorem, A.2(a_3), led to a study of the relation of relative interpretability between systems for its own sake. For a recent survey, see (Lindström 1997).

A.4 Translation.

Here, it must be admitted that, despite the many examples in practice of the sort dealt with in A.2(b), there is no useful general theory which does not already assimilate the notion of translation to proof-theoretic reduction. First of all, there is no settled definition of what constitutes a translation; some proposals were floated by Wang (1951) and Kreisel (1955). The minimal assumptions seem to be that we have a function f satisfying the following (1)-(3), as for relative interpretations.

(1) $$f : L_S \to L_T \quad \text{is effective,}$$

(2) $$S \vdash \phi \Rightarrow T \vdash f(\phi), \text{ and}$$

(3) $$f(\neg\phi) = \neg f(\phi).$$

To make sense of these, as background assumptions we need that the languages L_S of S and L_T of T are effectively specified, that negation is defined for every formula of each of these languages, and that each is based on a system of logic with given axioms and rules of inference, in terms of which the provability relation is defined. Another assumption which might be made is that f preserves all the propositional operations. This makes sense only if every such operation on formulas of L_S is also an operation on formulas of L_T, and that they function logically in the same way. That requirement is not met, for example, when the logic of S is classical and that of T is intuitionistic, as in the "negative", or "double-negation" translation A.2(b_1) of PA into HA, for there we take $f(\phi \vee \psi) = \neg\neg(f(\phi) \vee f(\psi))$. Further, in Gödel's functional translation A.2(b_2) of HA into the quantifier-free theory PR^ω, propositional operations on quantified formulas are not preserved. So in various cases of interest that occur in practice, the additional assumption is not met. Moreover, even when such an additional assumption is made, one obtains trivializing results. Namely, it was shown by Pour-El and Kripke (1967) that for each consistent recursively enumerable system S in first-order classical logic there is a primitive recursive translation f of S into the weak subsystem Q of PA ("Robinson's system"), which satisfies (1)–(3) *and* preserves all propositional operations.[14]

For the kinds of translations of which A.2 (b_1) and (b_2) are illustrative, it is more reasonable to assume that for each system S, we have an atomic sentence $\bot_S$ of L_S such that S proves $\bot_S$ just in case S is inconsistent. In the case of systems (such as in those examples) that contain a modicum of arithmetic, we can simply take for the sentence $\bot_S$ the sentence $0 = 1$. Then in place of (3) as a condition on translations, we take

(3′) $$f(\bot_S) = \bot_T \, .$$

Now define $f : S \leq T$ to hold in the sense of translation if f is a function satisfying (1), (2) and (3′), and $S \leq T$ if there is such an f. Then consistency transfers from T to S just as for relative interpretability, so at least in that respect, translations are "good", i.e. we have

(4) $$S \leq T \text{ and T consistent} \Rightarrow S \text{ consistent.}$$

Of course, that would also hold if we took $S \leq T$ to mean that there is a translation f satisfying (1)–(3), or even one preserving all propositional operations. So that notion of translation also has the value of insuring relative consistency (4), and the Pour-El/Kripke result does not affect that value, since it constructs a trivializing translation only for S already assumed to be consistent.

A further assumption that is reasonable to make on translations that squares with practice, and which brings us closer to proof-theoretic reduction, is that f extends to an effective map f^* from proofs in S to proofs in T (however these are represented) in such a way that

$$(2') \qquad Proof_{\mathrm{S}}(p, \phi) \Rightarrow Proof_{\mathrm{T}}(f^*(p), f(p)).$$[15]

In particular, combined with (3′), this has the consequence

$$(3^*) \qquad Proof_{\mathrm{S}}(p, \perp_{\mathrm{S}}) \Rightarrow Proof_{\mathrm{T}}(f^*(p), \perp_{\mathrm{T}}).$$

When we demand that not only should statements like (2′) or (3′) be true but also that they be established by restricted means of one sort or another, we are led closer to the notion of reduction treated next.

A.5 Proof-theoretic reduction.

