The Bulletin of Symbolic Logic
Volume 16, Number 1, March 2010

STRONG LOGICS OF FIRST AND SECOND ORDER

PETER KOELLNER

Abstract. In this paper we investigate strong logics of first and second order that have certain absoluteness properties. We begin with an investigation of first order logic and the strong logics ω-logic and β-logic, isolating two facets of absoluteness, namely, generic invariance and faithfulness. It turns out that absoluteness is relative in the sense that stronger background assumptions secure greater degrees of absoluteness. Our aim is to investigate the hierarchies of strong logics of first and second order that are generically invariant and faithful against the backdrop of the strongest large cardinal hypotheses. We show that there is a close correspondence between the two hierarchies and we characterize the strongest logic in each hierarchy. On the first-order side, this leads to a new presentation of Woodin's Ω-logic. On the second-order side, we compare the strongest logic with full second-order logic and argue that the comparison lends support to Quine's claim that second-order logic is really set theory in sheep's clothing.

This paper is concerned with strong logics of first and second order. At the most abstract level, a strong logic of first-order has the following general form: Let L be a first-order language and let $\Phi(x)$ be a formula that defines a class of L-structures. Then, for a recursively enumerable set T of sentences of L, and for a sentence φ of L set

$$T \models_{\Phi(x)} \varphi$$

iff for all L-structures M such that $\Phi(M)$,

$$\text{if } M \models T \text{ then } M \models \varphi.$$

To strengthen the logic one narrows the class of *test structures* M that are consulted. For example, if $\Phi(x)$ places no constraints whatsoever on the test structures, then the result is first-order logic; if $\Phi(x)$ holds only of models that correctly compute the natural numbers, then the result is ω-logic; if Φ holds only of models that are well-founded then, then the result is β-logic, etc.

In the second-order context there is an additional degree of variability, namely, the scope of the second-order domain. Let L' be a second-order

Received November 2, 2007.
I would like to thank the referee and Ilijas Farah for helpful comments on an earlier draft of this paper.

1079-8986/10/1601-0001/$4.60

extension of L and let $\Phi(x, y)$ be a formula that defines a class of L'-structures. Then, for a recursively enumerable set T of sentences in L' and for a sentence φ in L', set

$$T \models_{\Phi(x,y)} \varphi$$

iff for all L'-structures $\langle M, S\rangle$ such that $\Phi(M, S)$,

$$\text{if } \langle M, S\rangle \models T \text{ then} \langle M, S\rangle \models \varphi.$$

To strengthen the logic one narrows the class of test structures $\langle M, S\rangle$ that are consulted, only now there are two dimensions—one can restrict the first-order domain M and one can restrict the second-order domain S. The first restriction parallels the first-order case. But even if one allows all possible first-order domains, one can obtain a wide array of logics through variation along the second dimension. For example, at one extreme one can place no constraints on the second-order domain (apart from the minimal necessary constraint that $\langle M, S\rangle$ be such that S is closed under definability with parameters) and at the other extreme one can demand that $\langle M, S\rangle$ be such that S is the full powerset of M. The first extreme leads to Henkin's interpretation of second-order logic and the other extreme leads to full second-order logic.

It would be tendentious to maintain that every formula $\Phi(x)$ in the first-order case and every formula $\Phi(x, y)$ in the second-order case defines a logic. Most would agree that in the first-order case, when the formula $\Phi(x)$ places no constraint on the test structures, one obtains a genuine logic—first-order logic. And many would agree that in the second-order case, when the formula $\Phi(x, y)$ places no constraints on the test structures (apart from the minimal necessary constraint in the second-order setting), one obtains genuine logic—the Henkin version of second-order logic. However, doubts as to logicality arise as one places constraints on either the first-order part or the second-order part of the class of test structures consulted. For example, in the limiting case along the first dimension, the formula $\Phi(x)$ (or $\Phi(x, y)$) holds of a single test structure M (or of pairs $\langle M, S\rangle$ where M is fixed) and the resulting "logic" would simply be *truth in the model M*. And, in the limiting case along the second dimension, the formula $\Phi(x, y)$ singles out the powerset of M and many have thought that the mathematical entanglement[1] with the powerset relation is too significant to warrant classifying the "logic"—full second-order logic—as a genuine logic. The question motivating this work is: At what point in the restriction on the class of test structures does one clearly pass from something that can be called a logic to something that cannot?

I shall not be so much concerned with the question "What is logic?" This question is exceedingly difficult since logic has many facets. Instead I want to investigate an important feature of first-order logic and see how far one can

[1]The metaphor of "mathematical entanglement" is due to Charles Parsons.

go in establishing logics that share this feature. The key feature of first-order logic that I shall concentrate on is *absoluteness*. First-order logic has a high degree of absoluteness that enables a wide array of disputants to agree on what implies what and for this reason it can serve as a neutral framework in which a wide array of disputants can discuss their differences. The key point for the present discussion is that *absoluteness is relative*. In the case of first-order logic, the absoluteness of the logic is secured relative to weak background assumptions. When one moves to stronger background assumptions greater degrees of absoluteness become available and this enables one to set up strong logics that share the desired feature of absoluteness, only now absoluteness is secured relative to the stronger background assumptions. Provided disputants agree on these background assumptions they will agree on what implies what in the strong logic. For example, relative to ZFC, the logics ω-logic and β-logic are absolute, and relative to extensions of ZFC by large cardinal axioms much stronger logics become absolute. Our aim is to characterize the strongest logics of first and second order that are absolute (in a sense we shall make precise) relative to strong background assumptions (large cardinal axioms).[2]

The first task is to render precise the relevant notion of "absoluteness". This will be the main purpose of Sections 1 and 2. In Section 1 we shall investigate the many facets of the absoluteness of first-order logic. In Section 2 we shall start by investigating two traditional strong logics (ω-logic and β-logic) that share many of these features of absoluteness, only now absoluteness is secured relative to ZFC. These logics will serve as our guide in setting up stronger logics that are absolute relative to stronger background assumptions. We shall isolate an important facet of absoluteness—generic invariance—and then investigate a hierarchy of strong logics due to Woodin

[2]I do not wish to defend the positive claim that the strong logics in the two hierarchies I shall investigate are logics in *exactly* the same sense in which first-order logic is logic. For example, the hierarchy of strong logics of first-order that I consider starts with ω-logic and if this were to be regarded as having the same logical status as first-order logic then it would obscure certain foundational matters. For example, consider the approach that Carnap takes in [7] with regard his Language I. He starts with a notion of "consequence" that has two components—first, it incorporates the axioms of a Σ^0_1-complete system of arithmetic (in fact, a version of PRA), second, it involves the ω-rule. He then defines a sentence to be *analytic* if it is a consequence of the null set and to be *contradictory* if every sentence is a consequence of it. It is then straightforward to show that every sentence (in the mathematical fragment of Language I) is either analytic or contradictory. The truths of (the mathematical fragment of) Language I are thus analytic. According to Carnap, "by means of the concept 'analytic', an exact understanding of what is usually designated as 'logically valid' or 'true on logical grounds' is achieved" (p. 41). However, this route to securing the logicist thesis with regard to arithmetic is clearly compromised by the involvement of ω-logic. In short, ω-logic is entangled with mathematics to a degree that makes it unsuitable for the logicist project with respect to arithmetic. Nevertheless, ω-logic shares some of the features of first-order logic and it is useful to concentrate on these features and see how far the generalize.

that generalize ω-logic and β-logic and share this facet of absoluteness relative to strong background assumptions (large cardinal axioms). In Section 3 we shall isolate a second feature of absoluteness—faithfulness. Our aim in the remainder of the paper will be to characterize the strongest logics of first and second order that have these two facets of absoluteness relative to large cardinal axioms. In the remainder of Section 3 we shall do this for the first-order case and in Section 4 we shall do this for the second-order case. In the end this will lead (in the first-order setting) to a new motivation for Woodin's Ω-logic and (in the second-order setting) to a new perspective on full second-order logic.

The mathematical results that follow are quite elementary and I suspect that most of the points I shall make are familiar to those who have investigated the matter. I see this paper as a continuation of the discussion in [33], with which I am in agreement. My hope is that a comparison of full second-order logic with the hierarchy of strong logics considered in this paper will further reveal the problematic nature of full second-order logic.

§1. First-order logic and absoluteness.

1.1. Basic notions. We begin by recalling some basic features of first-order logic. Let L be a first-order language. The relation of *logical implication* is defined as follows: For T a recursively enumerable[3] set of sentences in L and φ a sentence in L,

$$T \models \varphi \text{ iff for every } L\text{-structure } M, \text{ if } M \models T \text{ then } M \models \varphi.$$

This relation is Π_1-definable over the universe of sets and so there is the potential for a high degree of mathematical entanglement. However, the Completeness Theorem provides us with a significant reduction. Recall that the relation of *logical provability* is defined as follows: For T a recursively enumerable set of sentences in L and φ a sentence in L,

$$T \vdash \varphi \text{ iff there is a proof of } \varphi \text{ from } T \text{ in system } S,$$

where S is one of many possible deduction systems that is adequate for first-order logic. The key feature of S is that it yields the Completeness Theorem: For T a recursively enumerable set of sentences in L and for φ a sentence in L,

$$T \models \varphi \text{ iff } T \vdash \varphi.$$

[3]In this paper we will generally assume that our theories T are recursively enumerable. However, much of what we say continues to hold if this assumption is relaxed. In particular, in the present definition and the definitions of the stronger logics that follow, the restriction to T that are recursively enumerable can be dropped.

For T that are recursively enumerable, the relation $T \vdash \varphi$ is Σ^0_1-definable. This provides us with a reduction of a notion which is apparently Π_1-definable over the universe of sets to one that is Σ^0_1-definable over the structure of the natural numbers.

This reduction is best possible. To explain this well-known result we need to introduce some basic notions: For sets of integers X and Y, let us write $Y \leqslant_T X$ to indicate that Y is Turing reducible to X. For a given class of formulas Γ, a set of integers X is a *complete* Γ *set of integers* if (i) X is definable by a formula in Γ and (ii) for every set of integers Y, if Y is definable by a formula in Γ, then $Y \leqslant_T X$. For a formula φ in L, let $\ulcorner \varphi \urcorner$ be the Gödel number of φ according to some fixed encoding. Finally, let

$$\mathcal{V}_{\mathrm{F}} = \{\ulcorner \varphi \urcorner \mid \varnothing \models \varphi\}.$$

This is simply the set of (Gödel numbers of) validities of first-order logic.

LEMMA 1.1. *$\mathcal{V}_{\mathrm{F}}$ is a complete Σ^0_1 set of integers.*

PROOF. We have already seen that $\mathcal{V}_{\mathrm{F}}$ is Σ^0_1-definable (by the Completeness Theorem). It remains to show that if X is a Σ^0_1-definable set of integers then $X \leqslant_T \mathcal{V}_{\mathrm{F}}$. Let $X \subseteq \omega$ be Σ^0_1-definable and let φ be the defining formula. The key point is that there is a formal system of arithmetic that is finitely axiomatizable, Σ^0_1-complete, and Σ^0_1-sound, namely, Robinson's system Q (see [6], p. 82). It follows that for each $n \in \omega$,

$$n \in X \text{ iff } \text{“}\mathrm{Q} \rightarrow \varphi(\bar{n})\text{”} \in \mathcal{V}_{\mathrm{F}},$$

where $\bar{n}$ is the canonical numeral designating n. This provides us with a Turing reduction $X \leqslant_T \mathcal{V}_{\mathrm{F}}$. ⊣

This shows that the set of logical validities is mathematically entangled to a minimal degree. However, as we shall see, the mathematical entanglement involved is innocuous since Σ^0_1 is highly *absolute*. To describe the various facets of absoluteness it will be useful to investigate the space of mathematical theories in an abstract setting.

