JAMES HAWTHORNE

MATHEMATICAL INSTRUMENTALISM MEETS THE CONJUNCTION OBJECTION

ABSTRACT. Scientific realists often appeal to some version of the *conjunction objection* to argue that scientific instrumentalism fails to do justice to the full empirical import of scientific theories. Whereas the conjunction objection provides a powerful critique of *scientific instrumentalism*, I will show that *mathematical instrumentalism* escapes the conjunction objection unscathed.

Our most sophisticated scientific theories are highly mathematical. Read literally, they assert the existence not only of physical things and properties like electrons, charges, and fields, but also of such non-physical things as numbers, vectors, mathematical functions, and sets. So *scientific realists*, if they are physicalists, should be *instrumentalists* with regard to at least some parts of our best scientific theories. For, if only the physical is real, then either mathematically endowed theories *must be false*, or (if they *may be true*) the sentences they imply that contain mathematical terms should not be taken literally. Physicalists should hold that the mathematical terms in a scientific theory are merely instrumental, an aid in the logical systematization of the theory's *real* subject matter.[1]

Physicalist's might adopt either of two distinct varieties of mathematical instrumentalism. *Formalists* take the mathematical expressions in theories to be literally meaningless but useful *syntactic tools*. *Fictionalists*, on the other hand, maintain that statements containing mathematical expressions make meaningful but deliberately *false* claims. The issues regarding mathematical instrumentalism that I will address in this paper apply equally to either version. Both varieties of instrumentalism share the view that mathematical assertions are not literally true, but mathematical expressions are an important aid in the logical systematization of the purportedly true, non-mathematical, substantive claims that theories make.

If mathematics plays only an instrumental role in scientific theories, might not other terms in these same theories also assume a merely instrumental part? *Scientific instrumentalists* believe so. Some believe that there is an identifiable theory/observation boundary, and attempt to draw the line between *empirically meaningful* and *purely instrumental* sentences along a divide between "observation sentences" (which contain only terms that purportedly refer to observable things) and "theoretical sentences" (which contain some terms that purportedly refer to unobservable things). Other versions of scientific instrumentalism attempt to draw a different line between *empirically meaningful* and *purely instrumental* sentences on the bases of other distinctions. But however scientific instrumentalists choose to draw this line, *scientific realists* chide them for not taking theories at face value. But are scientific realists who regard mathematics as a mere instrument in any position to reproach scientific instrumentalists for not taking theories literally? Only *mathematical realists*, it seems, may profess a strictly literal reading of theories.

Despite these considerations, I will argue that *mathematical instrumentalism* is defensible in a logically principled way that is not available to the *scientific instrumentalist*. To be precise, scientific realists often appeal to some version of the *conjunction objection* to argue that instrumentalism fails to do justice to the full empirical import of scientific theories (see, e.g., Boyd, 1973; Putnam, 1973). I will argue that whereas the conjunction objection provides a powerful critique against scientific instrumentalism, mathematical instrumentalism can escape the objection unscathed.

Phrased in terms of the theory/observation distinction, the conjunction objection contends that the "theoretical" part of a scientific theory carries hidden observational content not expressed by its observational subtheory. This hidden *excess content* shows itself when theories are conjoined. For, the observational logical consequences of the *conjunction of two theories* T_1 and T_2 may far exceed the observational logical consequences of the *conjunction of their observational subtheories*. Indeed, it can be proved that given any theory T_1, there exists an alternative theory T_3 that has precisely the same observational subtheory as T_1, but such that when T_3 is conjoined with some auxiliary theory T_2 it yields very different *excess* observational consequences than does the conjunction of T_1 with T_2.[2] Usually T_1 will have an infinite number of such "observationally equivalent" alternatives that each implies very different additional observational consequence when conjoined with some common auxiliary theory T_2.

The realist then argues that since the instrumentalist believes that theories T_1 and T_2 are merely instrumentally correct, she is not at all entitled to believe that their conjunction is also instrumentally correct. Until the *excess consequences* of the conjoint theory (T_1 & T_2) are empirically tested against the *excess consequences* of (T_3 & T_2), the most that an instrumentalist is entitled to believe are the consequences of the conjunction of the non-instrumental part of T_1 (which is identical to the non-instrumental subtheory of T_3) with the non-instrumental part of T_2. Thus, the realist argues, to the extent that the instrumentalist is disposed to believe the excess observational consequences of the conjoint theory (T_1 & T_2), as scientists usually do, she treats the "theoretical" parts of these theories realistically rather than instrumentally – as though each is *true* rather than merely empirically adequate.

Notice that the way in which the theory/observation distinction is specified makes no difference to the force of the conjunction objection. Notice also that precisely the same argument applies to any version of scientific instrumentalism however the line between the merely instrumental and the empirically meaningful is drawn. Thus, it appears that to whatever degree the conjunction objection favors a realist stance towards theoretical claims, it must equally favor a realist stance towards mathematical claims. For, when theories are conjoined, the mathematical parts of each may, it seems, contribute to the entailment of excess non-mathematical consequences in just the way that theoretical statements contribute to the entailment of excess observational consequences. So aren't scientific realists who wield the conjunction objection against scientific instrumentalism *obliged by a precisely analogous argument* to take mathematical statements realistically as well? Aren't they obliged to believe in the existence of the mathematical things that are said to exist by the scientific theories they believe? I will argue that they need not be. Mathematical instrumentalism can beat the conjunction objection in a logically principled way that is not available to a broader *scientific instrumentalism.*

Specifically, I will show that when *Representation theorems* of the kind proved by Field (1980) for Newtonian gravitation theory hold for each of a pair of mathematically endowed theories, then the conjunction of the mathematical versions of the theories[3] implies *only* those non-mathematical consequences that are already implied by the conjunction of their non-mathematical parts. That is, the conjunction of the mathematical theories is a *conservative extension*[4] of the conjunction of their non-mathematical subtheories. So, when Representation theorems hold, alternative mathematical theories T_1 and T_3 that agree on non-mathematical consequences may be viewed simply as different mathematical formulations of the same physical theory. When each is conjoined to an auxiliary theory T_2, the resulting theories (T_1 & T_2), and (T_3 & T_2) make precisely the same non-mathematical claims. By contrast, if the dividing line between instrumental sentences and realistic assertions is drawn "lower down," so that some non-mathematical terms are taken to be merely instrumental, then the conjunction of two instrumental theories will generally be a *non*-conservative extension of the conjunction of there non-instrumental parts. Thus, mathematical instrumentalists can overcome the conjunction objection in a logically principled way that is not generally available to theoretical instrumentalists.[5]

My aim in this paper is to state all of these claims precisely and to show that they are warranted. The treatment will be greatly facilitated by drawing on the resources of many-sorted languages containing *many-sorted quantifiers*. Their logic is a very modest extension of familiar one-sorted logic. In the next three sections I will discuss the utility of a many-sorted logic for a proper explication of instrumentalism, and I will introduce the relevant formal details of a many-sorted language and its semantics. Then, in Section 4 through 7 I will provide a precise treatment of the conjunction problem for instrumental theories and show how mathematical instrumentalism can overcome it.

1. SORTING OUT THE LANGUAGE

In "To Save the Phenomena" (1975, pp. 41–69) van Fraassen argues that instrumentalism as it is usually formulated does not successfully eliminate the apparent reference by theories to unobservable entities. Although his objection is technically correct, I think that the difficulty he raises is easily fixed. This fix is important to a proper treatment of instrumentalism, so I will discuss it in some detail.

The standard formulation of instrumentalism attempts to circumvent claims that a theory appears to make about dubious theoretical things by dividing the vocabulary of the theory into two classes of terms, terms that refer, and terms that merely function as an instrumental aid to the logic. These classes are usually taken to consist only of names (individual constants), function symbols, and predicate (including relation) symbols. Logical terms (including the identity symbol), *all variables* and quantifiers are counted among the non-instrumental vocabulary. Any sentence that contains an instrumental (e.g. theoretical) term is itself instrumental, and not to be taken literally. (Such sentences either make fictitious claims or are taken to be meaningless.) Only sentences with no occurrences of instrumental vocabulary purport to be true. For the instrumentalist, then, a theory T is only about its *non-instrumental subtheory* – the set of non-instrumental sentences that T logically implies. The rest of the theory is simply a tool to aid in the deduction of non-instrumental consequences.

Indeed, instrumentalists often appeal to the fact that if T is an axiomatizable first order theory, then the "tool" may in principle be dispensed with entirely and replaced with a Craigian re-axiomatization of the non-instrumental subtheory.

Van Fraassen points out that this version of instrumentalism will not work for most real scientific theories. He denotes the observational sub-vocabulary of a theory T by 'E', and denotes the observational sub-theory of T by 'T/E' (T restricted to vocabulary E). He then argues that the empirical import of a theory cannot be isolated in the syntactic fashion employed by instrumentalists (1975, 54–55):

> But any unobservable entity will differ from the observable ones in the way it systematically lacks observable characteristics. As long as we do not abjure negation, therefore, we shall be able to state in the observable vocabulary (however conceived) that there are unobservable entities, and, to some extent, what they are like. The quantum theory, Copenhagen version, implies that there are things which sometimes have a position in space, and sometimes have not. This consequence I have just stated without using a single theoretical term. Newton's theory implies that there is something (to wit, Absolute Space) which neither has position nor occupies a volume. Such consequences are by no stretch of the imagination about what there is in the observable world, nor about what any observable thing is like. The reduced theory T/E is not a description of part of the world described by T; rather, T/E is, in a hobbled and hamstrung fashion, the description by T of everything.

I think that van Fraassen is right. The *usual* sorting of vocabulary into observational and theoretical parts doesn't do the job required by instrumentalists. But we can easily fix the problem by adapting a device employed by Field (1980) for a somewhat different purpose.

In *Science Without Numbers* Field investigates the effect of adding a mathematical theory S to a non-mathematical theory T. He shows that $S + T$ is *Conservative* in the sense that it will only have non-mathematical consequences that are already implied by T alone. But in stating this result one must take care not to run afoul of a problem that is roughly the inverse of the difficulty raised by van Fraassen. The non-mathematical theory T might imply that everything has some physical properties (e.g. spatial location) not shared by the mathematical things posited by S; T may even rule out the existence of non-physical things altogether. So, if one isn't careful, $S + T$ might imply that the numbers have spatial locations, or $S + T$ may even be logically inconsistent.

Field avoids this problem by introducing a one-place predicate 'M', where 'Mx' stands for 'x is a mathematical entity'. Then, for a given non-mathematical theory T, he lets T^* be the sentences of T with all quantifiers relativized to the negation of 'M', so that $(x)P$ becomes $(x)(\sim Mx \supset P)$ and $(\exists x)P$ becomes $(\exists x)(\sim Mx \,\&\, P)$. Similarly, quantified expressions about mathematical things take the form $(x)(Mx \supset \cdots)$ and $(\exists x)(Mx \& \cdots)$. So, if T says that everything is located in space-time, then T^* says that all *non-mathematical* things are located in space-time, but is mute about the locations of mathematical things.