For simplicity, all systems S, T,... considered here are assumed to contain the system $\mathrm{I}\Sigma_1$, shown by Parsons (1970) to be conservative over the system PRA of Primitive Recursive Arithmetic. Furthermore, these systems are assumed to be primitive recursively axiomatized, and that when dealing with formalized versions of their proof predicates and provability predicates, we make use of canonically associated primitive recursive representations. Given S, we write $\mathrm{Proof}_{\mathrm{S}}(y, x)$ to express that y codes a proof in S of the formula coded by x, and $\mathrm{Prov}_{\mathrm{S}}(x)$ for $(\exists y)\ \mathrm{Proof}_{\mathrm{S}}(y, x)$. Finally we write $\mathrm{Con}_{\mathrm{S}}$ for $\neg\mathrm{Prov}_{\mathrm{S}}(0 = 1)$, to express the consistency of S. Where there is no ambiguity, we identify syntactic objects with their codes.

The essential step in arriving at the concept of proof-theoretic reduction of a system S to a system T is that this is given by an effective map f from proofs in S to proofs in T, not from formulas of L_{S} to formulas of L_{T}, as in A.3 and A.4. (Thus such f corresponds to the f^* described in 2.3 (2′) above.) In its greatest generality, as defined in (Feferman 1988), f need not be total. However, in practice, not only is it total, but it is also primitive recursive. For simplicity, that too is assumed here, though it is not assumed that f, even though total as a function, maps *every* proof in S to a proof in T. Given such f, let $\mathrm{End}(f(p))$ denote the end-formula of $f(p)$ when the latter is a proof in T. Special interest attaches to those end formulas which are preserved by f, in the following sense. Let Φ be any primitive class of formulas common to both L_{S} and L_{T}, and which includes the sentence $0 = 1$. Then we require that

$$\text{(1)} \qquad \text{for each } p, \phi, \text{ if } Proof_{\mathrm{S}}(p, \phi) \text{ and } \phi \in \Phi \text{ then } Proof_{\mathrm{T}}(f(p), \phi).$$

It is also an essential part of the notion of proof-theoretic reduction that not only do we meet (1), but we also prove it by some restricted means or other. Thus we consider provability of the formalization of (1) in a system W, where W is in general included in T. Thus define $f : \mathrm{S} \leq \mathrm{T}\ [\Phi]$ (in W) to hold if f satisfies (1) and

$$\text{(2)} \qquad \mathrm{W} \vdash \forall x, y[\mathrm{Proof_S}(y, x) \wedge \Phi(x) \rightarrow \mathrm{Proof_T}(f(y), x)],$$

where $\Phi(x)$ defines the class Φ. Then we put $\mathrm{S} \leq \mathrm{T}\ [\Phi]$ (in W) if there exists a primitive recursive f satisfying both (1) and (2). In words, S *is proof-theoretically reducible to* T *conservatively for* Φ, *provably in* W, if these hold. The two main cases of W considered below are $\mathrm{W} = \mathrm{I}\Sigma_1$ and $\mathrm{W} = \mathrm{T}$, and the associated proof-theoretic reducibility relation is then said to be *uniform* or *non-uniform*, respectively. Trivially, under the hypothesis (1) alone of $\mathrm{S} \leq \mathrm{T}\ [\Phi]$ (in W), we have that T is a conservative exension of S for formulas in Φ, i.e.,

$$\text{(3)} \qquad \phi \in \Phi \text{ and } \mathrm{S} \vdash \phi \Rightarrow \mathrm{T} \vdash \phi.$$

Applied to the equation $0 = 1$, assumed above to belong to Φ, this yields, no matter what W is taken:

$$\text{(4)} \qquad \mathrm{T} \text{ consistent} \Rightarrow \mathrm{S} \text{ consistent}.$$

But then the additional requirement (2) gives,

$$\text{(5)} \qquad \mathrm{W} \vdash \mathrm{Con_T} \rightarrow \mathrm{Con_S}.$$

We write $\mathrm{S} \leq_{\mathrm{RC}} \mathrm{T}$ (in W) for (5); in words, this expresses that S is *relatively consistent to* T, *provably in* W. Again this relation is said to be *uniform* or *non-uniform*, according as to whether $\mathrm{W} = \mathrm{I}\Sigma_1$ or $\mathrm{W} = \mathrm{T}$, resp.