1.2. Interpretability. We shall investigate the space of mathematical theories (recursively enumerable axioms systems) under the relation of *interpretability*. The informal notion of interpretability is ubiquitous in mathematics; for example, Poincaré provided an interpretation of two dimensional hyperbolic geometry in the Euclidean geometry of the unit circle; Dedekind provided an interpretation of analysis in set theory; and Gödel provided an interpretation of the theory of formal syntax in arithmetic.

We will employ the following formalization of this notion: Let T_1 and T_2 be recursively enumerable axiom systems. The system T_1 is *interpretable* in the system T_2 if, roughly speaking, the primitive concepts and the range of quantification of T_1 are definable in T_2 in such a way that for every theorem of T_1 the result of translating the theorem via the definitions is a theorem of T_2.[4] We shall write $T_1 \leqslant T_2$ to indicate that T_1 is interpretable in T_2, $T_1 < T_2$ to indicate that T_1 is interpretable in T_2 but not conversely, and $T_1 \equiv T_2$ to indicate that T_1 is interpretable in T_2 and conversely. In the last case, T_1 and T_2 are said to be *mutually interpretable*. The *interpretability hierarchy* is the collection of all theories (satisfying our three simplifying assumptions) ordered under the relation $\leqslant$. We now turn to a discussion of the structure of this hierarchy.

To begin with, there is a useful characterization of the relation $\leqslant$. To describe this we need a few more notions. A theory T is *reflexive* provided it proves the consistency of each of its finite fragments, it is Σ^0_1-*complete* provided it proves each true Σ^0_1-statement, and it is Σ^0_1-*sound* provided it does not prove a false Σ^0_1-statement. The following characterization of interpretability is fundamental: If T_1 is reflexive and Σ^0_1-complete, then

$$T_1 \leqslant T_2 \text{ iff } T_1 \subseteq_{\Pi^0_1} T_2,$$

where the right-hand side is shorthand for the statement that all Π^0_1-sentences provable in T_1 are provable in T_2.[5] It follows from this characterization and the second incompleteness theorem that for any theory T the theory $T + \mathrm{Con}(T)$ is strictly stronger than T, that is, $T < T + \mathrm{Con}(T)$. Moreover, it follows from the arithmetized completeness theorem that the theory $T + \neg\mathrm{Con}(T)$ is interpretable in T, hence, $T \equiv T + \neg\mathrm{Con}(T)$.[6]

The interpretability hierarchy is quite complex. For example, for a given theory T, using coding techniques it is possible to construct two theories T_1 and T_2 such that $T < T_1$ and $T < T_2$ and yet T_1 and T_2 are incomparable (that is, neither $T_1 \leqslant T_2$ nor $T_2 \leq T_1$). And, for any two theories T_1 and T_2 such that $T_1 < T_2$, using coding techniques it is possible to construct a third theory T such that $T_1 < T < T_2$. Thus, the interpretability hierarchy is neither linear nor well-founded.[7] The examples used to demonstrate the failure of linearity and well-foundedness are quite contrived and artificial. Remarkably, when one concentrates on those theories that "arise in nature" (in the normal course of mathematics) the theories are well ordered under interpretability.

We noted above that the second incompleteness theorem provides us with a case where the statement, namely, $\mathrm{Con}(T)$, leads to a jump in the interpretability hierarchy while its negation does not. There are also natural examples of this phenomenon (for example, large cardinal axioms). We shall

[4]For a more precise account of this and the other notions of interpretability theory, see [8], [20], and [34]. In the set theoretic discussion below I will be allowing Boolean-valued interpretations.

[5]See Theorem 6 on p. 103 of [20] and Fact 3.2 of [34].

[6]See [8].

[7]See [20] and [8].

refer to the kind of independence involved in the second incompleteness theorem as *vertical* independence. In contrast to vertical independence there is the kind of independence involved when one shows that a sentence φ is independent of a theory T by showing that

$$T \equiv T + \varphi \text{ and } T \equiv T + \neg\varphi.$$

Such a sentence φ is called an *Orey sentence* with respect to T. We shall refer to this kind of independence as *horizontal* independence. Using coding techniques one can construct such sentences. But again, just as in the case of vertical independence, there are also natural non-metamathematical examples (for example, CH is an Orey sentence with respect to ZFC).

1.3. Facets of absoluteness. With these notions in place we can now describe some of the facets of the absoluteness of first-order logic. The source of the absoluteness of first-order logic lies in the fact that the consequence relation is Σ^0_1 and almost all theories (even very weak ones) are Σ^0_1-complete. We shall spell this out in two stages, first assuming that the theories under consideration are Σ^0_1-sound and then relaxing this assumption, assuming only that the theories under consideration are consistent.

Let T_1 and T_2 be axiom systems that interpret PA.[8] Assume that both T_1 and T_2 are Σ^0_1-sound. The assumption that T_1 and T_2 interpret PA ensures that both theories are Σ^0_1-complete and that

$$T_1 \leqslant T_2 \text{ iff } T_1 \subseteq_{\Pi^0_1} T_2.$$

The assumption that T_1 and T_2 are Σ^0_1-sound ensures, in addition, that

$$T_1 =_{\Sigma^0_1} T_2.$$

There are two cases to consider: $T_1 \equiv T_2$ and $T_1 \not\equiv T_2$. In the case where $T_1 \equiv T_2$, it follows that

$$T_1 \vdash \text{“}T \vdash \varphi\text{” iff } T_2 \vdash \text{“}T \vdash \varphi\text{”}$$

and

$$T_1 \vdash \text{“}T \nvdash \varphi\text{” iff } T_2 \vdash \text{“}T \nvdash \varphi\text{”},$$

where here (and below) T is a recursively enumerable theory and φ is a sentence. Thus, in this case, T_1 and T_2 are in full agreement on both validity and non-validity. It the case where $T_1 \not\equiv T_2$ there are three subcases: Either $T_1 < T_2$, $T_2 < T_1$, or both $T_1 \nleqslant T_2$ and $T_2 \nleqslant T_1$. In either subcase it follows (from Σ^0_1-soundness) that

$$T_1 \vdash \text{“}T \vdash \varphi\text{” iff } T_2 \vdash \text{“}T \vdash \varphi\text{”}$$

[8]The choice of PA is merely one of convenience. Everything that we shall say applies under much weaker assumptions. For example, it suffices to assume only that both T_1 and T_2 interpret $I\Delta_0 + \text{Exp}$. See footnote 10 for further discussion.

and so there is agreement on validity. It also follows (by Σ^0_1-soundness) that there can be no disagreement concerning non-validity. However, it is possible that one theory sees instances of non-validity that the other theory does not. More precisely, in the subcase $T_1 < T_2$, T_2 will see instances of non-validity that T_1 does not see. Likewise, in the subcase $T_2 < T_1$, T_1 will see instances of non-validity that T_2 does not see, and in the subcase where $T_1 \not\leq T_2$ and $T_2 \not\leq T_1$, each theory will see instances of non-validity that the other theory does not see. But the important point is that although one theory sees instances of non-validity not seen by the other theory, there is no *disagreement*.[9]

Let us now weaken our background assumption by assuming only that T_1 and T_2 are consistent and not necessarily Σ^0_1-sound. In the case where $T_1 \equiv T_2$, we have agreement on non-validity (since $T_1 =_{\Pi^0_1} T_2$) but one theory, say T_1, may prove "$T \vdash \varphi$" where this is a false Σ^0_1-sentence. However, even in this case there will not be disagreement (since if T_2 were to prove the negation of this then so would T_1, in which case T_1 would be inconsistent). In the case where $T_1 \not\equiv T_2$, there can be disagreement. To take a trivial example, letting $T_1 = \mathrm{PA} + \neg\mathrm{Con}(\mathrm{PA})$ and $T_2 = \mathrm{PA} + \mathrm{Con}(\mathrm{PA})$, we have that

$$T_1 \vdash \text{“PA} \vdash 0 = 1\text{” and } T_2 \vdash \text{“PA} \nvdash 0 = 1\text{”}$$

and so there is a disagreement on validity (and hence on non-validity). In this case either T_1 is not Σ^0_1-sound (if $\mathrm{Con}(\mathrm{PA})$) or T_2 is inconsistent (if $\neg\mathrm{Con}(\mathrm{PA})$). Thus, if we merely assume that the theories are consistent, then disagreements on logic can arise in the case where $T_1 \not\equiv T_2$.

The above example does not pose a serious problem in terms of a reasonable foundational disagreements. The reason is that T_1 hardly counts as a reasonable theory since if one is committed to PA then one is committed to $\mathrm{Con}(\mathrm{PA})$. The question arises: Are there serious cases where disagreements arise on first-order logic? As a first pass, consider $T_1 = \mathrm{ZFC} + \neg\mathrm{Con}(\mathrm{ZFC}+$ "There is a measurable cardinal") and $T_2 = \mathrm{ZFC} +$ "There is a strong cardinal". For our present purposes, all that one needs to know about these theories is that T_2 proves $\mathrm{Con}(\mathrm{ZFC} +$ "There is a measurable cardinal") and hence the two theories disagree on logic. Moreover, in this case T_1 cannot be ruled out as unreasonable on the previous grounds since no "internally governed" iteration of consistency statements will reach $\mathrm{Con}(\mathrm{ZFC} +$ "There is a measurable cardinal"). Indeed there is at least one prominant set theoriest who believes T_1. Now, although there are people who *believe* T_1, there

[9]In this discussion we have not assumed that T_1 and T_2 involve enough set theoretic machinery to speak of the model-theoretic formulation of validity and non-validity, and so each theory has had to access the logical notions through their syntactic equivalents. In the case where T_1 and T_2 have the requisite model-theoretic resources, then, in the above discussion, "$T \vdash \varphi$" can be replaced by "$T \models \varphi$".

are not people who advocate T_1 as a foundational viewpoint, the reason being that if $\neg\mathrm{Con}(\mathrm{ZFC} + \text{“There is a measurable cardinal”})$ is true, then it is provable and so adding it as an axiom would be redundant; rather, it is something that one would strive to prove. As a second pass, consider $T_1 = \mathrm{ZFC} + V = L$ and $T_2 = \mathrm{ZFC} + \text{“There is a measurable cardinal”}$. For our present purposes, all that one needs to know about these theories is that T_1 proves “There does not exist a measurable cardinal” and so the two theories disagree. But although the two theories disagree, they do not disagree on first-order logic. In fact, in this case, one has much more than that $T_1 \subseteq_{\Pi^0_1} T_2$, one has that $T_1 \subseteq_{B(\Sigma^1_2)} T_2$, where $B(\Sigma^1_2)$ is the collection of all Boolean combinations of Σ^1_2-sentences. This example is representative of the general situation—when we restrict our attention to the “natural” theories that occur in foundational disputes, there is no disagreement on first-order logic.