The same approach may be employed to mark off purported theoretical (or other undesirable) entities that differ systematically from observable (or otherwise "desirable") things in regard to the sorts of properties they have or lack. To see how the relativization of the quantifiers solves the problem raised by van Fraassen, let 'U' be a new predicate such that 'Ux' expresses 'x is unobservable'. Let all quantifiers in T and its logical consequences be appropriately relativized, and call the resulting theory T^+. T^+ is just a version of T that makes clear which quantifiers are intended to range over observable things and which are meant to range over unobservables. If van Fraassen is right about there being systematic differences in the sorts of properties attributed by T to unobservables, then the alteration in T required to generate T^+ will be easily accomplished; if there are no such systematic differences, then the problem described by van Fraassen won't arise, so no relativization of quantifiers will be needed. (If T has a consequence containing an expression of form $(x)Px$ and the instrumentalist really intends the expression to say that all observable *and* unobservable things satisfy Px, then the corresponding relativized sentence in T^+ will be the conjunction $((x)(Ux \supset Px)\&(x)(\sim Ux \supset Px))$.)

Now the theory with just the observational consequences of T^+, which the instrumentalist wants to endorse, may be obtained as follows. First, take the subset of logical consequences of T^+ in which all quantifiers are relativized to '$\sim U$' and in which all *other* non-logical vocabulary (i.e. all vocabulary other than the occurrences of 'U' in relativized quantifiers) consists of observational vocabulary. Next, strip away all occurrences of '$\sim Ux \supset$' and '$\sim Ux\&$' (and excess parentheses) from these sentences. The resulting theory $T^\sim$ is just the theory that the instrumentalist had in mind. With regard to van Fraassen's quantum mechanics example, T^+ will imply '$(\exists x)(Ux\&x$ sometimes has position $\&x$ sometimes has no position)', so $T^\sim$ will *not* imply '$(\exists x)(x$ sometimes has position $\&x$ sometimes has no position)'.

2. A MANY-SORTED LANGUAGE AND ITS LOGIC

To overcome van Fraassen's objection the instrumentalist must mark a distinction between occurrences of quantifiers that play a merely instrumental role (or that range over "mere fictions") and those that purport to range over real things. Sorted quantifiers provide an equivalent, but more convenient way to mark this distinction than relativizing quantifiers to a special predicate 'U'. That is, one can replace all occurrences of variables and quantifiers that play an instrumental role in the theory T with variables of one sort (say, subscripted occurrences of 'u'); and replace all other occurrences of variables, the purportedly referential occurrences, with a second sort of variables (e.g. subscripted occurrences of 'v'). The result is an instrumentalist version of T stated in a 2-sorted language containing 2 sorts of quantifiers. The instrumentalist already supposes that the constant terms of the language (names, function symbols, and predicate symbols) can be divided into two sorts, so she may as well divide the variables (and quantifiers in which they occur) into the same two sorts. The theory $T^{\#}$ that results from sorting variables in T is essentially equivalent to the theory T^{+} that results from the relativization of quantifiers in T to the special predicate 'U'. (This is a well known fact about the relationship between many-sorted languages and the relativization of quantifiers in single sorted language (e.g., see Wang, 1964).) Moreover, the subset of consequences of $T^{\#}$ containing only non-instrumental vocabulary is just the theory $T^{\sim}$ described earlier. Many-sorted languages with sorted quantifiers possess essentially the same expressive power as one-sorted languages, but their use will greatly simplify my treatment of instrumentalism. So, in this section I will briefly describe the main features of some many-sorted languages and their logics.

The n-sorted language L (for some positive integer n) consists of the usual logical symbols (including the identity relation), and for each sort k (an integer such that $0 \leqslant k \leqslant n-1$), an infinite number of individual variables, individual constants, m-ary function symbols, and m-ary predicate symbols of sort k. If L is a second order language, it will also contain m-ary predicate variables of each sort k. L contains universal and existential quantifiers of each sort; (x) and $(\exists x)$ are universal and existential quantifier of the same sort (and order) as the variables x. Well-formed formulas and sentences are built up from the symbols of L in the usual way, with no special restrictions on sort (e.g. each m-ary predicate symbol followed by m names and constants, no matter what their sorts, is a formula).

For the purposes of this paper 2-sorted and 3-sorted languages will suffice. Usually I will regard theories as couched in a 2-sorted language. But when we consider the effects of combining the instrumental parts of a *pair* of theories T_1 and T_2 it will sometimes be useful to employ a 3-sorted language common to the two theories. In this context I will take sort-0 non-logical constants, sort-0 variables, and sentences containing only sort-0 vocabulary to be the non-instrumental vocabulary, and I will suppose that sort-0 vocabulary is common to T_1 and T_2. Sort-1 will be the instrumental vocabulary for theory T_1, and sort-2 will be the instrumental vocabulary of T_2.

When we focus on mathematical instrumentalism (and treat all non-mathematical, theoretical terms realistically) a 2-sorted language will suffice. Sort-0 will consist of the non-mathematical vocabulary, including variables that range over the domain of non-mathematical things. Sort-1 vocabulary will consist of mathematical terms (particularly the set-theoretic vocabulary '$\varnothing$' and '$\in$' and variables ranging over sets). Mathematical vocabulary will also include predicate and function symbols that relate mathematical things to non-mathematical things (e.g. the "mass of" function, which assigns numbers to bodies).

It is customary to specify the notion of *logical consequence* for a language in set-theoretic terms, in terms of *interpretations* or *models*. Ultimately this approach should be unsatisfactory to the mathematical instrumentalist, for it characterizes the notion of *logical consequence* in terms of a mathematical theory that the instrumentalist takes to consist either of meaningless symbols or terms refering to fictions. Furthermore, if the set-theoretic semantics is to apply to the instrumental part of the formal language in the same way that it applies to the rest of the language, then the intended *domain of discourse* for an interpretation must contain fictitious objects to which instrumental terms in the language "refer." If the instrumentalist is to avoid these difficulties, he will have to specify the *logical consequence relation* in an alternative fashion. One way might be to define *logical consequence* in terms of syntactic deduction rules; but this can only work for first order logic (since for logics of higher orders semantic consequence outruns any notion of consequence that may be captured by sound deduction rules). Field (1989, 1992) avoids the difficulties in another way. He develops a modal logic for the *logical necessity* and *logical possibility* operators and shows how the relevant metalogical results from standard semantics can be represented and proved in the modal logic. Thus, the mathematical instrumentalists may appeal to metalogical results without invoking mathematics.

In the remainder of this section I will briefly specify the notion of *logical consequence* for many-sorted languages in the usual set-theoretic way. This should provide a fairly familiar setting to aid the reader in understanding the many-sorted logic and its application to instrumentalism. In a later section I will show how to restate and prove the principal claims I will make in support of mathematical instrumentalism using only the resources of Field's modal metalogic.[6]

An *interpretation* of any n-sorted language employs n non-empty, disjoint domains (disjoint because no object is both observable and unobservable; no object is both mathematical and non-mathematical). Given any such n-tuple of domains $\langle D_0, D_1, \ldots, D_{n-1}\rangle$, let $\mathbf{D} = \bigcup_{i=0}^{n-1} D_i$. $\mathbf{D}$ is the super-domain that contains all members of each sort. An interpretation then consists of an assignment of an object of sort k to each individual constant of sort k, a function from $\mathbf{D}^m$ to D_k to each m-ary function symbol of sort k, and a subset of $\mathbf{D}^m$ to each m-ary predicate symbol (of any sort). Thus, observable things are allowed to satisfy expressions for "unobservable properties", and "unobservable things" are permitted to "possess properties" that count as observable (e.g. spatial location in van Fraassen's example). Of course a given theory may explicitly deny that the things of sort k satisfy a given expression Q simply by invoking the sentence $(x) \sim Qx$ with a variable x of sort k.

Individual variables are assigned objects by functions that assign to each variable of sort k an object from D_k. Satisfaction and truth under an interpretation behave as usual. In particular, when x is an individual variable of sort k and Px is a formula with x free, a value assignment α to variables satisfies $(x)Px$ (satisfies $(\exists x)Px$) just in case for every assignment (some assignment) β that differs at most from α in the member of D_k it assigns to x, β satisfies Px.

When considering mathematical instrumentalism I will resort to a 2-sorted language in which sort 0 contains all of the non-mathematical vocabulary and sort 1 contains the mathematical language (which I will suppose can all be defined in set-theoretic terms). The semantics will be the same as that just described for any n-sorted language, but with two emendations. First, in the case of the scientific instrumentalist's distinction between, for example, observational and non-observational terms we have permitted predicates in the observational vocabulary to apply to unobservables. However, mathematical instrumentalists need not count mathematical things within the extensions of non-mathematical predicates; so we will treat sort-0 predicate and function symbols applied to sort-1 terms as ill-formed. Secondly, Field (1980, 1989) advocates mathematical instrumentalism for both first order and second order languages, and I will follow suit. Hence, in the context of the discussion of mathematical instrumentalism the only emendations required in the previously specified semantics are these: the language will be 2-sorted; interpretations only assign subsets of D_0^m (rather than of $\mathbf{D}^m$) to m-ary sort-0 predicates, and only assign functions from D_0^m to D_0 to m-ary sort-0 function symbols; each assignment α to variables assigns a subset of D_0^m to m-ary sort-0 predicate variables and assigns a subset of $\mathbf{D}^m = (D_0 \cup D_1)^m$ to m-ary sort-1 predicate variables. Then, as is usual for second order quantifiers, when X^m is a predicate variable of sort k and $P(X^m)$ is a formula with X^m free, an assignment α to variables satisfies $(X^m)P(X^m)$ (satisfies $(\exists X^m)P(X^m)$) just in case for every assignment (some assignment) β that differs at most from α in the subset it assigns to X^m (from D_0^m for X^m of sort 0, or from $\mathbf{D}^m$ for X^m of sort 1), β satisfies $P(X^m)$.

In the explication of mathematical instrumentalism an *intended* interpretation will have a sort-0 domain D_0 that contains only non-sets. D_0 is the set of physical things that the theory is about. The *intended* sort-1 domain D_1 contains (fictitious) platonistic elements that are the sets, sets of sets, etc., built up from D_0. A natural way to build the intended sort-1 domain D_1 is this: let $D(0)$ be D_0; if ξ is any ordinal number, let $D(\xi + 1)$ be the union of $D(\xi)$ with the set of all subsets of $D(\xi)$; if λ is a limit ordinal, let $D(\lambda)$ be the union of all sets $D(\xi)$ such that $\xi < \lambda$; this recursive definition leads to a set $D(\gamma)$ where γ is a strongly inaccessible ordinal greater than ω; finally let the sort-2 domain D_1 be $D(\gamma) - D_0$.