In general, (provable) relative consistency is a weaker relation than proof theoretic-reducibility, and the non-uniform notions are weaker than the uniform ones. However, the following theorem shows that when Φ is taken simply to be the set $ClEq$ of closed equations $s = t$ of arithmetic, these are all the same (under our blanket assumption that all systems considered are primitive recursive and contain $\mathrm{I}\Sigma_1$).

Theorem 1. [16] *The following are equivalent:*

(i) $\mathrm{S} \leq \mathrm{T}$ [ClEq] (in T)

(ii) $\mathrm{S} \leq \mathrm{T}$ [ClEq] (in $\mathrm{I}\Sigma_1$)

(iii) $\mathrm{S} \leq \mathrm{T}$ $[\{0 = 1\}]$ (in $\mathrm{I}\Sigma_1$)

(iv) $\mathrm{S} \leq_{RC} \mathrm{T}$ (in $\mathrm{I}\Sigma_1$).

Proof. We show (i)⇒(iv)⇒(iii)⇒(ii)⇒(i). The last of these implications is of course immediate. We turn to the first one. Thus assume (i) , so that we have a primitive recursive function f satisfying

(a) $$\mathrm{T} \vdash \forall y[\mathrm{Proof}_{\mathrm{S}}(y, 0 = 1) \rightarrow \mathrm{Proof}_{\mathrm{T}}(f(y), 0 = 1)].$$

The matrix of this formula is equivalent in $\mathrm{I}\Sigma_1$ to a primitive recursive relation $R(y)$. We now make use of the following for any such R.

Lemma 1. *If* $\mathrm{T} \vdash \forall y R(y)$ *then* $\mathrm{I}\Sigma_1 + \mathrm{Con}_{\mathrm{T}} \vdash \forall y R(y)$.

The proof of this rests simply on the fact that we can prove in $\mathrm{I}\Sigma_1$ both $\mathrm{Prov}_{\mathrm{T}}(\forall y R(y))$ and $\exists y \neg R(y) \rightarrow \mathrm{Prov}_{\mathrm{T}}(\exists y \neg R(y))$.

Returning to (a), we thus have

(b) $$\mathrm{I}\Sigma_1 + \mathrm{Con}_{\mathrm{T}} \vdash \forall y\,[\mathrm{Proof}_{\mathrm{S}}(y, 0 = 1) \rightarrow \mathrm{Proof}_{\mathrm{T}}(f(y), 0 = 1)]\,.$$

But the sentence proved here implies $\mathrm{Con}_{\mathrm{T}} \rightarrow \mathrm{Con}_{\mathrm{S}}$, so we have

(c) $$\mathrm{I}\Sigma_1 \vdash \mathrm{Con}_{\mathrm{T}} \rightarrow \mathrm{Con}_{\mathrm{S}},$$

as required. Now assume (iv). Then we have

(d) $$\mathrm{I}\Sigma_1 \vdash \forall y \neg \mathrm{Proof}_{\mathrm{T}}(y, 0=1) \rightarrow \forall x \neg \mathrm{Proof}_{\mathrm{S}}(x, 0=1),$$

and hence

(e) $$\mathrm{I}\Sigma_1 \vdash \forall x \exists y \left[\mathrm{Proof}_{\mathrm{S}}(x, 0=1) \rightarrow \mathrm{Proof}_{\mathrm{T}}(y, 0=1)\right].$$

Thus by Parsons' conservation theorem for $\mathrm{I}\Sigma_1$ over PRA, there is a primitive recursive function f such that

(f) $$\mathrm{I}\Sigma_1 \vdash \forall x[\mathrm{Proof}_{\mathrm{S}}(x, 0=1) \rightarrow \mathrm{Proof}_{\mathrm{T}}(f(x), 0=1)],$$

i.e. (iii) holds. To show that (iii) implies (ii), working in $\mathrm{I}\Sigma_1$, suppose that x is a proof in S of a closed equation $s = t$. We can compute the values n, m of s, t primitive recursively, and then reduce the equation $s = t$ to $|n - m| = 0$. If $n \neq m$, this is equivalent to $0 = 1$, and we can then use f to carry over to a proof of $s = t$ in T. If $n = m$, this is just $0 = 0$, and that carries over trivially to T.