To summarize: In the case where we assume Σ^0_1-soundness we have that if $T_1 \equiv T_2$ then there is full agreement on logic and if $T_1 \not\equiv T_2$ then although one theory may see non-validities not seen by the other theory, there is no *disagreement* on logic. In the case where we merely assume that our theories are consistent, if $T_1 \equiv T_2$ then there is no disagreement on logic and if $T_1 \not\equiv T_2$ then there can be disagreement on logic but not in the case where T_1 and T_2 are “natural” theories such as those that occur in foundational debates. So we have been able to give a perfectly precise characterization of absoluteness across theories that are *horizontally displaced* $(T_1 \equiv T_2)$—there is no disagreement—but in the case of theories that are *not horizontally displaced* $(T_1 \not\equiv T_2)$ we have had to rely on the vague notion of a “natural” theory—there is no disagreement provided the theories are “natural”.[10]

§2. Some strong logics. In the previous section our base theory was a weak system of arithmetic (or, equivalently, a weak system of set theory). In this section we shall take as our base theory the much stronger theory ZFC

[10]In the above discussion we have focused on theories that interpret PA. It is important to note that the discussion carries through with respect to much weaker assumptions. The source of the absoluteness of first-order logic lies in Σ^0_1-completeness and this holds for theories much weaker than PA. In fact, Q is Σ^0_1-complete. Unfortunately, Q cannot prove this. In fact, Q is too weak to encode syntax. The most natural setting in which to encode syntax is $I\Delta_0 + \mathrm{Exp}$ and this theory is sufficiently strong to recognize the Σ^0_1-completeness of Q and hence of itself. There are even weaker theories in which one can encode syntax. For example, this is possible in Buss's S^1_2, although in this very weak setting the encoding of syntax is a quite subtle matter. The theory S^1_2 is also capable of proving Σ^0_1-completeness in the form appropriate to this setting. See Theorem 7.4 of [4] and Lemma 1.2.3 of [11]. Thus, it appears that as soon as one is able to make sense of the syntactic notions one has the resources to secure the central mathematical ingredient underlying the absoluteness of first-order logic.

and we shall supplement it with strong axioms (large cardinal axioms). The theories we shall consider have the feature that they enable one to establish progressively stronger logics which share many of the facets of absoluteness held by first-order logic.

2.1. Interpretability in set theory. There is a canonical way of climbing the hierarchy of interpretability into the higher reaches. It involves using theories formulated in the language of set theory, that is, the language of first-order logic with a binary relation, $\in$, for membership. A very weak such theory is the theory WS consisting of the following two axioms:

$$\exists x\,\forall y\,(y \notin x),$$
$$\forall x\,\forall y\,\exists z\,\forall u\,(u \in z \leftrightarrow u = x \vee u = y).$$

The theory is mutually interpretable with Q.[11] Strength is obtained by adding stronger closure principles and this leads to a hierarchy of set theoretic systems. A natural point in the hierarchy of systems is ZF − Infinity, which is mutually interpretable with PA. Stronger theories are then obtained by adding *large cardinal axioms*, such as the axioms asserting "there is an infinite cardinal", "there is a strongly inaccessible cardinal", "there is a Mahlo cardinal", "there is a measurable cardinal", "there is a strong cardinal", "there is a Woodin cardinal", "there is a supercompact cardinal", "there is a huge cardinal", "there is a rank to rank embedding", and so on.

Three remarkable facts emerge from the study of such theories. First, like all natural theories, large cardinal axioms are well-ordered under interpretability. More importantly, in the case of a large class of large cardinal axioms (those formulated in terms of elementary embeddings) the ordering is transparent. For example, the reason ZFC + "There is a strong cardinal" is stronger than ZFC + "There is a measurable cardinal" is that there are many measurable cardinals below each strong cardinal.

Second, the theories formulated in terms of large cardinals play a central role in the comparison of theories under interpretability. To compare two sufficiently strong theories T_1 and T_2 in terms of interpretability power one typically finds large cardinal axioms φ_1 and φ_2 and then shows that T_1 is mutually interpretable with ZFC + φ_1 and T_2 is mutually interpretable with ZFC + φ_2. The ordering of T_1 and T_2 can then be read off of the natural ordering on φ_1 and φ_2. Remarkably, in many cases this is the *only known way* to compare such theories—one apparently must pass through the large cardinal hierarchy. Thus, without loss of generality, we may restrict our attention to the language of set theory and consider theories of the form ZFC + φ where φ is a large cardinal axiom. These theories have a natural

[11] For an interpretation of WS (plus the axiom of extensionality) in Q see [26]. An interpretation of Q in WS is given in [25], building on [24], which in turn builds on [32]. See [35] for further results in this direction.

well-ordering and serve as a yardstick along which one can measure the "strength" of a theory in the hierarchy of interpretability.[12]

Third, the method of comparing theories in set theory has an important consequence concerning absoluteness. Let us first recall this method. As we noted above, one compares T_1 and T_2 by comparing each with a theory formulated in terms of large cardinals. Suppose then that we wish to show that T is mutually interpretable with $\mathrm{ZFC} + \varphi$ where φ is a large cardinal axiom. To show that $T \leqslant \mathrm{ZFC} + \varphi$ one starts with a model $\mathrm{ZFC} + \varphi$ and uses the *method of forcing* to produce a model of T. There are a number of ways of viewing the method of forcing. We shall focus on two. According to the first method, one starts with a countable transitive model M (sometimes (perhaps confusingly) also denoted 'V') of $\mathrm{ZFC} + \varphi$ and produces a model $M[G]$ satisfying $\mathrm{ZFC} + \varphi$, where G is an "M-generic" subset of a partial order $\mathbb{P}$ in M. According to the second method, one starts with V (a class size structure) and constructs a Boolean-valued model $V^{\mathbb{B}}$, where $\mathbb{B}$ is a complete Boolean algebra. What is important for our purposes is that the method is such that $T \leqslant \mathrm{ZFC} + \varphi$. To show that $\mathrm{ZFC} + \varphi \leqslant T$ one starts with a model of T and uses the *method of inner models* to build a model of $\mathrm{ZFC} + \varphi$. The important point for our purposes is that these two methods have as a consequence (via Shoenfield's absoluteness theorem[13]) that

$$T =_{B(\Sigma^1_2)} \mathrm{ZFC} + \varphi,$$

where recall that $B(\Sigma^1_2)$ is the collection of Boolean combinations of Σ^1_2 statements. More generally, if T_1 and T_2 are natural theories extending ZFC, and T_1 and T_2 are compared by the above methods, then

$$T_1 \leqslant T_2 \text{ iff } T_1 \subseteq_{B(\Sigma^1_2)} T_2.$$

This jump from Π^0_1- to $B(\Sigma^1_2)$-absoluteness opens up the possibility of setting up logics which are much stronger than first-order logic and which still have the desired features of absoluteness. We shall now consider two such logics.

2.2. ω-Logic. We begin with some definitions. Let $\mathrm{ZF_N}$ be the first N axioms of ZF under some canonical order. A model (M, E) satisfying $\mathrm{ZF_N}$ for some sufficiently large N is an *ω-model* if the natural numbers as computed in the model are (up to isomorphism) the true natural numbers, that is, $(\omega^M, E{\restriction}\omega^M) \cong (\omega, \in{\restriction}\omega)$. When we speak of an ω-model M it is to

[12]I should stress that large cardinal axioms are not the *only* way to climb the hierarchy of interpretability. In fact, for each theory of the form $\mathrm{ZFC} + \varphi$, where φ is a large cardinal axiom, the theories $\mathrm{PA} + \cup_{n<\omega}\mathrm{Con}_n(\mathrm{ZFC} + \varphi)$ and $\mathrm{ZFC} + \varphi$ are mutually interpretable, where $\mathrm{Con}_n(T)$ is the statement that no contradiction can be derived from the first n axioms of T (relative to some fixed enumeration). Thus, one could also use theories of the latter form to climb the hierarchy. Such theories are hardly natural and they are clearly derived from large cardinal axioms. I am only claiming that the most *natural* way to climb the hierarchy of interpretability is through large cardinal axioms.

[13]See [31] and [12], §13.

be understood that $M \models \mathrm{ZF_N}$ for some sufficiently large N, which can be fixed from the outset.

DEFINITION 2.1. Assume ZFC. Suppose T is a recursively enumerable set of axioms and φ is a sentence in the language of set theory. Then

$$T \models_\omega \varphi$$

iff for all ω-models M of $\mathrm{ZF_N}$,

$$\text{if } M \models T \text{ then } M \models \varphi.$$

This logic is called *ω-logic*. The relation $T \models_\omega \varphi$ is Π_1-definable over the universe of sets. However, the Löwenheim-Skolem theorem provides a significant reduction, since it shows that the result of restricting to ω-models which are countable yields an equivalent definition. Let

$$\mathcal{V}_\omega = \{\ulcorner \varphi \urcorner \mid \varnothing \models_\omega \varphi\}.$$

LEMMA 2.2. *Assume* ZFC. *Then $\mathcal{V}_\omega$ is a Π^1_1-complete set of integers.*

PROOF. In defining $T \models_\omega \varphi$ we may, by the Löwenheim-Skolem theorem, restrict to ω-models that are countable. It follows that $\mathcal{V}_\omega$ is Π^1_1-definable. It remains to show that if X is a Π^1_1-definable set of integers then $X \leqslant_T \mathcal{V}_\omega$.

The key point is that ω-models are correct about arithmetical statements and hence that Π^1_1-statements are downward absolute to ω-models. Thus, if φ is a true Π^1_1-statement then $\varnothing \models_\omega \varphi$. On the other hand, if φ is a Π^1_1-statement and $\varnothing \models_\omega \varphi$ then φ must be true, since, by reflection, letting γ be such that $V_\gamma \models \mathrm{ZF_N}$, we have that $V_\gamma \models \varphi$ and hence that φ. It follows that if $X \subseteq \omega$ is defined by the Π^1_1-formula $\varphi(x)$, then

$$n \in X \text{ iff } \varphi(\bar{n}) \in \mathcal{V}_\omega,$$

for each $n < \omega$, where $\bar{n}$ is a formal numeral designating n. This provides a Turing reduction of X to $\mathcal{V}_\omega$. ⊣

2.3. β-Logic. A model (M, E) is an *β-model* if the relation E is well-founded. Each such model is isomorphic to a transitive set (by a theorem of Mostowski) and we will routinely identify β-models with their isomorphic transitive copies.

DEFINITION 2.3. Assume ZFC. Suppose T a recursively enumerable set of axioms and φ a sentence in the language of set theory. Then

$$T \models_\beta \varphi$$

iff for all β-models M,

$$\text{if } M \models T \text{ then } M \models \varphi.$$

This logic is called *β-logic*. The relation $T \models_\beta \varphi$ is Π_1-definable over the universe of sets. Again, the Löwenheim-Skolem theorem provides a significant reduction, since it shows that the result of restricting to β-models which are countable yields an equivalent definition. Since well-foundedness is Π^1_1, the relation $T \models_\omega \varphi$ is thus Π^1_2-definable. Let

$$\mathcal{V}_\beta = \{\ulcorner \varphi \urcorner \mid \varnothing \models_\beta \varphi\}.$$

Lemma 2.4. *Assume* ZFC. *Then* $\mathcal{V}_\beta$ *is a* Π^1_2*-complete set of integers.*

Proof. The key point is that β-models are correct about Π^1_1-statements and so Π^1_2-statements are downward absolute to β-models. The rest of the proof is similar to that of Lemma 2.2. ⊣

2.4. Absoluteness. It follows from the above discussion that if T_1 and T_2 are "natural" theories extending ZFC and if $T_1 \leqslant T_2$ is established set theoretically, then

$$\text{if } T_1 \vdash \text{“}T \models_\omega \varphi\text{” then } T_2 \vdash \text{“}T \models_\omega \varphi\text{”}$$

and

$$\text{if } T_1 \vdash \text{“}T \not\models_\omega \varphi\text{” then } T_2 \vdash \text{“}T \not\models_\omega \varphi\text{”}.$$

Likewise, under the same conditions, by Lemma 2.4,

$$\text{if } T_1 \vdash \text{“}T \models_\beta \varphi\text{” then } T_2 \vdash \text{“}T \models_\beta \varphi\text{”}$$

and

$$\text{if } T_1 \vdash \text{“}T \not\models_\beta \varphi\text{” then } T_2 \vdash \text{“}T \not\models_\beta \varphi\text{”}.$$

If $T_1 < T_2$ then all of these inclusions are proper. To take a simple case consider $T_1 = \mathrm{ZFC}$ and $T_2 = \mathrm{ZFC}+$"There is a strongly inaccessible cardinal". Then T_2 proves Con(ZFC) and it proves "ZFC − Infinity $\models_\omega$ Con(ZFC)" since it thinks that ω-models are arithmetically correct. However T_1 cannot prove this since it also thinks that ω-models are arithmetically correct and so, if it could prove this, then it would be able to prove Con(ZFC), contrary to the second incompleteness theorem. A similar argument shows that T_2 proves "ZFC − Infinity $\not\models_\omega \neg$Con(ZFC)" while T_1 cannot.