When we treat mathematical instrumentalism with a 2-sorted language, the sort-1 mathematical relation symbol '$\in$' is to be interpreted on the whole super-domain $\mathbf{D} = D_0 \cup D_1$, as are all other predicates and function symbols that connect non-mathematical things with numbers and sets (e.g. the "mass of" function). This is *not* a special emendation to our general treatment of predicate and function symbols for n-sorted languages. The only special emendation of the earlier treatment that mathematical instrumentalism requires is the *restriction* of the sort-0 predicate and function terms to the sort-0 domain.

The notions of *logical truth* and *logical consequence* for an n-sorted languages L may then be defined in terms of interpretations of L in the usual way. If L is first order, *logical consequence* is recursively axiomatizable by the usual axioms and inference rules except that: (1) universal specification requires the insertion of a name of the same sort as the variable for which it is substituted; (2) universal generalization from Pc to $(x)Px$ will be legitimate under the usual conditions provided that x is a variable of the same sort as the constant c; (3) if x and y are variables of different sorts, then $(x)(y) \sim y = x$ is a logical truth. Indeed, the first order logic of an n-sorted language is equivalent to the logic of a single sorted language in which each quantifier is relativized to one of n distinct predicates 'S_i', and in which the following sentences are treated as (logical) axioms: $(x)(S_0x \vee \cdots \vee S_{n-1}x); (x) \sim (S_ix \,\&\, S_jx)$ (for each distinct i and j); $(\exists x)S_ix$ (for each i); and an axiom of from S_ic for each name c (which specifies the sort of c).

Use of a many-sorted language simplifies the description of the solution offered in the previous section to the problem raised by van Fraassen. For a given theory T the scientific instrumentalist will not only want to divide the constant terms of T's language into observational and theoretical terms. She should also specify two sorts of variable: one sort to range over observable things, the other sort to be part of the instrumental (theoretical) vocabulary. Next the instrumentalist must decide which occurrences of quantifiers in T are supposed to range over observable things and which purport to range over unobservables, and should replace each occurrence with the right sort of variable. (If some sentence or subformula of a sentence is really supposed to assert that *everything* of either sort satisfies formula Px, then occurrences of $(x)Px$ should be replaced with the conjunction $((y)Py \,\&\, (z)Pz))$, where y is of one sort and z is of the other.) Call this sorted version of T the theory $T^{\#}$. Now it is easy for the instrumentalist to specify which of the consequences of $T^{\#}$ she believes. She believes the observational consequences of $T^{\#}$ – the consequences that involve only observational vocabulary, including only observational variables. Supplied with appropriate sorts of quantifiers, van Fraassen's quantum mechanics example merely says that *there are unobservable* things that sometimes have a position in space and sometimes have no position. The occurrence of quantifiers of the sort 'there are unobservable things such that ...' within a sentence of a theory clearly marks the sentence as instrumental.

3. THEORIES IN TWO-SORTED LANGUAGES

Even the most committed scientific realist may grant that a theory he believes to be completely true can be *enhanced* by the addition of convenient fictions. The additional syntax may permit shortcuts in the computation of truths from truths, and a fictitious ontology tied to this syntax may offer conceptual shortcuts that aid in the modeling of complex physical systems. So, let us put aside for now disputes over *which* pieces of syntax are mere tools. Let us suppose that a community of philosophically like-minded scientists can distinguish merely instrumental terms in a given theory from those terms they intend to be referential, and that they can distinguish quantifiers they intend to range over real things from the instrumental sort. Let L_0 represent the vocabulary that the community understands realistically. A scientific theory T may be stated wholly in the realist vocabulary L_0, or the language of T may contain some additional instrumental vocabulary, vocabulary of some sort other than sort-0.

I will call any collection of sentences (in a given n-sorted language) a *theory*; and for any theory T, its language L_T will, by convention, consist of all vocabulary that actually occurs in T together with *all* vocabulary in L_0, the sort-0 part of the language. A theory T *may* be axiomatizable, but I will not generally assume so. Neither will I assume that T is logically closed. (So I'm not using the term 'theory' to mean a logically closed set of sentences, as logicians usually do. Nothing I will say hangs on this terminological difference; I will explicitly mark cases where logical closure is indicated.) The convention that theories have all of L_0 in common is merely a convenience – any theory that lacks part of L_0 can easily be extended to include it; and the non-instrumental vocabularies of all theories can certainly be pooled into one common vocabulary. Theories may, however, differ with regard to their instrumental vocabularies, and it will prove useful to keep their instrumental vocabularies distinct.

Several definitions will facilitate the investigation of the logical properties of instrumental theories.

DEFINITION 1 (The Logical Closure and Logical Equivalence of Theories).

(1.1) $|T|$ is the (semantic) *logical closure* of T (i.e. the set of logical consequences of T) under $L_T \cup L_0$ (the vocabulary occurring in T together with all vocabulary in L_0).

(1.2) For any language L, $|T|_L$ is the set of those sentences in $L \cup L_0$ that are in the (semantic) logical closure of T under the language $L \cup L_T \cup L_0$.

(1.3) Let L_1 and L_2 be the languages of T_1 and T_2. Then, by definition, T_1 is (semantically) *logically equivalent* to T_2 (abbreviated $T_1 \approx T_2$) iff $|T_1|_{L_1 \cup L_2} = |T_2|_{L_1 \cup L_2}$.

It will be convenient in comparing the contents of instrumental theories to compare their logical closures under the vocabulary occurring in T together with all of the non-instrumental vocabulary in L_0. We will have no need to restrict attention to any sub-vocabulary of L_0; so the definition of $|T|$ offered above will be more useful than taking the logical closure of T under just the vocabulary that occurs in it.

We will sometimes want to treat theories that differ in instrumental vocabulary. Thus, '$|T|_L$' can represent the logical closure of T under a possibly larger vocabulary $L \cup L_0$ that contains $L_T \cup L_0$. Later it will be useful to consider the combination of two theories T_1 and T_2 that have only the vocabulary of L_0 in common. Their instrumental parts will be peculiar to each of them. So, although T_1 and T_2 will each have a 2-sorted language, when we conjoin them we will move to a 3-sorted language where sort-0 is the vocabulary of their common non-instrumental parts, sort-1 is the instrumental vocabulary of T_1, and sort-2 is the instrumental vocabulary of T_2. Then $|T_1|_{L_1 \cup L_2}$ is the logical closure of T_1 under the 3-sorted language $L_1 \cup L_2 \cup L_0$. The definition of the expression '$|T|_L$' applies also when L contains a sub-vocabulary of $L_T \cup L_0$. In that case, $|T|_L$ is the result of first generating all logical consequences of T in the joint vocabulary of $L \cup L_T \cup L_0$ and then throwing away all sentences not in $L \cup L_0$.

The sole purpose of the instrumental part of a theory T is to help specify the non-instrumental sub-theory $|T|_{L_0}$. The next definition coins some useful terminology in this regard.

DEFINITION 2 (Literal Parts of Instrumental Theories).

(2.1) Define $\langle T \rangle = |T|_{L_0}$. $\langle T \rangle$ is the *literal component* of theory T.
(2.2) Call any set of sentences with only vocabulary in L_0 *literal*.
(2.3) T is called *trivially instrumental* if $|T| = |\langle T \rangle|_{L_T}$; otherwise, T in *non-trivially instrumental*.

For any theory T, $\langle T \rangle$ is called a *literal* theory because it contains only vocabulary in L_0; it is the subtheory of T that is to be *taken literally*. Sentences in $\langle T \rangle$ represent the ontologically respectable consequences of an instrumental theory. Notice that all literal theories count as trivially instrumental, since literal theories contain only non-instrumental vocabulary. More generally, if the logical closure under vocabulary $L_T \cup L_0$ of the literal subtheory $\langle T \rangle$ gives back *all* of $|T|$, then the instrumental vocabulary of T must play a completely superfluous role in T. In that case each sentence of T that employs instrumental vocabulary will either be a logical truth or logically equivalent to a sentence containing no instrumental vocabulary. Although such a theory is not literal, it is logically equivalent to it's literal subtheory $\langle T \rangle$, i.e. $T \approx \langle T \rangle$. In that case we call T *trivially instrumental*. If T is not logically equivalent to $\langle T \rangle, T$ will be called *non-trivially instrumental*.

Suppose a literal theory T_2 is extended by adding new sentences. The extended theory might be stated in an expanded vocabulary containing instrumental terms and quantifiers not in T_2. The extended theory T_1 is called an L_0-extension of T_2. Whereas T_2 is wholly in L_0, the broader theory T_1 may not be.

DEFINITION 3 (Extensions of Theories in L_0).

(3.1) T_1 is a (semantic) *L_0-extension* of T_2 <u>iff</u> T_2 is literal and $|T_2| \subseteq \langle T_1 \rangle$.
(3.2) T_1 is a (semantically) *conservative L_0-extension* of T_2 <u>iff</u> T_1 is an L_0-extension of T_2 and $\langle T_1 \rangle \subseteq |T_2|$ (thus, $|T_2| = \langle T_1 \rangle$).
(3.3) T_1 is a (semantically) *non-conservative L_0-extension* of T_2 <u>iff</u> T_1 is an L_0-extension of T_2, and $\langle T_1 \rangle \not\subseteq |T_2|$ (thus, $|T_2| \subset \langle T_1 \rangle$).

T_1 is a *conservative L_0-extension* of T_2 just when T_1 has exactly the consequences in the literal language L_0 that T_2 has. If T_1 is a non-conservative L_0-extension of T_2, it contains T_2 plus additional consequences in L_0. Notice that T_1 is a conservative L_0-extension of $\langle T_1 \rangle$ and only of $\langle T_1 \rangle$. That T_1 is a conservative L_0-extension of $\langle T_1 \rangle$ follows from $|\langle T_1 \rangle| = \langle T_1 \rangle$; that T_1 is a conservative L_0-extension of no other logically closed theory follows from the definition of *conservative L_0-extension*, since if T_1 is a conservative L_0-extension of T_2, then T_2 is literal and $|T_2| = \langle T_1 \rangle$.

For any given theory T couched in a 2-sorted language (whether first or second order), if the sorts correctly mark the non-instrumental/instrumental distinction, then $\langle T \rangle$ is just the part of T that is supposed to be taken literally. If T is first order and has a decidable set of axioms, then $\langle T \rangle$ will also be recursively axiomatizable, as an obvious extension of Craig's Theorem shows.

THEOREM 1 (Craig's Theorem). *If a first order theory $|T|$ is recursively axiomatizable, then so is $\langle T \rangle$.*

Instrumentalists often appeal to Craig's theorem in support of the idea that the instrumental part of a theory can in principle be jettisoned. The Craigian axiomatization of $\langle T \rangle$ is not pretty, and it is surely beyond our ability to make practical use of it. But this observation simply reinforces the *appeal* of the instrumental version of the theory. In principle, then, theories in first order logic need only employ instrumental vocabulary to tidy things up. For second order theories the instrumental component of the theory can assist in the specification of logical consequences that may not be captured by any *recognizable* axiomatization in the non-instrumental vocabulary. Even when there is a decidable set of second order axioms that semantically captures $|T|$, the intended non-instrumental subtheory $\langle T \rangle$ may not be specifiable even semantically by a decidable set of axioms in the non-instrumental vocabulary. Thus, for second order languages the instrument may be *required* in order to completely express the theory.