In practice, one is interested in establishing proof-theoretic reducibility with conservation for much wider classes Φ than closed equations. Usually, Φ is taken to include all Π^0_2 formulas of arithmetic; conservation of S over T for these tells us that S has no more provably recursive functions than T, and if $\mathrm{T} \subseteq \mathrm{S}$, that they have the same provably recursive functions. When the languages of S and T include analytic statements, one is interested in Φ which contain the class of Π^1_1 or even Π^1_2 formulas; conservation for the former class gives conservation with respect to provably recursive well-orderings, which are a central concern in ordinal analysis.[17] It is thus of interest, as was pointed out to me by Karl-Georg Niebergall, that quite generally for primitive recursive Φ, uniform and non-uniform proof-theoretic reducibility agree, at least for primitive recursive extensions S, T of $\mathrm{I}\Sigma_1$, as we have been assuming all along.

Theorem 2. *Suppose* Φ *is primitive recursive. Then* $\mathrm{S}{\le}\mathrm{T}\ [\Phi]$ *(in* T*) is equivalent to* $\mathrm{S} \le \mathrm{T}\ [\Phi]$ *(in* $\mathrm{I}\Sigma_1$*).*

Proof. Niebergall's proof is as follows. Suppose f is a primitive recursive function satisfying (1) above and (2) for $\mathrm{W} = \mathrm{T}$, i.e.,

(a) $$\mathrm{T} \vdash \forall x, y[\mathrm{Proof}_{\mathrm{S}}(y, x) \wedge \Phi(x) \rightarrow \mathrm{Proof}_{\mathrm{T}}(f(y), x)].$$

Denote by $R(x, y)$ the matrix of the sentence in (a); R is primitive recursive. Hence by the Lemma in the proof of Theorem 1, we have

(b) $$\mathrm{I}\Sigma_1 + \mathrm{Con}_{\mathrm{T}} \vdash \forall x, y R(x, y).$$

It follows that

(c) $$\mathrm{I}\Sigma_1 + \mathrm{Con}_{\mathrm{T}} \vdash \forall x[\mathrm{Prov}_{\mathrm{S}}(x) \wedge \Phi(x) \rightarrow \mathrm{Prov}_{\mathrm{T}}(x)].$$

But also (here's the trick),

(d) $$\mathrm{I}\Sigma_1 + (\neg\mathrm{Con}_{\mathrm{T}}) \vdash \forall x[\mathrm{Prov}_{\mathrm{S}}(x) \wedge \Phi(x) \rightarrow \mathrm{Prov}_{\mathrm{T}}(x)],$$

because if T is inconsistent, everything is provable from it. Hence, already

(e) $$\mathrm{I}\Sigma_1 \vdash \forall x[\mathrm{Prov}_{\mathrm{S}}(x) \wedge \Phi(x) \rightarrow \mathrm{Prov}_{\mathrm{T}}(x)].$$

Rewriting this as

(f) $$\mathrm{I}\Sigma_1 \vdash \forall x, y \exists z[\mathrm{Proof}_{\mathrm{S}}(y, x) \wedge \Phi(x) \rightarrow \mathrm{Proof}_{\mathrm{T}}(z, x)],$$

we can apply Parsons' (1970) conservation theorem here to obtain z as a primitive recursive function of x and y, provably in $\mathrm{I}\Sigma_1$, q.e.d.

A.6 Is proof-theoretical reducibility a "good" reducibility relation?

In his interesting contribution to this issue of *Erkenntnis*, Karl-Georg Niebergall argues that relative interpretability is the prime candidate for a general relation of reducibility between systems, partly on the grounds of its many nice properties, partly on the grounds that various alternative candidates that have been proposed are unsatisfactory in one respect or another, and partly on the grounds that those alternative candidates which *are* satisfactory are "not very different" from relative interpretability (Niebergall 2000, secs. 2 and 4.1). The "problematic" concepts of reducibility criticized op. cit. sec. 2 are not of concern to me here; except for the general notion of translation already dealt with in A.4 above, these are all model-theoretic in nature.