It turns out that when one moves to even stronger theories, greater degrees of absoluteness become available. As a consequence, as one climbs the hierarchy of "natural" theories the relation $T_1 \leqslant T_2$ holds if and only if $T_1 \subseteq_\Gamma T_2$ for progressively richer collections of formulas Γ. For example, when the theories T_1 and T_2 contain ZFC, one has that $T_1 \leqslant T_2$ if and only if $T_1 \subseteq_{B(\Sigma^1_2)} T_2$ and, at a certain further stage, one has that $T_1 \leqslant T_2$ if and only if $T_1 \subseteq_{B(\Sigma^1_3)} T_2$, and so on.

Recall that our goal is to investigate the extent to which one can establish strong logics that are "absolute" relative to large cardinal axioms. In the above discussion we have noted that as one climbs the hierarchy of interpretability along paths of "natural" theories, greater and greater degrees of absoluteness become available. However, this formulation involves the the vague notion of a "natural" theory. If we wish to prove results concerning the subject we shall have to remove this element of vagueness and isolate a precise version of absoluteness, one that is amenable to mathematical treatment. This is what we shall now do. But it is important to note that all of the results we prove continue to hold in the general setting provided one restricts to "natural" theories.

As we saw in Section 1.3 and above, the difficulty arises when one considers theories such that $T_1 \not\equiv T_2$—it is here that one has to restrict to "natural" theories in order to rule out the artificial counterexamples that arise from the second incompleteness theorem. In the case where the theories are horizontally displaced $(T_1 \equiv T_2)$ the situation is much clearer. The main technique in set theory for establishing horizontal independence is the method of forcing. It is this method that enables one to show that $\mathrm{ZFC} \equiv \mathrm{ZFC} + \mathrm{CH} \equiv \mathrm{ZFC} + \neg\mathrm{CH}$. This provides us with a broad and precise notion of absoluteness—generic invariance.

A logical relation $\models_\Phi$ is *generically invariant* relative to an extension $\mathrm{ZFC}^{(+)}$ of ZFC provided that for all recursively enumerable theories T and all sentences φ the following holds in $\mathrm{ZFC}^{(+)}$[14]: For all complete Boolean algebras $\mathbb{B}$

$$V \models \text{“}T \models_\Phi \varphi\text{”} \text{ iff } V^{\mathbb{B}} \models \text{“}T \models_\Phi \varphi\text{”}.$$

In other words, the logical consequence relation is not perturbed by passing to a generic extension. If we think of the models $V^{\mathbb{B}}$ as possible worlds, then this is tantamount to saying that the logical consequence relation is invariant across the possible worlds—all participants living in the various possible worlds agree on the logical consequence relation. This is the first of two facets of absoluteness that we shall isolate. The second will be isolated in Section 3.

All of the above logics are generically invariant relative to ZFC. We shall now investigate a hierarchy of logics due to Woodin that includes ω-logic and β-logic at the base and shares the feature of generic invariance, only now generic invariance is secured relative to much stronger background assumptions.

2.5. Generalization. We wish to generalize ω-logic and β-logic by placing stronger closure conditions on the (countable) test structures. In the case of ω-logic we restricted the test structures to those (countable) models that correctly computed ω. In the case of β-logic we restricted further to those (countable) models that are well-founded. Now, it might appear difficult to see how one could obtain a stronger logic by restricting the class of test structures even further, for we have already restricted to (countable) standard models. To be sure one can narrow further by demanding that these models satisfy a given theory—say, ZFC + "There is a supercompact cardinal"—but this will not lead to an essentially richer logic since the same

[14]In speaking of an *extension* we are including the degenerate case where $\mathrm{ZFC}^{(+)}$ is just ZFC.

effect can be achieved by looking at the β-consequences of the theory ZFC+ "There is a supercompact cardinal". What one should like is further closure conditions on the test structures themselves. Perhaps surprisingly there are such conditions that lead to essentially richer logics. To describe these closure conditions it will be useful to start by recasting the definition of β-logic in terms of a closure condition. This involves the notion of a *universally Baire* set of reals, which we now introduce.[15]

DEFINITION 2.5. Let λ be an ordinal. A *tree* on $\omega \times \lambda$ is a set $T \subseteq \omega^{<\omega} \times \lambda^{<\omega}$ such that for all pairs $(s, t) \in T$

(1) $\text{length}(s) = \text{length}(\tau)$ and
(2) for all $i < \text{length}(s)$, $(s{\restriction}i, t{\restriction}i) \in T$.

For T a tree on $\omega \times \lambda$, let

$$[T] = \{(x, f) \mid x \in \omega^\omega, f \in \lambda^\omega \text{ and } \forall k \in \omega\, (x{\restriction}k, f{\restriction}k) \in T\}$$

and let

$$p[T] = \{x \in \omega^\omega \mid (x, f) \in [T] \text{ for some } f \in \lambda^\omega\}.$$

Thus, $[T]$ is the set of "paths through T" and $p[T]$ is the "projection of $[T]$ onto the first coordinate".

DEFINITION 2.6. Let δ be a cardinal. A set $A \subseteq \omega^\omega$ is δ-*universally Baire* if for all partial orders $\mathbb{P}$ of cardinality δ there exist trees T and S on $\omega \times \lambda$, for some ordinal λ, such that

(1) $A = p[T]$ and
(2) if $G \subseteq \mathbb{P}$ is V-generic then in $V[G]$:

$$p[T] = \omega^\omega \smallsetminus p[S].$$

A set $A \subseteq \omega^\omega$ is *universally Baire* if it is δ-universally Baire for all cardinals δ. In other words, a set is universally Baire if for each partial order it admits a tree representation where the trees project to complements all generic extensions via the partial order. See [9] for further details.

The interest of the notion of a universally Baire set is that one is able to "track" such a set across generic extensions: For A a universally Baire and $G \subseteq \mathbb{P}$ a V-generic for some partial order $\mathbb{P}$, set

$$A_G = \bigcup \{p[T]^{V[G]} \mid T \in V \text{ and } A = p[T]^V\}.$$

We are now in a position to define the relevant notion of closure.

[15] This section is somewhat technical. The reader not familiar with the relevant notions is advised to overview it quickly on first reading to get the general picture and then proceed with the next section. See [17], [1], and the references therein for further discussion and proofs of the theorems of Woodin discussed below.

Definition 2.7. Suppose $A \subseteq \omega^\omega$ is universally Baire. Suppose M is an ω-model of $\mathrm{ZF_N}$. Then M is *A-closed* if for all partial orders $\mathbb{P} \in M$ and for all $G \subseteq \mathbb{P}$ which are V-generic, $V[G]$ satisfies

$$A_G \cap M[G] \in M[G].$$

It turns out that in ZFC one can establish that certain sets are universally Baire. The main result in this direction is Shoenfield's result on absoluteness, the proof of which shows that all Π^1_1-sets are universally Baire. Thus, we can consider the class of ω-models that are A-closed for all Π^1_1 sets A. The result of doing so leads to β-logic:

Theorem 2.8 (Woodin [1]). *Assume* ZFC. *Suppose* (M, E) *is an* ω*-model of* ZFC. *Then the following are equivalent*:

(1) (M, E) *is a* β*-model.*
(2) (M, E) *is* A*-closed for each* Π^1_1*-set* A.

It follows that β-logic (for extensions of ZFC) can be reformulated as follows: For T a recursively enumerable set of sentences extending ZFC and for φ a sentence, $T \models_\beta \varphi$ iff for all countable ω-models M, if M is Π^1_1-closed and $M \models T$, then $M \models \varphi$.

Thus, the notion of A-closure provides a generalization of well-foundedness and logics stronger than β-logic can now be obtained by restricting the class of test structures to those that are A-closed for all A in various richer pointclasses of universally Baire sets. Of course, this requires the existence of such pointclasses. Fortunately, strong background assumptions guarantee this:

Theorem 2.9 (Martin-Solovay [23]). *Assume that* $X^\sharp$ *exists for all sets* X. *Then all* Π^1_2*-sets are universally Baire.*

Moreover, some such strong background assumptions are necessary. In fact, if all Π^1_2-sets are universally Baire, then $X^\sharp$ exists for all sets X. See [9].

This theorem enables us to define the following strong logic:

Definition 2.10. Assume that $X^\sharp$ exists for all sets X. Suppose T is a recursively enumerable set of sentences and φ is a sentence. Then

$$T \models_{\Sigma^1_2} \varphi$$

iff for all countable transitive M,

$$\text{if } M \text{ is } \Sigma^1_2\text{-closed and } M \models T \text{ then } M \models \varphi.$$

With a bit of work this relation can be shown to be Π^1_3. The Martin-Solovay absoluteness theorem,[16] shows that if $X^\sharp$ exists for all sets X, then Π^1_3 statements are generically invariant. Thus, this logic is generically invariant under the background assumption used to define it.

[16] See [23] and [12], §14.

Stronger background assumptions ensure that even richer sets of reals are universally Baire and they secure the generic invariance required to set up stronger generically invariant logics. A critical stage is reached with the background theory ZFC + "There is a proper class of Woodin cardinals" since at this stage all sets of reals in $L(\mathbb{R})$ are universally Baire and the theory of such sets becomes generically invariant. We shall state this result in a general form.

THEOREM 2.11 (Woodin [17]). *Assume there is a proper class of Woodin cardinals. Then for each universally Baire set A the following hold*:

(1) *Every set of reals in $L(A, \mathbb{R})$ is universally Baire.*
(2) *If $\mathbb{P}$ is a partial order and $G \subseteq \mathbb{P}$ is V-generic, then*

$$(H(\omega_1)^V, \in, A) \prec (H(\omega_1)^{V[G]}, \in, A_G).$$

This puts us in a position to define a very strong generically invariant logic.[17]

DEFINITION 2.12 (Woodin [1]). Suppose T is a recursively enumerable theory and φ is a sentence. Then $T \vdash_\Omega \varphi$ if there exists a universally Baire set $A \subseteq \omega^\omega$ such that

(1) $L(A, \mathbb{R}) \models \mathrm{AD}^+$,
(2) each set in $\mathcal{P}(\mathbb{R}) \cap L(A, \mathbb{R})$ is universally Baire, and
(3) for all A-closed countable transitive M, for all $\alpha \in M \cap \mathrm{On}$,

$$\text{if } V_\alpha^M \models T \text{ then } V_\alpha^M \models \varphi.$$

In fact, under the assumption of a proper class of Woodin cardinals, the universally Baire sets admit of a stratification and this provides one with a hierarchy

$$\langle \Omega_\Gamma \mid \Gamma \subseteq \Gamma^\infty \rangle$$

of increasingly strong logics, where Γ^∞ is the totality of universally Baire sets. The details of this stratification and the above definition will not be important for our purposes. The important point it that these logics are generically invariant relative to large cardinal axioms. For the limiting case of $\vdash_\Omega$ the result to this effect is the following:

THEOREM 2.13 (Woodin [17]). *Assume there is a proper class of Woodin cardinals. Then for each complete Boolean algebra $\mathbb{B}$,*

$$V \models \text{“}T \vdash_\Omega \varphi\text{”} \textit{ iff } V^{\mathbb{B}} \models \text{“}T \vdash_\Omega \varphi\text{”}.$$

[17]The precise definition of the theory AD^+ will not be necessary for our purposes. See [16] for details.