Good scientific theories are hard to come by. If the use of avowedly instrumental devices aids the scientist in expressing those components of theories that he takes literally and believes may be true, then I see no reason to deny him their help. Scientists should be permitted all of the logical tools they can muster, including the use of instrumental bits of syntax or reference to convenient fictions. But when two theories that contain *avowedly instrumental* parts are used in conjungtion the theorist should be willing to take certain precautions, as the next section will make clear.

4. INSTRUMENTALISM AND THE CONJUNCTION OBJECTION

Suppose some scientific theory T_1 has a purely instrumental (perhaps fictitiuous) component and that we have come to believe that its non-instrumental subtheory $\langle T_1 \rangle$ is true. Consider, for instance, the case where the instrumental theory T_1 is a bit of "honest *ad hocery*" constructed for the sole purpose of enhancing the computability of consequences for some highly supported theory, T, about which we are realists. If T_1 is finely tuned to represent T precisely, then the set of non-instrumental consequences of T_1 should coincide with the logical closure of T (i.e. $\langle T_1 \rangle = |T|$); so T_1 will be a conservative L_0-extension of T. If an "alternative" to T_1, say T_3, is also designed as a mere computational tool for T, then it too should be a conservative L_0-extension of T. In this situation there need be no debate about which theory, T_1 or T_3, is *best*. They may each have advantages for modeling different kinds of physical systems to which T applies. But whatever the advantages of T_1 and T_3 in various contexts, their instrumental parts are either false or meaningless, and the instrumentalist *knows them to be so*. Thus, the instrumentalist has no reason to believe anything that either T_1 or T_3 might imply beyond what T asserts.

Now suppose that another instrumental theory T_2 is developed to cover a different range of physical phenomena than T_1 and T_3, and suppose that the evidence leads us to believe that its non-instrumental part $\langle T_2 \rangle$ is true. Normally we are warranted in believing the conjunction of two theories that we believe to be true; at least we are warranted in investigating their conjunction with some reasonable hope that it is true. In particular, the theory $\langle T_1 \rangle \cup \langle T_2 \rangle$ (which is logically equivalent to $\langle T_3 \rangle \cup \langle T_2 \rangle$) will usually have *new* non-instrumental consequences not implied by either $\langle T_1 \rangle$ or $\langle T_2 \rangle$ separately, and it is perfectly reasonable to expect these new consequences to be in agreement with observations. Consider, however, the direct conjunction of the instrumental theories, $T_1 \cup T_2$. Generally the new non-instrumental consequences of $\langle T_1 \cup T_2 \rangle$ will exceed those of $\langle T_1 \rangle \cup \langle T_2 \rangle$. But, insofar as the instrumentalist believes that the instrumental parts of T_1 and T_2 are *false or meaningless*, she has absolutely no basis on which to expect consequences in *excess* of $|\langle T_1 \rangle \cup \langle T_2 \rangle|$ to be true. Indeed, $\langle T_3 \cup T_2 \rangle$ (and an infinite number of other instrumental theories which, like T_3, agree with $\langle T_1 \rangle$) may differ radically in *excess consequences* from $\langle T_1 \cup T_2 \rangle$, (whereas, $\langle T_3 \rangle \cup \langle T_2 \rangle \approx \langle T_1 \rangle \cup \langle T_2 \rangle$). If T_1 has no better tie to reality than a host of alternative instruments, then there is no reason to think that its *excess* non-instrumental consequences have the least chance of being true. (It would, for instance, be ludicrous to spend large sums of research money in the hope of verifying *excess consequences* of $\langle T_1 \cup T_2 \rangle$, especially in cases where the proposed experiment is unlikely to yield any useful information should $\langle T_1 \cup T_2 \rangle$ be false.) Thus, realists argue, to the extent that one believes the *excess consequences* of $\langle T_1 \cup T_2 \rangle$ rather than simply the consequences of $\langle T_1 \rangle \cup \langle T_2 \rangle$, to that extent one treats these theories realistically rather than instrumentally – as though they were *true* rather than merely empirically adequate.

The following theorem and corollary establish that all non-trivially instrumental theories carry excess non-instrumental import that may show up when they are conjoined with other instrumental theories. This is the basis for the version of the conjunction argument presented above.

THEOREM 2 (Conjunctive Equivalence is Logical Equivalence). *If T_1 is non-trivially instrumental, $\langle T_1 \rangle$ is (semantically) incomplete, and $T_3 \not\approx T_1$, then for some finite non-trivially instrumental theory T_2 that is (semantically) consistent with both T_1 and T_3, $\langle T_3 \cup T_2 \rangle \neq \langle T_1 \cup T_2 \rangle$.*[7]

Richard Boyd (1973) draws on a version of Theorem 2 to argue that only logically equivalent theories are *really* empirically equivalent, empirically equivalent under conjunction with all possible auxiliary theories. The theorem says that if T_1 is an instrumental theory that leaves some non-instrumental assertions undetermined, then for any theory T_3 that *differs at all* from T_1 in its logical implications there is a possible auxiliary theory T_2 that in conjunction with T_3 yields *different non-instrumental consequences* than the conjunction of T_2 with T_1. Theorem 2 implies the following corollary.

COROLLARY (Non-conservativeness of Conjoint Instrumental Theories). *If T_1 is non-trivially instrumental and $\langle T_1 \rangle$ is (semantically) incomplete, then there is a finite non-trivially instrumental theory T_2 that is* *(semantically) consistent with* T_1 *such that* $T_1 \cup T_2$ *is a (semantically) non-conservative* L_0*-extension of* $\langle T_1 \rangle \cup \langle T_2 \rangle$.[8]

So, when T_1 and T_3 are distinct instrumental theories that agree on non-instrumental claims, they will yield non-equivalent conjoint theories (i.e. $|T_3 \cup T_2| \not\subseteq |T_1 \cup T_2|$) which will both be non-conservative L_0-extensions of a common non-instrumental subtheory (i.e. $\langle T_3 \rangle \cup \langle T_2 \rangle \approx \langle T_1 \rangle \cup \langle T_2 \rangle$). The instrumentalist may have good reason to expect that the conjoint theory $\langle T_1 \rangle \cup \langle T_2 \rangle$ is true, but only the realist about T_1 and T_2 should have any reason to believe that the *excess consequences* in $\langle T_1 \cup T_2 \rangle$ are true (rather than, say, those of $\langle T_3 \cup T_2 \rangle$, or merely the consequences of $\langle T_1 \rangle \cup \langle T_2 \rangle$).

There is one important special case in which the conjunction of instrumental theories implies precisely the same non-instrumental consequences as the conjunction of the non-instrumental subtheories. Suppose each of the two instrumental theories T_1 and T_2 are 2-sorted, but that they have only sort-0 in common. That is, suppose that the two theories can be construed as employing a common 3-sorted language, where T_1 is stated solely in terms of sort-0 and sort-1, and T_2 is stated solely in terms of sort-0 and sort-2, so that $L_{T_1} \cap L_{T_2} \subseteq L_0$. If the language of the theories is first order, the following theorem applies.

THEOREM 3 (Conservativeness Under Conjunction for Theories with Distinct Instrumental Parts). *For first order languages, if* $L_{T_1} \cap L_{T_2} \subseteq L_0$*, then* $T_1 \cup T_2$ *is a conservative* L_0*-extension of* $\langle T_1 \rangle \cup \langle T_2 \rangle$ *(i.e.,* $|\langle T_1 \rangle \cup \langle T_2 \rangle| = \langle T_1 \cup T_2 \rangle$*).*[9]

If the instrumental language employed by two theories is kept separate by sorting, and the only language that they share is about purportedly real things, then the theories may be used conjointly without exceeding the implications of their true subtheories. Indeed, if $\langle T_3 \rangle = \langle T_1 \rangle$ and $L_{T_3} \cap L_{T_2} \subseteq L_0$, then $\langle T_3 \cup T_2 \rangle = \langle T_1 \cup T_2 \rangle$,[10] and both are conservative L_0-extensions of $\langle T_1 \rangle \cup \langle T_2 \rangle$.

Suppose an instrumentalist wants to conjoin theories that were developed to account for disparate parts of the physical realm, but which have some instrumental vocabulary in common. And suppose she also wants to continue to take advantage of the instrumental features of these theories, and yet wishes to avoid unwarranted *excess consequences* which are mere artifacts of how the instruments fit together. Theorem 3 suggests the following strategy: before conjoining theories replace all common instrumental vocabulary (including variables) with new vocabulary that is distinct for each theory. Then Theorem 3 guarantees that the only new consequences that the conjunction will imply are those that are already consequences of the conjunction of their non-instrumental parts. Thus, one need not give up the convenience of the instrument in order to avoid unwarranted consequences. Since the instrumental vocabulary is merely a tool, the instrumentalist should have no qualms about making these syntactic alterations. The practicing scientist who knowingly employs convenient fictions (e.g. for computational or heuristic purposes) should be happy to forego consequences that are mere artifacts of those fictions, consequences that do not flow from the part of the theory that she takes to be true.

There is one important exception to the recommendation that the instrumental parts of theories should be kept syntactically distinct. Almost all instrumentalists will maintain that the mathematical parts of theories are part of the instrumental component. And although one could relabel the mathematical parts of two instrumental theories so that each has its own distinct vocabulary for numbers and sets, this tactic would be exceedingly unnatural. As it turns out, mathematical instrumentalists can beat the conjunction problem without going to such extremes.

5. MATHEMATICAL INSTRUMENTALISM BEATS THE CONJUNCTION OBJECTION

Mathematical instrumentalism is more tenable than scientific instrumentalism on two counts. First, for a given scientific theory it is usually relatively easy to distinguish the occurrences of quantifiers and non-logical vocabulary that belong to mathematics from those that purport to refer only to a non-mathematical realm. By contrast, scientific instrumentalists generally have a much harder time finding a *principled* way of distinguishing the "observable" (or otherwise "empirically kosher") parts of the vocabulary of a given scientific theory from the purely instrumental part. Secondly, although the conjunction objection establishes that truly instrumental scientific theories should not be conjoined (unless they share no instrumental vocabulary), the conjunction objection does not apply to mathematical instruments. For, if a mathematically endowed scientific theory is *properly fleshed out* with qualitative relations (e.g. various betweenness and congruence relations) that reflect the physical import of the mathematical component of the theory (e.g. a distance function, a mass density function, a gravitational potential function), then the theory will not contain *excess* non-mathematical content – no *excess* consequences can emerge when such a theory is conjoined with other such mathematically endowed theories. The point of the present section is to establish this claim.