My view is that because of the examples in A.2(b) and (c) above, Niebergall's thesis is prima-facie wrong. First of all, by the recursive build-up of the interpretation function f, *the definition of relative interpretability makes sense only if the formulas of L_T are closed under the logical operations of L_S.* This immediately excludes the reducibility examples (b_2) and (c_3). Secondly, even where we have both S and T formalized in the same basic logic, in particular in the first-order classical predicate calculus, the reductions in examples (c_1) and (c_2) can't be accounted for in terms of relative interpretability, as Niebergall himself acknowledges. This is because if S is relatively interpretable in T and S is finitely axiomatized, then the interpretation sends S into a finitely axiomatized subtheory of T. In that case, if T is essentially reflexive, i.e. proves the consistency of every one of its finite subtheories, we have $T \vdash Con_S$. The conclusion is that if S is finitely axiomatized and T is consistent and essentially reflexive and $T \vdash Con_T \leftrightarrow Con_S$, then S is not relatively interpretable in T. These hypotheses are met in both examples (c_1) and (c_2), of BG in ZF and ACA_0 in PA, resp.

For ease of comparison in the further discussion, let me repeat Niebergall's proposed axioms from his sec. 3 (op. cit.) for a "good" reducibility relation $S \leq T$ (following his numbering of them):

(PRL1) $S \subseteq T \Rightarrow S \leq T$.

(PRL2) $S \leq T$ & $T \leq U \Rightarrow S \leq U$.

(PRL3) $S \leq T$ & T consistent $\Rightarrow$ S consistent.

(PRL4) $S \leq T$ & E finite & $E \subseteq S \Rightarrow$ there exists F finite

with $F \subseteq T$ and $E \leq F$.

(PRL5) For E, F finite, $E \leq F \Rightarrow I\Sigma_1 \vdash Con_F \rightarrow Con_E$

All of these axioms are satisfied by the relation $\leq$ of relative interpretability, but as just pointed out (and as Niebergall himself acknowledges), (PRL4) fails for proof-theoretic reduction for the very familiar cases of BG over ZF and ACA_0 over PA. This is the only one of the five proposed axioms with which I have any dispute.[18]

Let me now turn to the specific criticisms of proof-theoretic reductions in (Niebergall 2000, sec 4.1). These are mainly directed at the non-uniform reduction relation and at the associated non-uniform relative consistency relation. The suggestion to use the latter apparently originates with Kreisel (1968), p. 368. Basically, his argument was that when considering the role of W in

(1) $$\mathrm{W} \vdash \mathrm{Con_T} \to \mathrm{Con_S},$$

one can't expect to take $\mathrm{W} = \mathrm{I}\Sigma_1$, or for that matter, any other uniform base system, since Gödel's incompleteness theorems undermined Hilbert's program to take something like that system as a privileged base. Once one gives up uniformity, this does not mean that T *must* be taken for W but, as a general requirement, it is not clear what other choice could be made.[19] However, under natural additional conditions that are met in practice, one can insure that uniform relative consistency holds as well. One such condition was given in Theorem 1 of sec. A.5, namely that we have a primitive recursive function f that, provably in T, converts any proof of a closed equation in S (or more particularly, the equation $0 = 1$) to a proof of the same in T.

Niebergall's criticism of the non-uniform relative consistency relation is that it is not transitive, in contrast to the uniform relation. This is given (op. cit., Theorem 4.1, and in full detail in Niebergall 1999, Theorem 2.5) by an ingenious example of three primitive recursive extensions S, T, U of $\mathrm{I}\Sigma_1$ such that $\mathrm{S} \leq_{RC} \mathrm{T}$ (in T) and $\mathrm{T} \leq_{RC} \mathrm{U}$ (in U), but not $\mathrm{S} \leq_{RC} \mathrm{U}$ (in U).[20] The particularities of his example are not needed for the discussion here, except to remark that they fail to have other properties met in practice. One of these is conservation with respect to Π^0_2 sentences. Since the statement of relative consistency is one such, it is easily seen that we have:

(2) $$\mathrm{S} \leq_{RC} \mathrm{T} \text{ (in T)},\ \mathrm{T} \leq_{RC} \mathrm{U} \text{ (in U) and U } \Pi^0_2\text{–conservative over T} \Rightarrow$$

$$\mathrm{S} \leq_{RC} \mathrm{U} \text{ (in U)}.$$

As it happens, in Niebergall's example (loc. cit.), U is not even Π^0_1 conservative over T.