§3. Generically invariant logics of first order. The logics in the above hierarchy of strong logics resemble first-order logic in that they have strong absoluteness properties, only now these properties are secured relative to increasingly strong background theories. In each case the absoluteness of the logic has many facets. We shall isolate two important facets of absoluteness—*generic invariance* and *faithfulness*—and characterize the strongest logic of first order that shares these features. This will actually lead to a countable collection of strong logics and a particularly important element of this collection will be singled out for special attention and called *generically invariant logic*.

Our approach is both *local* in that the definition is given with respect to the background theory ZFC without large cardinal assumptions and *flexible* in that as one increases the strength of the background theory with large cardinal axioms the logical consequence relation becomes progressively stronger. In particular, for each background theory considered in the previous section, the corresponding logic will be subsumed by generically invariant logic. A particularly important stage is reached when one comes to the theory ZFC + "There is a proper class of Woodin cardinals", for at this stage generically invariant logic coincides with Woodin's Ω-logic. Our definition of generically invariant logic provides a new perspective on Woodin's Ω-logic, one that is made available by recent developments in the theory of forcing.

In the next section we shall do the same for logics of second order, characterizing in a local and flexible fashion the strongest logic of second order which is generically invariant and faithful relative to strong background assumptions. It will turn out that generically invariant logic of first order is (provably in ZFC) equivalent to generically invariant logic of second order in that (provably in ZFC) the two are Turing equivalent.

3.1. Two facets of absoluteness. Recall that the general form of a first-order strong logic is as follows: Let L be a first-order language and let $\Phi(x)$ be a formula that defines a class of L-structures (possibly including Boolean-valued structures). Then, for each recursively enumerable set T of sentences in L and for each sentence φ of L,

$$T \models_\Phi \varphi$$

iff for all L-structures M such that $\Phi(M)$,

$$\text{if } M \models T \text{ then } M \models \varphi.$$

The structures M that satisfy Φ are the *test structures* that the logic consults to certify logical implication and validity. As above, we shall restrict our attention to the case where L is the language of set theory.

The strong logics of the previous section all have this general form. In addition they have two facets of absoluteness, the first of which has already been introduced.

DEFINITION 3.1. Suppose $\models_\Phi$ is a strong logic and $\mathrm{ZFC}^{(+)}$ is an extension of ZFC.

(1) The strong logic $\models_\Phi$ is *generically invariant in* $\mathrm{ZFC}^{(+)}$ if, for each recursively enumerable set of sentences T and for each sentence φ, the following is a theorem of $\mathrm{ZFC}^{(+)}$: Suppose $\mathbb{B}$ is a complete Boolean algebra. Then

$$V \models \text{“}T \models_\Phi \varphi\text{” iff } V^{\mathbb{B}} \models \text{“}T \models_\Phi \varphi\text{”}.$$

(2) The strong logic $\models_\Phi$ is *faithful in* $\mathrm{ZFC}^{(+)}$ if, for each sentence φ, the following is a theorem of $\mathrm{ZFC}^{(+)}$: For all complete Boolean algebras $\mathbb{B}$,

$$\text{if } V \models \text{“}\varnothing \models_\Phi \varphi\text{” then } V^{\mathbb{B}} \models \varphi.$$

Each of the logics of the previous section is defined relative to a sufficiently strong generically invariant background theory. Letting $\mathrm{ZFC}^{(+)}$ be this theory we have that the corresponding logic is generically invariant and faithful in $\mathrm{ZFC}^{(+)}$.

The importance of these two conditions is this: Think of the models $V^{\mathbb{B}}$ as the *possible worlds* from the point of view of the common background theory. The disputants in a foundational dispute can disagree on which of these worlds is the true universe of sets. However, when they employ a logic to articulate their differences the logic should be such that each party agrees on (i) what implies what and (ii) the fact that logical validities are true; in other words, the conditions of generic invariance and faithfulness should hold relative to the common background theory.

Our goal is to characterize in ZFC the strongest strong logics that are generically invariant and faithful. As motivation for our approach it is useful to contrast two ways in which generic invariance might be secured. The first way was the way it is secured in the previous section: Here one first shows that the logical consequence relation has a given complexity and then one shows that the background theory secures the generic invariance of statements with this level of complexity. For example, the theory ZFC + “$\forall X\ X^\#$ exists” enables one to define the relation $\models_{\Sigma^1_2}$ and it ensures that this relation is generically invariant. But there is also a second way—one can try to secure generic invariance by *brute force* by having the various universes $V^{\mathbb{B}}$ consult the same collection of test structures. There is an obvious difficulty with this. For there appears to be an asymmetry in that although V can access $V^{\mathbb{B}}$ (and so capture the test structures that it consults) it is less clear that $V^{\mathbb{B}}$ can access V (and hence the test structures that it consults). Surprisingly, it turns out that this hurdle can be overcome.

3.2. The generic-multiverse. In the case where M' is a generic extension of M let us say that M is a *generic refinement* of M'. The *generic-multiverse* *generated by the universe M* is the collection $\mathbb{V}_M$ obtained by starting with M and closing under generic extensions and generic refinements.[18]

We wish to render this informal characterization more precise. This is readily done when one occupies an *external* vantage point. For example, suppose that M is a countable transitive model of ZFC. In this case one can easily talk about generic extensions and generic refinements of M. A generic extension of M is a set of the form $M[G]$ where $G \subseteq \mathbb{P}$ is M-generic for some partial order $\mathbb{P} \in M$. Thus, $\mathbb{V}_M$ can be characterized as the smallest set such that $M \in \mathbb{V}_M$ and, for each pair of countable transitive models $(N, N[G])$ such that $N \models \text{ZFC}$ and $G \subseteq \mathbb{P}$ is N-generic for some partial order in $\mathbb{P} \in N$, if either N or $N[G]$ is in $\mathbb{V}_M$, then both N and $N[G]$ are in $\mathbb{V}_M$. So characterized, $\mathbb{V}_M$ is simply a set of countable transitive models of ZFC.

It will be important for our purposes to give an *internal* characterization of the generic-multiverse. Here one is to think of the multiverse as being generated by the background universe V. This leads to two difficulties: First, from this perspective one cannot think of a generic extension as having the form $V[G]$ where G is an additional object that is adjoined to the universe of sets. This hurdle is easily overcome by treating generic extensions as Boolean-valued models, that is, as having the form $V^{\mathbb{B}}$, where $\mathbb{B}$ is a complete Boolean algebra. In this way one obtains a characterization that is internal to V. The second problem is more subtle. This is the problem of talking about generic refinements of generic extensions of V and of generic refinements of generic extensions of generic refinements of generic extensions of V and so on through all finite alternations.

The basic question is whether there is a way of talking about V in a generic extension $V[G]$. It turns out that there is. This is fundamental result in the theory of forcing, which is a corollary of independent work of Laver and Woodin and was noted independently by Laver, Reitz and Woodin. See [18] and [38].

THEOREM 3.2. *Suppose V is a transitive model of* ZFC. *Suppose $V = N[G]$ where $G \subseteq \mathbb{P} \in N$ is N-generic. Then N is definable in V with the parameter $\mathcal{P}(\delta)^N$, where $\delta = (|\mathbb{P}|^+)^V$.*

The formula $\psi(x, y)$ that defines the ground model is fixed beforehand and applies uniformly across models. For a given parameter a, let

$$N_{\psi(a)} = \{x \in V \mid V \models \psi(x, a)\}.$$

Thus, for certain parameters a, if the hypothesis of the theorem is met then V will satisfy "I am a generic extension of $N_{\psi(a)}$". In this way V can access its generic refinements.

This result can be used to capture the idea of truth across the generic-multiverse. A formula φ is *true across the generic-multiverse* $\mathbb{V}_M$ (*or* $\mathbb{V}$) if it

[18]The generic-multiverse was introduced in [38]. See that paper for a detailed discussion.

is true in each universe of $\mathbb{V}_M$ (or $\mathbb{V}$). In such a circumstance we shall write: $\mathbb{V}_M \models \varphi$ (or $\mathbb{V} \models \varphi$). For each sentence φ in the language of set theory, one can uniformly define a sentence φ^* which holds in V if and only if φ holds in each universe in $\mathbb{V}_M$ (or $\mathbb{V}$), namely, let φ^* be the sentence

> It is not the case that there exists a Boolean algebra $\mathbb{B}$ such that $V^{\mathbb{B}}$ satisfies: There exists a parameter a such that V is a generic extension of $N_{\psi(a)}$ and $N_{\psi(a)} \models \neg\varphi$.

We claim that this computes truth across the multiverse: In the case where the multiverse $\mathbb{V}_M$ is generated by a countable transitive model M:

LEMMA 3.3. $M \models \varphi^*$ *iff* $\mathbb{V}_M \models \varphi$.[19] *In other words, to compute truth across the multiverse it suffices to consult only generic refinements of generic extensions.*[20]

PROOF. Let M_0 be in the generic-multiverse $\mathbb{V}_M$. Suppose that M' is in the generic-multiverse $\mathbb{V}_M$ and $M' \models \varphi$. We claim that there exists M'' which is a generic refinement of a generic extension of M_0 such that $M'' \models \varphi$. A counter-example to this claim about M_0 and φ is a sequence of the form $\langle M_0, M_1, \ldots, M_n\rangle$ of successive generic refinements and generic extensions such that (i) $M_n \models \varphi$ and (ii) for every M'' which is a generic refinement of a generic extension of M_0, $M'' \not\models \varphi$. Let M_0 and φ be such that there is a counter-example $\langle M_0, M_1, \ldots, M_n\rangle$ with n minimal (ranging over all counter-examples for all M_0 and φ for which the claim fails). By the minimality of n, the claim holds of M_1 and φ. There are two cases. The first case is where M_1 is a generic extension of M_0. Since a generic extension of a generic extension of M_0 is a generic extension of M_0 it follows that the claim holds of M_0 and φ. The second case is where M_1 is a generic refinement of M_0. Since the claim holds of M_1 and φ we can let M_2' be a generic extension of M_1 via $\mathbb{Q}$ and let M_3' be generic refinement of M_2' such that $M_3' \models \varphi$. Now, M_2' satisfies "I am a generic extension of a model of φ". It follows that M_1 satisfies "There is a condition $q \in \mathbb{Q}$ which forces the statement "I am a generic extension of a model of φ" ". Let $G \subseteq \mathbb{Q}$ be M_0-generic. Since M_0 is a generic extension of M_1, $M_0[G]$ is a generic extension of $M_1[G]$. It follows that the claim holds for M_0 and φ, which completes the proof. ⊣

3.3. Generically invariant logic. The results in the previous section show that each universe of the generic-multiverse (and hence each model of the form $V^{\mathbb{B}}$) can access all universes of the generic-multiverse (in that it can define truth across the generic-multiverse). This enables us to design by *brute*

[19]In the case where the multiverse $\mathbb{V}$ is that given by a proper class model V this equivalence continues to hold with 'V' in place of 'M' and '$\mathbb{V}$' in place of '$\mathbb{V}_M$' only now $\mathbb{V}$ is understood in the internal sense discussed above.

[20]This observation is due to Woodin. See [38]. The referee pointed out that it is not generally true that every universe in the multiverse is a generic refinement of a generic extension of the model generating the multiverse.

force logics that are generically invariant—we simply ensure that all generic extensions $V^{\mathbb{B}}$ consult the same collection of test structures.

We now wish to characterize (in ZFC) the strongest logic of first order that is generically invariant and faithful in ZFC. To do so let us proceed abstractly. Assume that $\models_\Phi$ is a strong logic which is generically invariant and faithful in ZFC. There are strong versions of these notions that the strong logic must satisfy.

Definition 3.4. Suppose $\models_\Phi$ is a strong logic and $\mathrm{ZFC}^{(+)}$ is an extension of ZFC.