I will treat mathematical instrumentalism in the context of 2-sorted languages, as described in Section 2. The non-logical vocabulary and the quantifiers are divided into the non-mathematical (sort-0) sort and the mathematical (sort-1) sort. The language may either be first or second order. All definitions in Section 3 remain as stated, but now L_0 (the sort-0 vocabulary) encompasses all and only the non-mathematical part of the language. '$\langle T \rangle$' now represents the result of taking the (semantic) logical closure of T and throwing away all sentences containing mathematical vocabulary.

Various scientific theories may employ somewhat different parts of mathematics. A given theory might, for example, employ only the axioms for real numbers together with functions that map parts of the world to numbers. But since all standard mathematical theories can be constructed (or emulated) within set theory, it is reasonable to suppose that all of the mathematics that a scientific theory employs may be subsumed under the usual set-theoretic *definitions* (or emulations) of mathematical terms.

If set theory is to be of any use to a scientific theory it will have to be an *applied* set theory, a set theory in which there is a set of the purportedly *real things* (i.e. *urelements*). Also, for any description (open sentence) in the non-mathematical vocabulary there should be a set of all the (n-tuples of) *real things* that satisfy it. For definiteness let applied ZFC (applied ZF with the axiom of choice) be the set theory in which the mathematics of the sciences is couched. In applied first order ZFC all instances of the *separation* and *replacement* axiom schemata, including those applied to expressions containing non-mathematical vocabulary, are axioms. I will use the expression 'S_{L_0}' to refer to ZFC applied the sort-0 vocabulary. In the context of first order scientific theories 'S_{L_0}' will represent first order applied ZFC (see Montague 1965 and Suppes 1960 for the technical details); in the context of second order scientific theories 'S_{L_0}' will designate second order applied ZFC (see Montague 1965). Set-theoretic vocabulary and quantifiers over sets are of sort-1; sort-0 contains quantifiers over non-sets, second order quantifiers over properties and relations among non-sets, and all other non-mathematical vocabulary.

In *Science Without Numbers* Field (1980) proved that the conjunction of applied set theory (either first or second order) with any non-mathematical theory (of the same order) yields a semantically conservative extension of the non-mathematical theory. We may state Field's Conservativeness theorem for set theory in terms of the 2-sorted language as follows:

THEOREM 4 (Conservativeness of Set Theory). $\langle T \rangle \cup S_{L_0}$ *is a conservative* L_0*-extension of* $\langle T \rangle$.

Here T is either first or second order, and the set theory is of the same order as T. If T is non-mathematical, then $\langle T \rangle$ is just $|T|$ (the semantic logical closure of T) and the theorem says that $(T \cup S_{L_0})$ is a conservative L_0-extension of T. When T is a mathematical scientific theory the Conservativeness theorem does not apply to it directly, but only to its non-mathematical subtheory $\langle T \rangle$.

Scientific theories are usually mathematically endowed from inception. The exercise of first stripping scientific theories of their native mathematics and then (re)introducing set theory into the remaining non-mathematical theory may appear to be pointless. So Theorem 4 may seem to have little or no interesting implications for real scientific theories. But in *Science Without Numbers* Field finds significant work for set-theoretic Conservativeness to do.

Field constructs a *reasonably attractive* non-mathematical axiomatization for a theory N that is intended to capture all of the non-mathematical consequences of a standard mathematical version of second order Newtonian gravitation theory, P. In both P and N, as Field expresses them in (1980), the second order quantifiers range only over arbitrary regions of space-time. Field then uses the Conservativeness of set theory together with a Representation theorem to *prove* that N does indeed capture all of the non-mathematical consequences of P, i.e. $|N| = \langle P \rangle$. In "On Conservativeness and Incompleteness" (1989, Ch. 4) Field shows how to get a similar result for a first order version of mathematical Newtonian gravitation theory, P_0^-. He constructs a *reasonably attractive* axiomatization for a non-mathematical first order theory N_0, and proves that $|N_0| = \langle P_0^- \rangle$. In a moment we will see how set-theoretic Conservativeness theorem and the Representation theorems work together to prove that N (N_0) captures precisely the non-mathematical part of P (P_0^-).

For my purposes the precise details of Field's Representation theorems are not important. What is important is that when a representation theorem of this kind holds, it establishes the following *logical relationship*, which I'll call 'Rep', between a mathematical theory T_1 and a non-mathematical theory T_2.

DEFINITION 4 (Representation Condition).

$$\text{Rep}(T_1 : T_2) \quad \underline{\text{iff}} \quad \begin{array}{l} (1)\ T_1 \text{ is an } L_0\text{-extention of } T_2, \text{ and} \\ (2)\ |T_1| \subseteq |T_2 \cup S_{L_0}|. \end{array}$$

In order for $\text{Rep}(T_1 : T_2)$ to hold, the first clause of the Representation Condition requires that every sentence in T_2 is a non-mathematical logical consequences of T_1 (i.e. $|T_2| \subseteq \langle T_1 \rangle$). The second clause is much more substantive. It requires that adding applied set theory to T_2 yields all of the consequences of the mathematical version of scientific theory T_1.

For a given pair of theories T_1 and T_2, $\text{Rep}(T_1 : T_2)$ is a very strong claim. It does not derive from the Conservativeness of set theory, but (when it holds) must be proved in its own right for individual pairs of theories. Field's Representation theorems show that $\text{Rep}(P : N)$ holds for the second order versions of Newtonian gravitation theory described above, and that $\text{Rep}(P_0^- : N_0)$ holds for the first order version. Thus, $|N| \subseteq |P| \subseteq |N \cup S_{L_0}|$ and $|N_0| \subseteq |P_0^-| \subseteq |N_0 \cup S_{L_0}|$ (where 'S_{L_0}' represents applied set theory of the appropriate order for each case). Field's Conservativeness theorem for set theory implies that $\langle N \cup S_{L_0} \rangle = |N|$ and $\langle N_0 \cup S_{L_0} \rangle = |N_0|$. Combining the Representation and Conservativeness theorems then yields $\langle P \rangle = |N|$ and $\langle P_0^- \rangle = |N_0|$ – i.e. P is a conservative L_0-extension of N, and P_0^- is a conservative L_0-extension of N_0.

The conservative extension results just stated are only of interest because they show that N and N_0 provide *attractive axiomatizations* of the non-mathematical parts $\langle P \rangle$ and $\langle P_0^- \rangle$ of the mathematical theories P and P_0^-, respectively. If we simply wanted to find *some* non-mathematical theory of which P is a conservative extension, attractive or not, $\langle P \rangle$ is always ready to hand, and Field's result are irrelevant to the issue. For, it is already perfectly obvious that P is a conservative L_0-extension of $\langle P \rangle$. Indeed, in the case of the recursively axiomatizable first order theory P_0^-, we even know that $\langle P_0^- \rangle$ is recursively axiomatizable (without appeal to N_0) via Craig's theorem. So it seems that the whole point of Field's version of mathematical instrumentalism, its only advantage over the old fashioned kind (e.g. relying on Craig's theorem) is this: given an arbitrary scientific theory, T, we generally will not know whether its non-mathematical subtheory $\langle T \rangle$ is *nicely axiomatizable*; but whenever a nice collection of axioms N conjoined with applied set theory S_{L_0} *reproduces* the original theory T, we can infer that the non-mathematical subtheory of T is *nicely axiomatizable* by N. Field's approach, however, can be enlisted to do much more for the cause of mathematical instrumentalism than provide a way to prove that some theories are nice axiomatizations of the non-mathematical parts of others.

Let us put aside the issue of whether the non-instrumental part $\langle T \rangle$ of a theory T has a *nice* axiomatization, and let's just consider what the relation $\text{Rep}(T : \langle T \rangle)$, when it holds, implies about T. Why should we care about whether Rep(T : $\langle T \rangle$) holds? After all, T is automatically a conservative L_0-extension of $\langle T \rangle$. So what is to be gained when the Representation Condition holds for T? To understand the importance of Rep(T : $\langle T \rangle$), suppose that it fails for some theory T. Indeed, we need not look far to find an example of such a theory. Let P_0 be the first order platonistic version of Newtonian gravitation theory P gotten by replacing each second order axiom of P with an axiom schema (see Field, 1980, Ch. 9; 1989, Ch. 4). Although Field established that Rep(P : N) holds, Shapiro (1983) proved that Field's first order version N_0 of the second order non-mathematical Newtonian theory N does not satisfy Rep(P_0 : N_0). And Field (1989, p. 133) observes that a similar argument shows that there is no "natural" non-mathematical subtheory of P_0 for which a Representation theorem holds. But, what is more important for our purposes, Rep(P_0 : $\langle P_0 \rangle$) *does not hold* either. Applying set theory to $\langle P_0 \rangle$ does not reproduce all of the mathematical-physical implications of P_0.

Why does Rep(P_0 : $\langle P_0 \rangle$) fail? Because the mathematical part of P_0 connects up mathematical things with physical things in a way that implies more about the world than would result from simply applying set theory to $\langle P_0 \rangle$.[11] And although this extra connection between the mathematical and the physical does not show up directly in $\langle P_0 \rangle$, it can show up when P_0 is conjoined with other mathematical theories. Field only regains a Representation theorem for a weaker version of first order Newtonian gravitation theory, the theory P_0^- mentioned earlier. Rep(P_0^- : $\langle P_0^- \rangle$) does hold; and $\langle P_0^- \rangle$ turns out to be the nicely axiomatizable theory N_0 mentioned above. (Although P_0^- is not as strong as P_0, P_0^- is an intuitively plausible version of first order Newtonian gravitation theory; it is precisely the sort of mathematical theory that Burgess (1984) shows is a conservative extension of a synthetic subtheory.) Thus, although T is always an L_0-extension of $\langle T \rangle$, the other claim that 'Rep(T : $\langle T \rangle$)' makes about T, that $|T| \subseteq |\langle T \rangle \cup S_{L_0}|$, is far from trivial.

The real payoff of Rep(T : $\langle T \rangle$), when it holds, is that it establishes the non-mathematical part of T as sufficiently *fleshed out* that no extra connections between the mathematical and the physical are coded into the formalism of T. So no excess non-mathematical consequences can emerge when T is conjoined with other *fleshed out* theories. The following theorem established this claim.

THEOREM 5 (Conservativeness Under Conjunction for Representable Theories). *If* Rep(T_1 : N_1) *and* Rep(T_2 : N_2), *then* $T_1 \cup T_2$ *is a conservative* L_0*-extension of* $N_1 \cup N_2$.[12]

Theorem 5 tells us that the non-mathematical subtheory of the conjunction of two theories $\langle T_1 \cup T_2 \rangle$ has precisely the (semantic) consequences of the conjunction of each of their non-mathematical parts $|N_1 \cup N_2|$, provided only that each respective pair of theories satisfies the Representation Condition. Here each N_i may be nicely axiomatizable or not. In any case, $\mathrm{Rep}(T_i : N_i)$ guarantees that $|N_i| = \langle T_i \rangle$ (see the first full paragraph after Definition 4).