The second of Niebergall's criticisms is directed at non-uniform proof-theoretic reducibility, again because of an example of failure of transitivity (op. cit., Theorem 4.3, and in full detail in Niebergall 1999, Theorem 2.20). It follows from Theorem 2 in the preceding section that the counterexample *can't* be for extensions of $I\Sigma_1$, since non-uniform proof-theoretic reducibility *is* transitive for such, by the equivalence with the uniform relation. Indeed, this second counterexample (again ingenious) is for systems which are extensions of the weaker system $I\Delta_0$ + Exp.[21] As a result of this second counter-example, there may thus be a concern as to the use of the notion of non-uniform proof-theoretic reducibility when applied to systems which are in this weaker range. But that does not argue against *uniform* proof-theoretic reducibility modified to $I\Delta_0$ + Exp or perhaps even some weaker system as a base. Since my concerns in this paper are with the challenges to reductive proof theory having to do with systems that are vastly stronger than $I\Sigma_1$, the problems with those lie elsewhere. However, such careful and thought-provoking investigations of reducibility relations as Niebergall has carried out are greatly to be valued for their contribution to the development of proper conceptual foundations of proof theory in general.

Acknowledgments. I would like to thank Wilfried Buchholz, Gerhard Jäger, Michael Rathjen, Anton Setzer, and Thomas Strahm for providing technical information and helpful comments, especially on section 6; I wish also to thank Karl-Georg Niebergall for the same with respect to the Appendix section A.5.

Dept. of Mathematics
Stanford University
Stanford CA 94305 USA
sf@csli.stanford.edu

---

[1] I must admit I have not studied the literature pro and con scientific reductionism to any significant extent. I am sympathetic, for example, to the anti-reductionist arguments presented by Dupré(1993).

[2] By functional foundations, I mean the attempts by Church and Curry to develop mathematically all-encompassing theories whose variables range over arbitrary total functions.

[3] For a broader picture of proof-theoretical concerns and approaches (not only those which are reductive in character), see the appendixes to (Takeuti 1987) and the handbook (Buss 1998).

[4] See (Buchholz 1991) for details.

[5] In order to concentrate on the pattern, we are being deliberately vague about details here and below.

[6] For example, in Problem 2 of his famous list of mathematical problems at the meeting of the International Congress of Mathematicians, Paris, 1900; cf, e.g. (Feferman 1998), p. 13 for discussion.

[7] I have discussed Kreisel's program vs. Hilbert's program in (Feferman 1996), pp. 267–269.

[8] Incidentally, GC is in Π^0_1 form, and it is known that RH can be brought to that form, while TP is in Π^0_2 form.

[9] My presumption—partially in view of (Buchholz 1997, 1999)—is that whatever applies to Rathjen's work, insofar as the discussions in this paper are concerned, applies as well to Arai's work.

[10] The criteria for constructivity of intuitionistic systems of explicit mathematics are informal and less demanding than those for constructive type theory in the sense of Martin-Löf.

[11] Examples of mathematical notions that cannot be expressed in full in set theory: *truth of a statement of the language of set theory*, and *the category of all categories*. There are other common notions which are only modeled (not explicated) in set theory, such as that of *infinitesimal*, or *random variable*.

[12] I am such a one.

[13] Non-set-theoretic.

[14] A precursor to that result was an argument due to Kreisel, presented in (Feferman 1960), pp. 85–86, which showed that there is such f satisfying (1)–(3); though stated there for translations into PA, the argument works equally well for translations into Q.

[15] This was already suggested in (Kreisel 1955), p. 31.

[16] This theorem has been observed independently by Niebergall (1999), 2.14–2.15. An essential step in the argument, stated as Lemma 1 in the proof here, is due to a remark of Kreisel, unpublished by him as far as I know; it is to be found in (Feferman 1988), p. 369.

[17] See (Feferman 1988, or 1993) for a wide range of examples of such proof-theoretic reductions.