(1) The strong logic $\models_\Phi$ is *strongly generically invariant in* $\mathrm{ZFC}^{(+)}$ if, for each recursively enumerable set T and each sentence φ, the following is a theorem of $\mathrm{ZFC}^{(+)}$: Suppose T is a recursively enumerable set of sentences and φ is a sentence. Then

$$V \models \text{“}T \models_\Phi \varphi\text{” iff } \mathbb{V} \models \text{“}T \models_\Phi \varphi\text{”}.$$

(2) The strong logic $\models_\Phi$ is *strongly faithful in* $\mathrm{ZFC}^{(+)}$ if, for each sentence φ, the following is a theorem of $\mathrm{ZFC}^{(+)}$:

$$\text{if } V \models \text{“}\varnothing \models_\Phi \varphi\text{” then } \mathbb{V} \models \varphi.$$

Lemma 3.5. *Suppose* $\mathrm{ZFC}^{(+)}$ *is an extension of* ZFC *that is preserved under generic extensions and generic refinements. Then*

(1) *the strong logic* $\models_\Phi$ *is generically invariant in* $\mathrm{ZFC}^{(+)}$ *iff it is strongly generic invariant in* $\mathrm{ZFC}^{(+)}$ *and*

(2) *the strong logic* $\models_\Phi$ *is faithful in* $\mathrm{ZFC}^{(+)}$ *iff it is strongly faithful in* $\mathrm{ZFC}^{(+)}$.

Proof. The proof of the lemma is immediate from the definitions. ⊣

Thus, *any* strong logic $\models_\Phi$ which is generically invariant and faithful in ZFC, must also be *strongly* generically invariant and *strongly* faithful in ZFC. Now, the second constraint—strong faithfulness—shows that the strongest possible logic that is provably generically invariant and faithful in ZFC will be such that the validities are contained in the generic multiverse truths. But as we saw in the previous section, the notion of generic-multiverse truth is itself generically invariant and faithful in ZFC. Thus, generic-multiverse truth is itself the strongest logic that is generically invariant and faithful in ZFC.

This last statement requires some qualification. For metamathematical reasons pertaining to the undefinability of truth one cannot define truth across the generic-multiverse $\mathbb{V}_M$ in M (or $\mathbb{V}$ in V). However, it is possible to define generic-multiverse truth at a given level of complexity. For example, for each n we would like to define $\varnothing \models_{\mathrm{GI}_n} \varphi$ iff for all $W \in \mathbb{V}$, if $W_\alpha \prec_n W$ then $W \models \varphi$. In other words, this logic takes as test structures all rank initial segments of universes of the generic-multiverse that are Π_n-substructures of the parent universe. By our earlier considerations, this relation is definable in ZFC and clearly, in ZFC, we have that

$$\varnothing \models_{\mathrm{GI}_n} \varphi \text{ iff } \mathbb{V} \models \varphi,$$

for each Π_n-sentence φ. One problematic feature of this approach is that we have defined our test structures by accessing truth in the parent universe and this seems to undermine the claim to being a true logic. In the case of $n = 2$, however, there is a perfectly satisfactory version:

Definition 3.6. Assume ZFC. Let $\langle T_n \mid n < \omega \rangle$ enumerate the recursively enumerable theories and let $\langle \psi_i \mid i < \omega \rangle$ enumerate the sentences in the language of set theory. For each $n, i \in \omega$, let φ_n^i be the statement

$$\text{For all } \alpha, \text{ if } V_\alpha \models T_n \text{ then } V_\alpha \models \varphi_i.$$

Let $(\varphi_n^i)^*$ be the associated sentence such that $(\varphi_n^i)^*$ holds if and only if $\mathbb{V} \models \varphi$. (See Section 3.2). Since $(\varphi_n^i)^*$ is uniform in i and n and since $(\varphi_n^i)^*$ is of bounded complexity we can define

$$T_n \models_{\mathrm{GI}} \psi_i \text{ iff } T((\varphi_n^i)^*),$$

where T is a bounded truth predicate of sufficiently complexity.

Letting

$$\mathcal{V}_{\mathrm{GI}} = \{\ulcorner \varphi \urcorner \mid \varnothing \models_{\mathrm{GI}} \varphi\}$$

we have that $\mathcal{V}_{\mathrm{GI}}$ is Turing equivalent to the set of Π_2-generic-multiverse truths.

To summarize: The hierarchy of logics $\models_{\mathrm{GI}_n}$ comprise the strongest logics that are generically invariant and faithful in ZFC. Of these a particularly important strong logic is $\models_{\mathrm{GI}}$. We shall call this *generically invariant logic*.

3.4. Strength. It is important to note that we have defined generically invariant logic $\models_{\mathrm{GI}}$ in ZFC without reference to strong background assumptions. Moreover, as one increases the strength of the background theory through the addition of large cardinal axioms the logic increases in strength. In particular, as one increases the background theory the logics of the previous section are successively encompassed. For example, during the first two stages we have:

Lemma 3.7. *Assume* ZFC. *Then for each recursively enumerable theory T and for each sentence φ,*

$$\textit{if } T \models_\beta \varphi \textit{ then } T \models_{\mathrm{GI}} \varphi.$$

Lemma 3.8. *Assume* ZFC *and that $X^\#$ exists for each set X. Then for each recursively enumerable theory T and for each sentence φ,*

$$\textit{if } T \models_{\Sigma^1_2} \varphi \textit{ then } T \models_{\mathrm{GI}} \varphi.$$

And as one continues to increase the large cardinal background assumptions, one marches up the hierarchy of strong logics from the previous section.

Eventually, when one reaches the background theory ZFC + "There is a proper class of Woodin cardinals", the strong logic $\vdash_\Omega$ is encompassed. This is an immediate corollary of work of Woodin which we now describe.[21] First we need a definition.

Definition 3.9 (Woodin). Suppose T is a recursively enumerable theory and φ is a sentence. Then $T \models_\Omega \varphi$ iff for all complete Boolean algebras $\mathbb{B}$ and for all ordinals α,

$$\text{if } V_\alpha^{\mathbb{B}} \models T \text{ then } V_\alpha^{\mathbb{B}} \models \varphi.$$

In this definition the collection of test structures that the logic consults is *smaller* than that of the case of generically invariant logic. The issue, of course, is whether the above logic is generically invariant. In the case of $\models_{\mathrm{GI}}$ this feature was build in by design but in the case of $\models_\Omega$ it is far from obvious.

Theorem 3.10 (Woodin [1]). *Assume* ZFC *and that there is a proper class of Woodin cardinals. Suppose T is a recursively enumerable theory and φ is a sentence. Then for each complete Boolean algebra $\mathbb{B}$*

$$V \models \text{“}T \models_\Omega \varphi\text{”} \textit{ iff } V^{\mathbb{B}} \models \text{“}T \models_\Omega \varphi\text{”}.$$

Corollary 3.11. *Assume* ZFC *and that there is a proper class of Woodin cardinals. Suppose T is a recursively enumerable theory and φ is a sentence. Then*

$$T \models_\Omega \varphi \textit{ iff } T \models_{\mathrm{GI}} \varphi.$$

Thus, we have recast Ω-logic in terms of the strongest logic that is generically invariant and faithful. Turning to the provability notion the key result is the following:

Theorem 3.12 (Woodin [1]). *Assume* ZFC *and that there is a proper class of Woodin cardinals. Suppose T is a recursively enumerable theory and φ is a sentence. Then*

$$\textit{if } T \vdash_\Omega \varphi \textit{ then } T \models_\Omega \varphi.$$

Corollary 3.13. *Assume* ZFC *and that there is a proper class of Woodin cardinals. Suppose T is a recursively enumerable theory and φ is a sentence. Then*

$$\textit{if } T \vdash_\Omega \varphi \textit{ then } T \models_{\mathrm{GI}} \varphi.$$

[21] For proofs of the basic theorems on Ω-logic see [1].

Thus, generically invariant logic scales up to subsume the strongest logic of the previous section.

Let us next address the question of the complexity of $\mathcal{V}_{\mathrm{GI}}$. Again the answer to this question depends on the background theory. For example, if $V = L$ then it is straightforward to see that $\mathcal{V}_{\mathrm{GI}}$ is Turing equivalent to the complete Π_2-set of natural numbers. The situation becomes more interesting in the context of large cardinal axioms. Here the following fundamental conjecture is the conjectured completeness theorem for the "syntactic" notion $\vdash_\Omega$ with respect to the semantic notion $\models_\Omega$.

Ω Conjecture (Woodin). *Assume there is a proper class of Woodin cardinals. Then for each sentence φ,*

$$\varnothing \models_\Omega \varphi \textit{ iff } \varnothing \vdash_\Omega \varphi.$$

Theorem 3.14 (Woodin). *Assume that there is a proper class of Woodin cardinals and the Ω Conjecture is true. Then $\mathcal{V}_\Omega$ (and hence $\mathcal{V}_{\mathrm{GI}}$) is definable in $H(\delta_0^+)$, where δ_0 is the least Woodin cardinal.*

Assuming that an additional conjecture holds (the AD^+ Conjecture) one has that $\mathcal{V}_\Omega$ (and hence $\mathcal{V}_{\mathrm{GI}}$) is definable in $H(\mathfrak{c}^+)$, where $\mathfrak{c}$ is the cardinality of the continuum. See [38]. On the other hand if there is a proper class of Woodin cardinals and the Ω Conjecture fails then it is still possible (given our current understanding) that $\mathcal{V}_{\mathrm{GI}}$ Turing equivalent to the Π_2-complete set of truths in V. The simplest way this could happen is if there is a sentence (or a recursive set of axioms) A such that $V \models \mathrm{ZFC} + A$ and

$$\{\varphi \in \Pi_2 \mid V \models \varphi\} = \{\varphi \in \Pi_2 \mid \mathrm{ZFC} + A \models_\Omega \varphi\}.$$

§4. Generically invariant logics of second order. We now turn to strong logics of second order. Let L be a first-order language—such as the language of PA or the language of ZFC—and let L' be the associated second-order language obtained by adding quantifiers of second-order. An L'-structure has the form $\langle M, S\rangle$ where M is an L-structure and S is a sufficiently closed subset of $\mathcal{P}(M)$. The second-order variables are interpreted to range over S. To say that S is *sufficiently closed* (with respect to the structure M) is to say that S is closed under definability over M with parameters; in other words, that $\langle M, S\rangle$ satisfies the comprehension axioms.

Recall that the general form of a second-order logical consequence relation is as follows: Let L be a first-order language and let L' be the second-order extension. Let $\Phi(x, y)$ be a formula that defines a class of L'-structures. Then, for a recursively enumerable set T of sentences in L' and for a sentence φ in L', set

$$T \models_{\Phi(x,y)} \varphi$$

iff for all L'-structures $\langle M, S\rangle$ such that $\Phi(M, S)$,

$$\text{if } \langle M, S\rangle \models T \text{ then } \langle M, S\rangle \models \varphi.$$

There are now two degrees of variability in the formula $\Phi(x, y)$ which defines the test structures: One can restrict to richer first-order domains M and one can restrict to richer collections of second-order domains S or one can do both. In what follows, we shall allow arbitrary structures for the first-order domain and concentrate on variation along the second dimension.

4.1. General. The weakest version of second-order logic is given by a formula $\Phi_{\mathrm{H}}(x, y)$ which holds of all L'-structures and places no additional constraints on either M or S (apart from the minimal necessary requirement that S be closed under definability over M with parameters). Let

$$\mathcal{V}_{\mathrm{HS}} = \{\ulcorner\varphi\urcorner \mid \varnothing \models_{\Phi_{\mathrm{H}}} \varphi\}.$$

Here 'HS' stands for 'Henkin Second-Order Logic'.