Conservation under conjunction, as expressed by Theorem 5, lends credence to the physicalist contention that the mathematics employed in science can, and should, function as a mere instrument. Scientific instrumentalism cannot generally attain a similar *conservation under conjunction result* unless the instrumental languages of the theories to be conjoined share no instrumental vocabulary (and Theorem 3 applies). Theorem 5 suggests that the physicalist should strive to develop theories for which the Representation Condition is satisfied. Then she may rest assured that the theory insinuates no spooky non-physical connection by which the mathematical realm influences the physical.

Theorem 5 may give some comfort to scientific instrumentalists, too. The scientific instrumentalist should want to inhibit the generation of excess non-instrumental consequences when theories are conjoined (since they are mere artifacts of the particular instruments being combined). Theorems 3 and 5 together recommend some constraints that will ensure that *excess* consequences will not flow from the instrumental parts of his theories. Suppose two first order theories T_1 and T_2 each satisfy a Representation Condition (i.e. $\mathrm{Rep}(T_1 : \langle T_1 \rangle)$ and $\mathrm{Rep}(T_2 : \langle T_2 \rangle)$ hold). And suppose that T_1 and T_2 also have non-mathematical instrumental terms and quantifiers, but have none in common. Then Theorems 3 and 5 together imply that the conjunction of the two theories will entail only non-instrumental consequences entailed by the conjunction of their non-instrumental subtheories.[13] So, when preparing to conjoin theories T_1 and T_2 for which $\mathrm{Rep}(T_1 : \langle T_1 \rangle)$ and $\mathrm{Rep}(T_2 : \langle T_2 \rangle)$ hold, one may beat the conjunction problem merely by ensuring that the non-mathematical parts of the instrumental vocabularies are kept distinct.

6. MATHEMATICAL INSTRUMENTALISM THE CONJUNCTION OBJECTION IN MODAL METALOGIC

The Conservativeness of set theory (Theorem 4) and Conservativeness Under Conjunction for Representable Theories (Theorem 5) rely on model theory for their proofs. The notion of *logical consequence* they employ is defined in the usual set-theoretic way in terms of set-theoretic *inter-*

pretations of a formal language. So if a mathematical instrumentalist appeals to these theorems in support of her views, she invokes a *theory of the logical consequence relation* that she must ultimately believe to be *untrue*. Has the mathematical instrumentalist any reason to believe metalogical claims, including claims about conservativeness, derived as they are from set-theoretic semantics?

Field deals with this kind of problem by reformulating metalogic in a way that does not draw on set theory. Field introduces a modal operator for *logical truth* that provides a way to represent the *logical consequence relation* without presupposing set theory (see Field, 1989, Ch. 3; 1992). He then shows how important metalogical results on which a mathematical instrumentalist would like to draw can be reproduced in the resulting *modal metalogic*. In particular, Field represents the Conservativeness of set theory (Theorem 4) modally. In this section I will show how to secure a version of the *Conservativeness Under Conjunction* theorem (Theorem 5) in modal metalogic.

To capture the notion of logical consequence without appealing to set theory Field introduces a modal operator '$\Box$' for *logical necessity* into the object language (either first or second order) in which theories are expressed, and he sets down axioms for it. These axioms cannot, of course, be complete; but neither is set theory a complete recursively axiomatizable theory. The axioms for the logical necessity operator are those for quantified S-5 together with an anti-essentialist axiom – i.e. an axiom which implies that each formula $\Box P$ (where P may have x free) is logically equivalent to $\Box(x)P$. And Field introduces several additional axioms that are intuitively plausible for logical necessity.

Field (1992) represents the claim that set theory is conservative as a sentence schema for sentences in the language of modal metalogic. He labels the schema '(C#)'; schema (C$) in the following thesis is equivalent to (C#) (but adapted to many-sorted logic). I will call the claim that all instances of (C$) hold 'Thesis (C$)'. The thesis asserts a modal version of the Conservativeness of set theory (Theorem 4).[14]

THESIS (C$) (Modal Version of the Conservativeness of Set Theory). *Let H and A be any sentences of L_0, and let SS be any conjunction of a finite number of the axioms of the set theory S_{L_0}. All instances of the following schema hold:*

$$\text{(C\$)} \qquad \Box((SS \& H) \supset A) \supset \Box(H \supset A).$$

For first order logic S_{L_0} has an axiom schema, and so an infinite number of (potential) axioms. 'SS' represents any finite conjunction of these axioms. For second order logic S_{L_0} has a finite list of axioms, so we may simply let SS be their conjunction.

In order to capture the Conservativeness Under Conjunction theorem in modal metalogic we must first state modal versions of the notions *conservative L_0-extension* and Rep($T_1 : T_2$).

DEFINITION 5 (Extensions of Theories in L_0 for Modal Metalogic).

(4.1) T_1 is a *modal L_0-extension* of T_2 iff

(1) T_2 is a decidable list of sentences (axioms) in L_0 and T_1 is a decidable list of sentences (axioms) in a 2-sorted language that includes L_0 as one sort, and
(2) for each axiom H of T_2 there is a conjunction K of axioms of T_1 such that $\Box(K \supset H)$.

(4.2) T_1 is a *modal conservative L_0-extension* of T_1 iff

(1) T_1 is a modal L_0-extension of T_2, and
(2) for all sentences A in L_0 and any conjunction K of axioms of T_1, if $\Box(K \supset A)$, then there is a sentence H, a conjunction of axioms of T_2, such that $\Box(H \supset A)$.

If we are to avoid using set theory in the specification of logic, then we can no longer think of theories as arbitrary *sets* of sentences. But no useful scientific theory will be just an arbitrary set of sentences. We now may think of theories as specified by some list of axioms. A theory may have either a finite or a (potentially) infinite list of axioms, perhaps specified by axiom schemata. At most we need only require that there is some effective way to decide whether any given sentence is an axiom of the theory.

Given any two theories T and N specified in terms of a decidable list of axioms, the previous definition says that for T to be a *modal conservative L_0-extension* of N, each axiom H of N must be *necessitated* by some conjunction K of axioms of T (i.e. $\Box(K \supset H)$), and every L_0 sentence that is necessitated by a finite conjunction of T's axioms must also be necessitated by some finite conjunction of N's axioms. So, for any list of axioms N, $S_{L_0} + N$ is a *modal conservative L_0-extension* of N (by Thesis (C$)). (When theories Q and R have infinite lists of axioms let '$Q + R$' denote the theory resulting from alternating between their lists of axioms.)

The modal version of the Representation Condition (Definition 4) runs as follows:

DEFINITION 6 (Modal Representation Condition). M-Rep($T : N$) iff

(1) T is a modal L_0-extension of N, and
(2) for each axiom K of T there is a sentence H formed from a conjunction of axioms of N and a conjunction SS of axioms of the set theory S_{L_0} such that $\Box((SS \,\&\, H) \supset K)$.

If T is an axiomatized first order theory, then clause 1 of the condition is always satisfied for the N that is the Craigian axiomatization of the L_0 part of T. Even so, clause 2 will not always hold for a Craigian N (recall the discussion of P_0 in Section 5).

As with Rep($T_1 : T_2$) in Section 5, M-Rep($T : N$) makes a substantive claim. It claims that the mathematical part of T does not connect up mathematical things with physical things in a way that implies more about the world than would result from simply adding set theory to its non-mathematical subtheory N. When M-Rep($T : N$) holds, the non-mathematical part of T is sufficiently fleshed out that no excess non-mathematical consequences (due to extra connections between the mathematical and the physical) can emerge when other mathematical theories are conjoined with T. The next theorem establishes this claim.

THEOREM 6 (Modal Version of Conservativeness Under Conjunction). *If* M-Rep($T_1 : N_1$) *and* M-Rep($T_2 : N_2$), *then* $T_1 + T_2$ *is a modal conservative* L_0*-extension of* $N_1 + N_2$.[15]

Thus, when theories satisfy the *Modal Representation Condition* the mathematical instrumentalist is entitled to believe that their mathematical parts are *merely* instrumental and carry no *hidden excess content* about the physical world that may emerge under conjunction with other (auxiliary) theories.

7. CONCLUSION

Instrumentalism is a view about language, logic, and ontology, but it is usually motivated by epistemological concerns, by views about how we assign meanings to words and how we can come to know truths about the things to which our words refer. The mathematical instrumentalist need not eschew everything that philosophers have called *abstract*. He need not be a nominalist. He may, for instance, hold that there are a host of natural physical kinds and relations instantiated by physical objects, and that we refer to them and know truths about them through the causal influences of their physical instances.

One reason often cited by mathematical instrumentalists for denying the existence of pure mathematical objects like sets and numbers is that mathematical things do not participate in the physical world (whereas natural kinds do). Mathematical things, if they exist, are not only far removed from our senses, but, being non-physical, they can have no causal influence on us. Hence, no physicalist theory of reference can tie words to set-theoretic objects. And even if sets and numbers exist and we can somehow refer to them, still they can have no causal influence on, nor any *physical* relationship to the physical world. And if no causal or other physical relationship applies, then the physical world must be completely autonomous from the mathematical realm,[16] so:

(1) there is no way for physical beings like us to come to *know* anything about them;
(2) physical theories should be autonomous *in principle* from mathematical terms – i.e. the non-mathematical parts of a true physical theory should capture *all* of its physical content.

And indeed theories are autonomous from mathematics in principle when the *Rep* relations hold.

A Representation theorem for a mathematically endowed theory establishes that the theory satisfies a Rep relation, and this in turn guarantees that all of its non-mathematical content is explicitly expressed by its non-mathematical sentences. When the *Rep* condition holds the mathematics employed in a theory harbors no *excess hidden physical content* coded into the mathematical sentences. One may safely conjoin the mathematical versions of a pair of theories (when Rep($T_1 : \langle T_1 \rangle$) and Rep($T_2 : \langle T_2 \rangle$) hold) without fear that unintended, unsupported, excess physical content will emerge. Thus, the mathematical terms in such theories do indeed play a *purely instrumental* role as part of a convenient systematizing scheme.

If a *scientific instrumentalist* sincerely maintains that the theoretical component of a theory is just a useful tool that reflects reality no better than observationally equivalent instruments, then she should consider any excess observational consequences that are produced when theories are conjoined to be mere artifacts of the instruments she happens to have at hand. She should give no more credence to these excess observational consequences than to the alternative excess consequences produced by conjoining different instrumental representations of observationally equivalent theories. If theoretical expressions in scientific theories are supposed to be merely instrumental, then it seem reasonable to insist that they should be *as merely instrumental* as mathematical expressions turn out to be when *Rep* relations holds.