[18] A fine point to raise about (PRL1) is that if we don't have a proof that $S \subseteq T$ then we can't satisfy $S \leq T$ in the case of proof-theoretic reducibility.

[19] Kreisel's argument (loc. cit) for doing so is worth quoting; in it he uses the symbol '$\wp_0$' to represent finitist reasoning in something like Hilbert's original sense, for which PRA or $I\Sigma_1$ is now often taken as a surrogate. He says that the requirement that relative consistency proofs ($\text{ConS} \rightarrow \text{ConS}'$) be given in $\wp_0$ is "one of the hangovers from Hilbert's programme... This depends on the tacit conviction that, some day, ConS would be proved in $\wp_0$. But, a chain being as weak as its weakest link, a philosophically more meaningful requirement is to use any of the principles that one expects to need for proving ConS. *A posteriori* some 'justification' can be given for the old requirement by showing that it is automatically fulfilled under more reasonable conditions."

[20] My notation for the relations involved is somewhat different from Niebergall's , but the reader should not have any trouble matching them up.

[21] It should be noted that the necessary proof-theoretical notions can be formalized in $I\triangle_0 + \text{Exp}$, as is necessary to make sense of the relation of non-uniform proof-theoretic reducibility between its extensions.

References

P. Aczel, 1978. The type theoretic interpretation of constructive set theory. In *Logic Colloquium '77*, (A. MacIntyre et al, eds.), North-Holland, Amsterdam, 55–66.

T. Arai, 1996. Systems of ordinal diagrams. (Preprint)

T. Arai, 1997. Proof theory for theories of ordinals II: Σ_1-stability. (Preprint)

P. Benacerraf and H. Putnam (eds.), 1983. *Philosophy of Mathematics*, 2nd edn. Cambridge Univ. Press, Cambridge, U.K.

W. Buchholz, 1991. Notation system for infinitary derivations. *Archive for Mathematical Logic 30*, 277–296.

W. Buchholz, 1997. Explaining Gentzen's consistency proof within infinitary proof theory, in *KGC'97*, Lecture Notes in Computer Science 1289, 4–17.

W. Buchholz, 1999. Explaining the Gentzen-Takeuti reduction steps: a second order system. (Preprint; to appear in *Archive for Mathematical Logic*)

W. Buchholz, S. Feferman, W. Pohlers, W. Sieg, 1981. *Iterated Inductive Definitions and Subsystems of Analysis*, Lecture Notes in Mathematics 897.

S.R. Buss (ed.), 1998. *Handbook of Proof Theory*, North-Holland, Amsterdam.

J. Dupré, 1993. *The Disorder of Things*. Harvard Univ. Press, Cambridge.

S. Feferman, 1960. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae 49*, 35–92.

S. Feferman, 1975. A language and axioms for explicit mathematics. In J.N. Crossley (ed.), *Algebra and Logic,* Lecture Notes in Mathematics 450, 87–139.

S. Feferman, 1979. Constructive theories of functions and classes. In M. Boffa et al (eds.), *Logic Colloquium '78*, North-Holland, Amsterdam, 159–224.

S. Feferman, 1984. Foundational ways. In Perspectives in Mathematics, (W. Jäger et al,eds.), Birkhäuser, Basel, 147–185. (Reprinted in (Feferman 1998) 94–104.)

S. Feferman, 1988. Hilbert's program relativized. Proof-theoretical and foundational reductions. *J. Symbolic Logic 53*, 364–384.

S. Feferman, 1993. What rests on what? The proof-theoretical analysis of mathematics. In *Philosophy of Mathematics, Vol. I.*, (J. Czermak, ed.), Verlag-Hölder-Pichler-Tempsky, Vienna, 147–171. (Reprinted in (Feferman 1998), 187–208.)

S. Feferman, 1996. Kreisel's "unwinding" program. In *Kreiseliana*, (P. Odifreddi, ed.) A.K. Peters Ltd., Wellesley, 247–273.