THEOREM 4.1 (Henkin [10]). *The system S_2 is sound and complete for $\mathcal{V}_{\mathrm{HS}}$.*

This provides a nice parallel with the case of first-order logic in that in either case when one imposes no conditions on the test structures (beyond the minimal condition required in the second-order case) then the result is a logic that is highly absolute. In fact, the two logics are Turing equivalent to the Σ^0_1-complete set of integers.

To obtain stronger second-order logics one considers richer second-order domains. There are two extreme cases at opposite ends of the spectrum: In the first case the second-order domain S of $\langle M, S\rangle$ is constrained to be the *minimal* domain, namely $\mathrm{Def}(M)$; in the second case the second-order domain is constrained to be the *maximal* domain, namely, $\mathcal{P}(M)$. Let us begin with the first and work our way through the spectrum to the second.

4.2. Minimal. Let $\Phi_{\mathrm{D}}(x, y)$ be a formula such that $\Phi_{\mathrm{D}}(M, S)$ iff $\langle M, S\rangle$ is an L'-structure and $S = \mathrm{Def}(M)$, that is, the set of subsets of M that are definable over M with parameters. Thus, whereas above the permissible test structures were of the form $\langle M, S\rangle$ where S could be *any* collection of subsets of M that is closed under definability over M with parameters, we are now paring down and demanding that S be the minimal such collection. Let

$$\mathcal{V}_{\mathrm{DS}} = \{\varphi \mid \varnothing \models_{\Phi_{\mathrm{D}}} \varphi\}.$$

Here 'DS' stands for 'Definable Second-Order Logic'. Let us say that a language L is "sufficiently rich" if it contains either the language of first-order arithmetic $(0, S, +, \cdot)$ or the language of set theory.

THEOREM 4.2 (Lindström [19]). *Suppose L is a sufficiently rich language with finitely many predicate, function and constant symbols. Let $\mathcal{V}_{\mathrm{DS}}$ be defined as above with respect to L. Then $\mathcal{V}_{\mathrm{DS}}$ is the complete Π^1_1 set of integers.*[22]

[22] Lindström also shows that such a language is sufficiently rich to characterize the structure of the natural numbers up to isomorphism. It should be noted, however, that in the case where L has infinitely many predicate, function or constant symbols, the resulting logical relation can be much simpler than Π^1_1, in fact, it can admit of a complete (recursively enumerable) axiomatization. See [19] for details.

Thus, $\mathcal{V}_{\text{DS}}$ is Turing equivalent to $\mathcal{V}_\omega$ and the relation $\models_{\text{DS}}$ is generically invariant in ZFC.

4.3. Generically invariant second-order logic. It is of interest to ask (just as we did in the first-order context) how strong one can make the logic while retaining absoluteness relative to a strong background theory consisting of large cardinal axioms. In particular, what is the strongest version of second-order logic which is generically invariant against the backdrop of (current) large cardinal axioms?

The argument from the beginning of Section 3 can be adapted to the second-order case: To ensure that the logic is generically invariant we consult test structures from across the generic-multiverse. To increase the strength of the logic there are two parameters that we can vary—the first-order domain and the second-order domain. For the second-order domain there are many options: For example, one can take all subsets of the first-order domain M that appear in the constructible universe built over M, or one can take all subsets of the first-order domain M that appear in one of the many fine-structural inner models built over M, and so on. In each case one can devise a generically invariant second-order logic by brute force by using the technique of Section 3. Of course, the maximal choice is to simply take the second-order domain to be the full powerset of M as computed in the universe in the generic multiverse from which M is selected, that is, if M is in $W \in \mathbb{V}$ then we take the corresponding test structure to be $\langle M, \mathcal{P}(M)^W, \in \rangle$.

Suppose T is a recursively enumerable set of L' sentences and φ is an L' sentence (where recall that L' is the language of second-order set theory). Let $\models_{\text{FS}}$ be the relation of full second-order logic. Let

$$T \models_{\text{GIS}} \varphi$$

iff

$$\mathbb{V} \models \text{“}T \models_{\text{FS}} \varphi\text{”}.$$

As we shall see in the next subsection the relation $T \models_{\text{FS}} \varphi$ is Π_2-definable and so the definition of $\models_{\text{GIS}}$ is well-formed. Another option is to strengthen the first-order domain by restricting to rank initial segments. Let

$$T \models_{\text{GIS}'} \varphi$$

iff

$$\mathbb{V} \models \text{“For all } \alpha, \text{ if } \langle V_\alpha, V_{\alpha+1}, \in \rangle \models T \text{ then, } \langle V_\alpha, V_{\alpha+1}, \in \rangle \models \varphi\text{”}.$$

(It follows from the techniques of Section 3 that these definitions are well-formed.) In each case ‘GIS’ stands for ‘Generically Invariant Second-Order Logic’.

Theorem 4.3. *Assume* ZFC. *Then*

$$\mathcal{V}_{\text{GI}} \equiv_T \mathcal{V}_{\text{GIS}} \equiv_T \mathcal{V}_{\text{GIS}'}.$$

Generically invariant second-order logic is the strongest version of second-order logic that is generically invariant relative to large cardinal assumptions. An immediate corollary is that relative to this semantics, second-order ZFC does not settle CH, in contrast to the case of full second-order logic.

The above theorem shows that there is an intimate connection between generically invariant logic of first order and generically invariant logic of second order—the move to second-order does not lead to an increase in strength. As we shall see in the next section, in the case of full second-order logic there is also an intimate connection with a "logic" of first order, namely, the one based on the collection of sets structures consisting of true rank initial segments of V. This is the pseudo-logic that we ruled out in Section 3.3.

4.4. Full second-order logic. Let us now move to the other extreme and maximize the second parameter, letting $\Phi_{\mathrm{F}}(x, y)$ be a formula such that $\Phi_{\mathrm{F}}(M, S)$ iff $\langle M, S\rangle$ is an L'-structure and $S = \mathcal{P}(M)$. Let

$$\mathcal{V}_{\mathrm{FS}} = \{\ulcorner\varphi\urcorner \mid \varnothing \models_{\Phi_{\mathrm{F}}} \varphi\}.$$

Here 'FS' stands for 'Full Second-Order Logic'. To simplify matters, in what follows we shall concentrate on the case where L is the language of set theory. So our structures have the form $\langle M, \mathcal{P}(M), \in\rangle$, where we have now made the additional structure of M explicit.

Notice that although we have ensured that the second-order domain is maximal we have considered arbitrary first-order domains. One might think that further strength could be obtained by narrowing the class of first-order domains. To take an extreme case let $\mathcal{V}_{\mathrm{FS}'}$ be defined as above except where now we demand that M be of the form V_α for some ordinal α. In the second-order context ZFC and its variants can be formulated as single sentences. Let $\mathrm{ZC}^{(\mathrm{VN})}$ be the second-order sentence consisting of the second-order sentence ZC (Zermelo set theory with AC) conjoined with the sentence expressing "for all ordinals α, V_α exists". This sentence has the feature that $\langle M, \mathcal{P}(M), \in\rangle \models \mathrm{ZC}^{(\mathrm{VN})}$ iff $M = V_\lambda$ for some limit ordinal $\lambda > \omega$. Now let φ be an arbitrary sentence and let φ' be the sentence expressing "for all $\alpha \geqslant \omega$, $V_\alpha \models \varphi$". It follows that $\varphi \in \mathcal{V}_{\mathrm{FS}'}$ iff $(\mathrm{ZC}^{(\mathrm{VN})} \to \varphi') \in \mathcal{V}_{\mathrm{FS}}$. Thus, within $\mathcal{V}_{\mathrm{FS}}$ one can achieve the effect of restricting to first-order structures of the form V_α and so there is no loss of generality in working with arbitrary domains.

Notice that

$$\langle M, \mathcal{P}(M), \in\rangle \models \mathrm{ZC}^{(\mathrm{VN})} + \Sigma_1\text{-Replacement}$$

if and only if $M = V_\lambda$ for some $\lambda > \omega$ such that $|V_\lambda| = \lambda$. Moreover, by condensation, each such λ is such that $V_\lambda \prec_{\Sigma_1} V$ and so a Π_2 sentence is true in V iff for each such λ, $V_\lambda \models \varphi$. Thus, for each Π_2 sentence, we have that

$$V \models \varphi$$

if and only if

$$(\mathrm{ZC}^{(\mathrm{VN})} + \Sigma_1\text{-Replacement} \to \varphi) \in \mathcal{V}_{\mathrm{FS}}.$$

In this way Π_2 truth in V can be computed from $\mathcal{V}_{\mathrm{FS}}$.

THEOREM 4.4 (Väänänen [33]). *Assume* ZFC. *Then $\mathcal{V}_{\mathrm{FS}}$ is the complete Π_2 set of integers.*

PROOF. We first show that $\mathcal{V}_{\mathrm{FS}}$ is Π_2-definable over V. The natural definition of $\varphi \in \mathcal{V}_{\mathrm{FS}}$ is

$$\forall x\, (\mathrm{Str}(x) \to \mathrm{Sat}_\varphi(x))$$

where $\mathrm{Str}(x)$ is the Σ_2-statement asserting that x is a structure of the relevant type and $\mathrm{Sat}_\varphi(x)$ is the Σ_2-statement asserting that x is a structure of the relevant type and φ is true in it. This statement is Π_3. To reduce the complexity we simply write

$$\forall \lambda > \omega\, \big(|V_\lambda| = \lambda \to \forall x \in V_\lambda\, (V_\lambda \models (\mathrm{Str}(x) \to \mathrm{Sat}_\varphi(x)))\big).$$

We now show that if $X \subseteq \omega$ is Π_2-definable over V then X is continuously (in fact, recursively) reducible to $\mathcal{V}_{\mathrm{FS}}$. Suppose

$$X = \{n \in \omega \mid \forall x\, \exists y\, \varphi(n, x, y)\}.$$

For each $n \in \omega$, let φ_n be the statement $\forall x\, \exists y\, \varphi(\bar{n}, x, y)$, where $\bar{n}$ is a name for n. It follows that $n \in X$ iff

$$\text{“}\mathrm{ZC}^{(\mathrm{VN})} + \Sigma_1\text{-Replacement} \to \varphi_n\text{”} \in \mathcal{V}_{\mathrm{FS}}.$$

Thus, X is Turing reducible to $\mathcal{V}_{\mathrm{FS}}$. ⊣

It is instructive to compare the situation with that of Ω-logic. The above result shows that

$$\mathrm{Th}_{\Pi_2}(V) \equiv_T \mathcal{V}_{\mathrm{FS}}.$$

Likewise, in the scenario considered at the end of Section 3.4 (where one has a radical failure of the Ω Conjecture), we saw that

$$\mathrm{Th}_{\Pi_2}(V) \equiv_T \mathcal{V}_{\Omega}.$$

However, although in this scenario $\mathcal{V}_{\mathrm{FS}}$ and $\mathcal{V}_\Omega$ have the same complexity there is a major difference—$\mathcal{V}_\Omega$ is *generically invariant* against the backdrop of the theory ZFC + “There is a proper class of Woodin cardinals”, while $\mathcal{V}_{\mathrm{FS}}$ is *generically fragile* against the backdrop of any theory of the form ZFC + φ where φ is a (current) large cardinal axiom. This might seem counter-intuitive.

The point is that the reduction

$$(\mathcal{V}_{\mathrm{FS}})^V \equiv_T \mathrm{Th}_{\Pi_2}(V)$$

is provable in ZFC and is uniform. Letting $V[G]$ be a set generic extension of V we have that

$$(\mathcal{V}_{\mathrm{FS}})^{V[G]} \equiv_T \mathrm{Th}_{\Pi_2}(V[G])$$

by *exactly the same Turing reduction*. In other words, $\mathcal{V}_{\mathrm{FS}}$ *co-varies* with the Π_2-theory under forcing. Consequently, $\mathcal{V}_{\mathrm{FS}}$ is highly generically fragile. It has *no* degree of absoluteness—*any* non-trivial forcing perturbs the consequence relation. In contrast, in the scenario involving the radical failure of the Ω Conjecture, $\mathcal{V}_\Omega$ and the Π_2-theory do not co-vary. The former is generically invariant while the latter is generically fragile. The reduction

$$(\mathcal{V}_\Omega)^V \equiv_T \mathrm{Th}_{\Pi_2}(V)$$

involves the *fixed* universe V.