The argument I have offered for the superiority of mathematical instrumentalism over the more usual kind of scientific instrumentalism draws on *semantic* logical concepts to make the case. Commenting on Field's work in *Science Without Numbers*, some philosophers (e.g. Shapiro, 1983, 1993; and Hellman, 1989) have called into question the appropriateness of using semantic logical notions to carry out the instrumentalist project. More generally, these philosophers argue that Field's approach fails on the following grounds: (1) the semantic notions of logical consequence and conservativeness are inappropriate for a defense of nominalism; (2) although mathematics is proof-theoretically conservative for first order logic, it is not so for second order logic; (3) Field's Representation theorem for Newtonian gravitation theory holds for the second order version of the theory, but not for its first order counterpart. Thus, they conclude, the *required* proof-theoretic version of set-theoretic Conservativeness cannot be combined with an appropriate Representation theorem to support Field's claim that the mathematical part of the Newtonian theory is purely instrumental. I will briefly address each of these points and the attendant issue of whether mathematical instrumentalism succeeds in overcoming the so-called Quine–Putnam indispensability argument – the argument that scientists should be committed to the existence of mathematical things because mathematics is indispensable to the sciences.

First, I too think that Field's approach to mathematical instrumentalism fails as a convincing defense of *nominalism*. As I see it the so-called "nominalistic" physics is committed to natural physical kinds and relations (e.g. is-a-region, has-mass-density-between, is-congruent-to) whether or not it quantifies over them. But the real worry raised by those who criticize Field for using semantic notions is not nominalism *per se*, but whether a *mathematical instrumentalist* can use a semantic notion of logical consequence in making the case that mathematics may be treated as a fiction. Doesn't the semantic notion of logical consequence depend on mathematics (i.e. set theory) for its definition?

Of course, when the logic involved is first order all results in Sections 4 and 5 can be restated and proved in proof-theoretic terms. But I think that even the *is-provable-from* relation does not provide the appropriate *concept* of logical consequence. Deduction would be a mere syntactic game were it not thought to soundly reflect the conception of logical consequence as the relation of truth preservation.

Although it has become common practice to treat the (semantic) logical consequence relation as *defined* in terms of set theory, I think this is a mistake. The set-theoretic "definitions" of logical consequence and logical consistency serve only to model (i.e. emulate) these notions, not ultimately to *define* them. If set theory provided the *fundamental definitions* of logical consistency and logical consequence, then to be a logical consequence of the axioms and definitions of set theory is just to satisfy the set-theoretic definition of logical consequence. Is the *logical relationship* that holds between set-theoretic theorems and the axioms just an additional relationship between sets? Does the assertion that a set-theoretic theorem C is a *logical consequence* of set-theoretic axioms B amount to the claim that every (set-theoretically defined) model M of B is a model of C? If so (and if the assertion is true), then presumably the model-theoretic claim is itself a *logical consequence* of the axioms of set theory – i.e. in every model M' of the axioms of set theory, every model M of B in the model M' is a model of C. And presumably *this* claim about models of models of set theory is in turn a *logical consequence* of the set-theoretic axioms. And if this iteration effect for a set-theoretic "definition" of semantic logical consequence isn't troubling enough, there is a related quandary about how the assertion that set theory itself is logically consistent is to be understood? Does it mean that there is a set-theoretic model of the set-theoretic axioms (a domain containing all sets or all classes)? When we consider the question of whether the axioms of a set theory are logically consistent, are we simply raising a question about the truth of a (logically?) stronger set-theoretic claim about what sets exist, as the set-theoretic "semantic" definition of *logical consistency* implies?[17]

It seems to me that logical consistency and logical consequence are *fundamental* semantic logical concepts which, rather than presuppose mathematics, are logically prior to it; mathematics *presupposes* logic rather than *defines* it. Clearly for most purposes the mathematical instrumentalist can reason *in accord with* these logical concepts without any appeal to truths of model theory. And when required, the mathematical instrumentalist can also reason *about* logical concepts without any appeal to mathematical things. Field shows how in (1989, Ch. 3), where he represents the logical concepts in terms of modal operators. This provides the mathematical instrumentalist with a modal version of metalogic, which I adapted to my purpose in Section 6.

The main philosophical idea behind modal metalogic is that we have pre-set-theoretic intuitions about the semantic concepts of logical consistency and logical consequence – i.e. logically possible truth and logically guaranteed truth preservation. For most purposes we need not formalize these notions, but merely restrict our deductive inferences to obey rules that accord with them. When it becomes useful to *reason about* the logical notions we need not move to a "definition" of them in terms of set theory. Rather, we may represent them directly in the object language in terms of modal operators, and write down intuitively plausible truths about them as axioms for the *modal metalogic*. From these axioms we can deduce truths about these logical notions in terms of the modal operators using intuitively sound inference rules that reflect our concept of logical consistency and consequence as they apply to the modal sentences themselves. Of course we cannot completely axiomatize modal metalogic, but we cannot completely axiomatize the set-theoretic notion of logical consistency either. (And just as some of our intuitions about the truths of set theory have been subject to revision, so it is *conceivable* that some intuitively plausible principles about logical consequence and logical consistency *may* need revision.)

The modal versions of the main results, Section 6, establish that mathematical instrumentalism can overcome the conjunction objection, and establish this in a manner that does not presuppose the existence of mathematical things. The mathematical instrumentalist can employ *semantic logical notions* to justify the belief that mathematics is merely instrumental without any reliance on set-theoretic semantics. In light of the modal approach to metalogic the second objection commonly raised against mathematical instrumentalism shifts to the issue of whether the *modal semantic* conservativeness of set theory is the right kind of conservativeness for the purposes of the mathematical instrumentalist. For first order logic, applied set theory is both modally (semantically) conservative and proof-theoretically conservative. Applied set theory adds neither new consequences to a non-mathematical theory, nor makes additional consequences provable. Thus, at least for first order theories the mathematics used by a theory T is indeed dispensable in principle when M-Rep$(T : N)$ holds for an axiomatized version N of $\langle T \rangle$.

The modal semantic conservativeness theorem also holds for second order logic. It shows that applied set theory adds no new (semantic) logical consequences to a non-mathematical theory. However, adding applied set theory can supplement the deductive power of a second order theory so that consequences that were not provable without the mathematics become provable with its aid. This might seem like a remarkable gift – the same set of logical consequences, and more of them provable – but this is really no different than introducing additional logically sound deduction rules to the necessarily incomplete set of rules of a given deduction system for second order logic.

I think the above considerations show that at least for first order theories, when *M-Rep* holds for a theory the mathematical instrumentalist successfully circumvents the Quine–Putnam argument; mathematics may be practically indispensible, but is not needed by science *in principle*. (If we were cognitively more powerful we might get along quit well without using any mathematics at all with first order theories.) Does modal semantic conservativeness together with *M-Rep* (when it holds) suffice to mitigate the Quine-Putnam argument applied to second order theories? It seems to me that it does. For, when *M-Rep* holds for a second order theory, the theory does not need mathematics in order to *say* (via logical consequence) what is true. Rather, the mathematics is only needed to assist us in computing consequences that the theory already implies. So (when *M-Rep* holds) one can believe that the theory *is true* without being in any way committed to mathematical things – one can express the theory completely non-mathematically in a way that maintains in tact all of its semantic logical consequences. And because one knows that using the syntax of set theory for deducing consequences is (modally) semantically conservative, one is free to employ mathematics for deduction, knowing full well that whatever one deduces was already a consequence of the theory that one believes. The use of mathematics commits one to no more than the existence of a (modal) sematically conservative syntactic device.

Finally, as to the issue of whether the appropriate Representation theorems hold – for the second order mathematical version of Newtonian gravitation theory P and Field's non-mathematical version of it N, M-Rep$(P : N)$ holds, so there is no problem. (Notice, too, that since I am not defending nominalism, the second order version of N, which only quantifies over regions, should be unobjectionable.) Regarding the first order versions of Newtonian gravitation theory – the mathematical version P_0 is too strong for Rep$(P_0 : \langle P_0 \rangle)$ to hold; but the weaker formulation of P_0 that Field calls P_0^- has much to recommend it over P_0 (recall the discussion in Section 5 above). Field argues persuasively in (1989, Ch. 4) that P_0^- is all of Newtonian gravitation theory that a scientist should want. For one thing, P_0^- is a very plausible rendition of a first order scientific theory, as evidenced by the fact that it is precisely the kind of mathematical scientific theory suggested independently by Burgess (1984). Furthermore, the theory P_0^- appears to capture all but extremely *arcane* consequences of P_0. P_0^- lacks only those physical consequences of P_0 that derive from the way in which P_0 connects up physical things to *arcane properties* of the real numbers, mathematical properties that are provable in full first order set theory but not provable in the first order theory of collections of quadruples of real numbers (see Field 1989, Ch. 4). In addition, M-Rep$(P_0^- : N_0)$ holds for the *nicely axiomatizable* first order version N_0 of Field's non-mathematical second order theory N. Thus, the physicalist should have no qualms about choosing P_0^- over P_0 as a first order rendition of his mathematical physics.

There is of course no *guarantee* in advance that for each of our best true scientific theories T, there will be a non-mathematical theory T' such that M-Rep$(T : T')$ holds. Physicalism is a contingent thesis. Although the physicalist believes that for true physical theories the *M-Rep* relation will indeed hold, he should always be willing to back this belief regarding a given, well supported scientific theory with a clear statement of a non-mathematical version of the theory, and with a Representation theorem. If we should ever become convinced that some mathematically endowed scientific theory T is wholly true but that no Representation result holds for T, then I think we may indeed have good reason to believe in the existence of mathematical things, and to believe that the world is not entirely physical after all.[18]

NOTES

[1] The physicalist might also entertain the prospect of *defining away* the theory's mathematical vocabulary in terms for just physical things and properties. But this logicist strategy has little chance of succeeding for our best current scientific theories.

[2] I will provide a proof of this claim in Section 4.

[3] I will use the term 'mathematical theory' as shorthand for 'mathematically endowed scientific theory'; when referring to a theory of pure mathematics I'll say 'pure mathematical theory'.

[4] Logicians sometimes employ the term 'extension' in a syntactic, proof-theoretic way. That is, T' is a syntactic *extension* of T just in case every sentence *deducible* from T is *deducible* from T' (in some specified deduction system). In this paper a semantic reading of the term 'extension' will be more useful: T' is a (semantic) *extension* of T just in case every (semantic) logical consequence of T is a logical consequence of T'. (Boolos and Jeffrey, 1989, pp. 105–107, 174, 243–244 employ only the semantic reading.) Terms like 'logical consequence' and 'extension' will always have the semantic sense in this paper.

[5] The reader should not construe my defense of mathematical instrumentalism as a defense of *nominalism*. There is enough that is mysterious about mathematical things (e.g. how can they possibly influence the physical?) to warrant reservations about their reality without calling into question the reality of natural physical properties. So, although I will draw heavily on formal results Field proves in *Science without Numbers* – where Field employs this work in defense of nominalism – I will only address the more moderate position that mathematics is instrumental. Indeed, Field emphasizes the defense of mathematical instrumentalism and de-emphasizes nominalism in his (1989). I will discuss this issue a bit more fully toward the end of the paper.