S. Feferman, 1998. *In the Light of Logic*. Oxford Univ. Press, New York.

S. Feferman, 1999. Does mathematics need new axioms? *American Mathematical Monthly 106*, 99–111.

G. Gentzen, 1969. *The Collected Papers of Gerhard Gentzen* (M.E. Szabo, ed.). North-Holland, Amsterdam.

G. Jäger and W. Pohlers, 1982. Eine beweistheoretische Untersuchung von ($\triangle^1_2$-CA) + BI und verwandte Systeme. *Sitzungsberichte der Bayerischen Akademie der Wissenschaften*, 1–28.

G. Jäger and T. Studer, 1999. Extending the system T_0 of explicit mathematics: the limit and Mahlo axioms. (Preprint)

G. Kreisel, 1955. Models, translations and interpretations. In *Mathematical Interpretation of Formal Systems* (Th. Skolem et al, eds.), North-Holland, Amsterdam, 26–50.

G. Kreisel, 1958. Hilbert's programme. *Dialectica 12*, 346–372. (Revised, with Postscript in (Benaceraff and Putnam 1983), 207–238.)

G. Kreisel, 1968. A survey of proof theory. *J. Symbolic Logic 33*, 321–388.

G. Kreisel, 1987. Proof theory: some personal recollections. Appendix to (Takeuti 1987), 395–405.

P. Lindström, 1997. Interpretability in reflexive theories—a survey. *Theoria 63, Part 3*, 182–209.

P. Martin-Löf, 1984. *Intuitionistic Type Theory.* Bibliopolis, Naples.

K.-G. Niebergall, 1999. Proof-theoretical reducibility. (Preprint.)

K.-G. Niebergall, 2000. On the logic of reducibility: axioms and examples. (This journal.)

P. Oppenheim and H. Putnam, 1958. The unity of science as a working hypothesis. In *Minnesota Studies in the Philosophy of Science, Vol. 2* (H. Feigl et al, eds.). Univ. of Minnesota Press, Minneapolis.

C. Parsons, 1970. On a number-theoretic choice scheme and its relation to induction. In *Intuitionism and Proof Theory* (A. Kino et al, eds.). North-Holland, Amsterdam, 459–473.

W. Pohlers, 1998. Subsystems of set theory and second order number theory. In (Buss 1998), 209–335.

M.B. Pour-El and S. Kripke, 1967. Deduction-preserving "Recursive Isomorphisms" between theories. *Fundamenta Mathematicae 61*, 141–163.

M. Rathjen, 1991. Proof-theoretic analysis of KPM. *Archive for Mathematical Logic 30*, 377–403.

M. Rathjen, 1994. Proof theory of reflection. *Annals of Pure and Applied Logic 68*, 181–224.

M. Rathjen, 1995. Recent advances in ordinal analysis: Π^1_2-CA and related systems. *Bull. Symbolic Logic 1*, 468–485.

M. Rathjen, 1998. The superjump in Martin-Löf type theory, in *Proceedings of the Logic Colloquium '98, Lecture Notes in Logic, 13.*

K. Schütte, 1960. *Beweistheorie.* Springer-Verlag, Berlin.

K. Schütte, 1977. *Proof Theory.* Springer-Verlag, Berlin.

A. Setzer, 1998. Extending Martin-Löf type theory by one Mahlo universe. (Preprint; to appear in *Archive for Mathematical Logic.*)

S. Simpson, 1988. Partial realizations of Hilbert's program. *J. Symbolic Logic 53*, 349–363.

S. Simpson, 1998. *Subsystems of Second Order Arithmetic.* Springer-Verlag, Berlin.

W.W. Tait, 1981. Finitism. *J. of Philosophy 78*, 524–546.

G. Takeuti, 1975. *Proof Theory.* North-Holland, Amsterdam.

G. Takeuti, 1987. *Proof Theory* (2nd. edn.). North-Holland, Amsterdam.

A. Tarski, A. Mostowski, and R.M. Robinson, 1953. *Undecidable Theories.* North- Holland, Amserdam.

S. Tupailo, 2000. Realization of constructive set theory into explicit mathematics: a lower bound for impredicative Mahlo universe. (Preprint.)

H. Wang, 1951. Arithmetic translations of axiom systems. *Transactions American Math. Soc. 71*, 283–293.

F. Ye, 1999. *Strict Constructivism and the Philosophy of Mathematics.* Ph. D. Dissertation, Dept. of Philosophy, Princeton University.