[6] Field (1980, 1989) discusses mathematical instrumentalism in terms of both proof-theoretic and semantic notions. Some philosophers have argued that Field's use of the semantic notions of logical consequence and conservativeness are inappropriate for a

defense of mathematical instrumentalism. I ask the reader to suspend judgement about such issues until after the formal results have been presented. I will discuss these issues in the concluding section of the paper.

[7] Proof. Suppose T_1 is non-trivially instrumental and that $\langle T_1\rangle$ is incomplete. Suppose for every finite T_2 consistent with both T_1 and T_3, $\langle T_3 \cup T_2\rangle \approx \langle T_1 \cup T_2\rangle$. Clearly $\langle T_3\rangle = \langle T_1\rangle$ (i.e. let T_2 be a tautology). So to prove $T_3 \approx T_1$ we only need show that any consequence C_1 of T_1 not in $\langle T_1\rangle$ is in $|T_3|$, and any consequence C_3 of T_3 not in $\langle T_3\rangle$ is in $|T_1|$.

Let C_1 be a consequence of T_1 not in $\langle T_1\rangle$, and let O be a sentence of L_0 with neither O nor $\sim O$ in $\langle T_1\rangle$. Neither $(C_1 \supset O)$ nor $(C_1 \supset\sim O)$ is in $|T_1|$. But $\langle T_1\cup\{(C_1 \supset O)\}\rangle$ contains O; $\langle T_1 \cup \{(C_1 \supset\sim O)\}\rangle$ contains $\sim O$. So, $\langle T_3 \cup \{(C_1 \supset O)\}\rangle$ contains O; $\langle T_3 \cup \{(C_1 \supset\sim O)\}\rangle$ contains $\sim O$. Then T_3 entails both $((C_1 \supset O) \supset O)$ and $((C_1 \supset\sim O) \supset\sim O)$. Thus, T_3 entails C_1.

Let C_3 be a consequence of T_3 not in $\langle T_3\rangle$. The previous argument establishes that there is such a C_3, namely C_1; so T_3 is non-trivially instrumental. Since $\langle T_3\rangle = \langle T_1\rangle$, $\langle T_3\rangle$ must be incomplete too. An argument just like the preceding one (with the subscripts '1' and '3' switched) establishes that T_1 entails C_3.

[8] Just substitute '$\langle T_1\rangle$' for 'T_3' in Theorem 2.

[9] Proof. First, a careful examination of a proof of the Craig Interpolation Theorem (see, e.g., Boolos and Jeffrey, 1989, Ch. 23) shows that it applies to the logic of many-sorted languages specified in Section 2. That is, if A and C are sentences of an n-sorted language in which sorts don't overlap, and A logically entails C, then there is a sentence B containing only individual constants, function and predicate symbols shared by A and C, *and* containing only quantifiers of the sorts shared by A and C, such that A logically entails B and B logically entails C.

Now assume that $L_{T_1} \cap L_{T_2} \subseteq L_0$ and let O be a sentence in $\langle T_1 \cup T_2\rangle$; we show that O is in $|\langle T_1\rangle \cup \langle T_2\rangle|$ as follows. O is of sort-0 and is logically entailed by $T_1 \cup T_2$. From the compactness of first order logic it follows that there are a pair of sentences C_1 and C_2 entailed by T_1 and T_2, respectively, such that $(C_1 \& C_2)$ logically entails O. Then C_1 entails $(C_2 \supset O)$. So there is a sort-0 sentence P entailed by C_1 and entailing $(C_2 \supset O)$ (since the latter sentences have only L_0 vocabulary and quantifiers in common). But then C_2 must entail $(P \supset O)$. Thus, P is in $\langle T_1\rangle$, and $(P \supset O)$ is in $\langle T_2\rangle$. So, O is in the logical closure of $\langle T_1\rangle \cup \langle T_2\rangle$; this establishes that $\langle T_1 \cup T_2\rangle \subseteq |\langle T_1\rangle \cup \langle T_2\rangle|$. Clearly $|\langle T_1\rangle \cup \langle T_2\rangle| \subseteq \langle T_1 \cup T_2\rangle$ holds, since $\langle T_1\rangle$ and $\langle T_2\rangle$ are each entailed by $\langle T_1 \cup T_2\rangle$.

[10] Since $\langle T_3 \cup T_2\rangle = |\langle T_3\rangle \cup \langle T_2\rangle| = |\langle T_1\rangle \cup \langle T_2\rangle| = \langle T_1 \cup T_2\rangle$.

[11] What P_0 has that $\langle P_0\rangle \cup S_{L_0}$ lacks is the mereological completeness schema that Field calls (C_p): for any non-empty *set* of entities (e.g. of space-time regions) there is an entity u that is their mereological sum. The weaker theory P_0^- only implies the schema (C_s): if $(\exists x)Fx$, then there is an entity u that is the mereological sum of the entities y such that Fy (formula F and variables x and y are in L_0). Indeed, $|P_0| \subseteq |\langle P_0^-\rangle \cup S_{L_0} \cup (C_p)|$. See (Field 1989, Ch. 4) for details.

[12] Proof. Suppose Rep($T_1 : N_1$) and Rep($T_2 : N_2$). $(|N_1| \cup |N_2|) \subseteq (|T_1| \cup |T_2|) \subseteq (|N_1 \cup S_{L_0}| \cup |N_2 \cup S_{L_0}|) \subseteq |(N_1 \cup N_2) \cup S_{L_0}|$; so $|N_1 \cup N_2| \subseteq \langle T_1 \cup T_2\rangle \subseteq \langle(N_1 \cup N_2) \cup S_{L_0}\rangle$. And $(N_1 \cup N_2) \cup S_{L_0}$ is a conservative L_0-extension of $(N_1 \cup N_2)$, so $\langle(N_1 \cup N_2) \cup S_{L_0}\rangle = |N_1 \cup N_2|$. Thus, $|N_1 \cup N_2| = \langle T_1 \cup T_2\rangle$.

[13] Notice that the 'L_0' in Theorem 3 represents just the non-instrumental part of the vocabulary; the 'L_0' of Theorem 5 may be taken to represent the L_0 of Theorem 3 taken together with all of the non-mathematical instrumental vocabulary. Each of the two

theories T_1 and T_2 should properly be 3-sorted, each with a distinct non-mathematical instrumental component. When the theories are combined the entire language becomes 4-sorted: sort-0 is their common non-instrumental language, sort-4 is their common mathematical vocabulary, and sort-1 and sort-2 are the non-mathematical instrumental vocabularies unique to T_1 and T_2, respectively.

[14] Thesis (C#) is labeled as a 'thesis' rather than a 'theorem' because although (C#) may be derived from intuitively plausible axioms for the modal logic of logical necessity, (C#) may itself be used as an axiom for this logic (and one of the other axioms can then be derived). See Field (1992).

[15] Proof. Suppose M-Rep($T_1 : N_1$) and M-Rep($T_2 : N_2$). T_1 and T_2 are modal L_0-extension of N_1 and N_2, respectively. So for any axiom H of $N_1 + N_2$ there is a conjunction K of axioms of T_1+T_2 such that $\Box(K \supset H)$. Thus, T_1+T_2 is a modal L_0-extension of N_1+N_2. Now let A be in L_0, and K_1 and K_2 be finite conjunctions of axioms from T_1 and T_2, respectively. Suppose $\Box((K_1\&K_2) \supset A)$. Then $\Box(K_1 \supset (K_2 \supset A))$. M-Rep($T_1 : N_1$) implies that, for a conjunction H_1 of axioms of N_1 and a conjunction SS of axioms of S_{L_0}, $\Box((SS\&H_1) \supset K_1)$; then $\Box((SS\&H_1) \supset (K_2 \supset A))$; so $\Box(K_2 \supset ((SS\&H_1) \supset A))$. M-Rep($T_2 : N_2$) implies that, for a conjunction H_2 of axioms of N_2 and a conjunction SS' of axioms of S_{L_0}, $\Box((SS'\&H_2) \supset K_2)$; so $\Box((SS'\&H_2) \supset ((SS\&H_1) \supset A))$. Thus, $\Box((SS\&SS')\&(H_1\&H_2)) \supset A)$. Therefore, $\Box((H_1\&H_2) \supset A)$ (by (C$)), where $(H_1\&H_2)$ is a conjunction of axioms of N_1+N_2.

[16] See Benacerraf (1973) and Field (1989, 18–30) for arguments of this sort.

[17] See Field's discussion in (1989, pp. 30–38 and Ch. 3). Field claims that the logical consequence relation is neither a proof-theoretic nor a semantic notion, but rather is a primitive logical notion. By calling it a *primitive* Field means that logical consequence does not get its meaning from a definition. I agree, but I still consider logical consequence to be a *semantic* notion (like truth and reference). Set-theoretic "semantics" only emulates genuine semantic concepts.

[18] I am greatful to Chris Swoyer, Hartry Field, and an anonymous referee for insightful comments and suggestions.

REFERENCES

Benacerraf, P. 1973: 'Mathematical Truth'. *J. Philos.* **19**, 661–679.

Burgess, J. 1984: 'Synthetic Mechanics'. *J. Philos. Logic* **13**, 379–395.

Boolos, G. and R. Jeffrey 1989: *Computability and Logic*, 3rd edition. Cambridge: Cambridge Univ. Press.

Boyd, R. 1973: 'Realism, Underdetermination, and a Causal Theory of Evidence'. *Noûs* **7**, 1–12.

Field, H. 1980: *Science Without Numbers: A Defense of Nominalism.* Oxford: Basil Blackwell; Princeton: Princeton Univ. Press.

Field, H. 1989: *Realism, Mathematics and Modality.* Oxford: Basil Blackwell.

Field, H. 1992: 'A Nominalistic Proof of the Conservativeness of Set Theory'. *J. Philos. Logic* **21**, 111–123.

Hellman, G. 1989: *Mathematics Without Numbers.* Oxford: Oxford Univ. Press.

Montague, R. 1965: 'Set Theory and Higher Order Logic'. In *Formal Systems and Recursive Functions*, eds J. N. Crossley and M. Dummett. Amsterdam: North-Holland, 131–148.

Putnam, H. 1973: 'Explanation and Reference'. In *Conceptual Change*, eds G. Pearce and P. Maynard. Dordrecht: Reidel, 199–221. Reprinted as Chapter 2 of H. Putnam 1975: *Mind, Language and Reality: Philosophical Papers*, vol. 2. Cambridge: Cambridge Univ. Press.

Shapiro, S. 1983: 'Conservativeness and Incompleteness'. *J. Philos.* **80**, 521–531.

Shapiro, S. 1993: 'Modality and Ontology'. *Mind* **102**, 455–481.

Suppes, P. 1960: *Axiomatic Set Theory*. Princeton: van Nostrand.

van Fraassen, B. 1975: *The Scientific Image*. Oxford: Clarendon Press.

Wang, H. 1964: *A Survey of Mathematical Logic*. Amsterdam: North-Holland.

Department of Philosophy,
University of Oklahoma,
Norman, OK 73019-0535,
U.S.A.