The point I wish to make is that from the present perspective the problem is not with the complexity of full second-order logic—it has the same complexity as Ω-logic (in the scenario under consideration) and Ω-logic is not problematic since it is generically invariant relative to large cardinal axioms. Rather, the problem with full second-order logic is that it is generically fragile even against the backdrop of all (current) large cardinal axioms, the reason being that $\mathcal{V}_{\mathrm{FS}}$ is uniformly Turing equivalent to the Π_2 truths of the universe of sets in such a way that the two co-vary, in other words, full second-order logic is "coupled into" Π_2-truth in the ambient universe.

It is useful to compare the difference between Ω logic and full second-order logic in the context of a foundational dispute. Suppose two parties disagree on what holds in the true universe. One holds A while the other holds A'. They will still agree on Ω-logic. Their disagreement on V—in particular, $\mathrm{Th}_{\Pi_2}(V)$—will not infect their agreement on logic. In the case of second-order logic, in order to agree on $\mathcal{V}_{\mathrm{FS}}$ the two parties would have to agree on $\mathrm{Th}_{\Pi_2}(V)$.

4.5. Expressive power. From the present perspective the problem with full second-order logic is not mathematical entanglement alone since all of the strong logics we have considered are mathematically entangled and, in the limiting case, it is possible that the strongest logics in the hierarchies we have considered (generically invariant first-order logic and generically invariant second-order logic) have the same complexity as full second-order logic. Rather, the problem lies with the *kind* of mathematical entanglement involved. In the case of full second-order logic there is a radical failure of absoluteness due to a direct "coupling" with truth in the universe of sets. Instead of being an invariant constant against which differences about Π_2-truth in the universe of sets can be articulated (as is the case with all of the strong logics we have considered) full second-order logic co-varies with Π_2-truth in the universe of sets and so to agree on the logic is to agree on Π_2-truth in the universe of sets.

Now one might maintain that although full second-order logic is unsuitable for the role of a neutral background framework it is nonetheless indispensable for another reason, namely, because we need its expressive resources to secure categoricity and thereby rule out unintended interpretations. Indeed one might maintain that since we implicitly accept that we have such categorical characterizations full second-order logic is something that we implicitly accept.

I would like now to briefly discuss this and related issues.[23] In short, I think that the idea that full second-order logic can secure categoricity in some absolute sense—a sense that is not available without full second-order logic—is an illusion.

To begin with one should note that categorical characterizations can be given in second-order logics where the semantics is less than full. In fact, the minimal version of second-order logic that we have considered—Lindström's definable second-order logic—is already sufficiently rich in expressive resources to provide a categorical characterization of the natural numbers. What Lindström shows is that if L is a language extending the language of arithmetic, L_A, then there is an L-sentence φ such that for every L-structure M, M is a DS-model of φ iff the reduct of M to L_A is isomorphic to $\mathbb{N}$. Thus, full second-order logic is not needed to give a categorical characterization of the natural numbers. One might protest that the sentence φ in Lindström's theorem is a rather complicated sentence involving partial satisfaction predicates yet what captures our idea that the natural numbers are unique up to isomorphism is the categoricity result pertaining to second-order arithmetic, PA_2, the proof of which is direct and transparent.

But what really is the significance of this categoricity result and what role is full second-order logic playing in it? Let us first look at a very simple instance. Suppose one is given two structures M and N which satisfy PA_2. The isomorphism is simply definable over these structures—one simply maps the zero element of M to the zero element of N and inductively maps successors to successors. To show that this map is one-to-one and onto one simply needs to interpret the second-order quantifiers over M to include the domain of N and conversely. Does this rule out unintended interpretations? Every time we consider a non-standard interpretation this will be readily seen and every time we consider a new structure that satisfies PA_2 to establish the isomorphism anew we simply need to broaden the interpretation of the second-order quantifiers to include the domain of the new structure. Thus, we have a "template" for a solution.[24]

[23]There is an extensive literature on these additional topics and without doing the subject justice I merely wish to add a few remarks.

[24]For more on this subject see [27] and §§48–49 of [28]. Similar remarks carry over to the set theoretic context—see [22].

However, some yearn for greater assurance, something that will rule out all non-standard interpretations in one fell swoop. Because of this one is tempted to pin down the range of possibilities. The most natural way to do this is to articulate the semantics of second-order arithmetic in terms of first-order set theory. The trouble is that if one ties the interpretation of second-order arithmetic to first-order set theory in this way then when first-order set theory sways under non-standard interpretations so will second-order arithmetic.

The dilemma is this: Either one articulates the fullness of the second-order domain by appeal to the full powerset operation or one refrains from doing this and instead relies on our common understanding of the second-order language without mention of set theory. In the first case one will have given a first-order formulation of the second-order theory and so the problem of non-categoricity will resurface. In the second case one might either gesture toward the "template solution" or go so far as to simply "acquiesce in one's mother tongue" (to borrow a famous phrase from Quine ([29], p. 201)). But the "template solution" does not involve full second-order logic in the sense we have been discussing and if the tactic of acquiescence is available in the second-order context then it is also available in the first-order context—one can simply take the language of arithmetic and set theory at face value. In summary it appears to be an illusion that full second-order logic can somehow secure categoricity in an absolute sense that is unavailable through other means.

I would like finally to say something about the so-called plural interpretation of second-order languages. This interpretation is really a translation. It originated with Boolos[25] and involves translating sentences of second-order logic into statements involving plural quantifiers in English. One reason this is thought to be important is that the translations appear not to involve ontological commitment to sets and hence this ontological innocence is thought to transfer to second-order logic. For example, the sentence "some critics are such that they admire only one another" does not appear to involve ontological commitment to a set of critics. Similarly, the sentence "there are some sets none of which is a member of itself and of which each set that is not a member of itself is one" does not appear to involve ontological commitment to the Russell set. However, as Parsons has observed, in these cases the class in question is definable and it is routine to handle definable classes without ontological commitment.[26]

For our purposes the important question is whether the translation supports full second-order logic in contrast, say, to definable second-order logic or any of the versions of second-order logic in the hierarchy we have investigated. What has been provided is a translation of one *language* into another

[25] See his [2] and [3].

[26] For more on the subject see [21] and §13 of [28].

and this does not touch the question of which *logic* is supported. The above examples do not argue for more than definable second-order logic. It would be hard indeed to argue that full second-order logic is vindicated through this translation. Consider the sentence "there are some numbers such that ... " where what follows is a paraphrase of the definition of $0^{\#}$. It is hard to believe that this sentence is ontologically innocent. The truth of this sentence is a substantive mathematical fact about the existence of a real number.

§**5. Closing.** There is a correspondence between the hierarchy of strong logics of first order and the hierarchy of strong logics of second order. At the base of each hierarchy lies the logic based on the *most general* class of test structures—first-order logic and Henkin second-order logic. These logics are Turing equivalent to one another and each is to the complete Σ^0_1 set of integers. They are sound and complete. To climb each hierarchy one narrows the class of test structures—in the first case by constraining the first-order domain to involve richer structures and in the second case by constraining the second-order domain to involve richer powersets. And again there is a correspondence. For example, the minimal step we considered in the first-order setting involved narrowing the test structures to ω-models and the minimal step we considered in the second-order setting involved narrowing the second-order domain by taking the definable powerset. The resulting logics—ω logic and definable second-order logic—are Turing equivalent to each other and each is Turing equivalent to the complete Π^1_1 set of integers. The correspondence continues as one moves to stronger and stronger logics. Staying within the realm of generically invariant and faithful logics we characterized the strongest logics in each hierarchy—generically invariant first-order logic and generically invariant second-order logic. Once again, these logics are Turing equivalent.

One can continue to narrow the collection of test structures to the point where generic invariance is lost. To take two extreme examples, in the first-order setting one can constrain the class of test structures to contain only the rank initial segments V and in the second-order setting on can constrain the class of test structures in such a way that S is the full powerset of M. The result in the first-order case is Π_2-truth in V and the result in the second case is full second order logic. Once again, these "logics" are Turing equivalent, each being Turing equivalent to Π_2-truth in the universe of sets. Here the mathematical entanglement has become so great that generic invariance is lost. In fact, any non-trivial forcing perturbs the "logic". The "logic" is generically fragile.

One might wonder what role the above strong logics could play in a foundational dispute. Assume ZFC + "There is a proper class of Woodin cardinals" and consider Ω-logic (equivalently, generically invariant logic). This logic in this context has the feature that for every statement φ of $L(\mathbb{R})$, either

$$\varnothing \models_\Omega \varphi \text{ or } \varnothing \models_\Omega \neg\varphi.$$

So, one way to *prove* such a statement in Ω-logic is to show that it is *consistent via forcing*.[27] People with radically different viewpoints who nonetheless agree on the background theory will agree on this completeness feature of Ω-logic and they will both have the methods of forcing at their disposal. They can thus settle disputes by using forcing to establish the consistency of various statements. Now compare the situation in second-order logic. In order to agree on all questions of second-order satisfiability the two parties would have to first agree on all questions of Π_2-truth. For further instances of the utility of Ω-logic in foundational disputes see [13], [14], [15], [37], and [38].

The above strong logics are modeled on forcing and one might think that this is a limitation. I should stress once again that I am not arguing for the positive proposal that these logics are to be regarded as having the same status as first-order logic. I am just claiming that they share two central facets of absoluteness that enable them to play a significant role in foundational disputes. These two facets are closely modeled on forcing. Should other independence methods arise—say, a method of "super-forcing"—then it would be of interest to investigate logics sharing other facets of absoluteness. But notice that this would broaden the domain of models across which one would demand invariance, not narrow it. This would not effect the negative case against full second-order logic. For that case the narrower the domain the stronger the case.

The advocate of second-order logic might think that generic invariance is too high a demand and that one should weaken the demand by asking for a narrower domain across which the logic is required to be invariant. But the point is that full second-order logic has *no* degree of absoluteness. We can only agree on full second-order logic when we have narrowed down the possibilities for Π_2-truth about the universe of sets to a *point* and this amounts to narrowing the domain down to V. For this reason it appears that Quine was right to think that full second-order logic is set theory in sheep's clothing.

REFERENCES

[1] JOAN BAGARIA, NEUS CASTELLS, and PAUL LARSON, *An Ω-logic primer*, ***Set theory*** (Joan Bagaria and Stevo Todorcevic, editors), Trends in Mathematics, Birkhäuser, Basel, 2006, pp. 1–28.

[2] GEORGE BOOLOS, *To be is to be a value of a variable (or to be some values of some variables)*, ***The Journal of Philosophy***, vol. 81 (1984), no. 8, pp. 430–449.

[3] ———, *Nominalist platonism*, ***Philosophical Review***, vol. 94 (1985), no. 3, pp. 327–344.

[4] SAMUEL BUSS, ***Bounded arithmetic***, Bibliopolis, Napoli, 1986.

[27] Compare the case of a theory (such as the real closed fields) which is complete in first-order logic. In such a context to show that a statement φ is a theorem it suffices to show that it is consistent with the axioms.

[5] ———, First-order proof theory of arithmetic, In Handbook of Proof Theory [6], 1998, pp. 79–147.

[6] Samuel Buss (editor), ***Handbook of proof theory***, Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1998.

[7] Rudolf Carnap, ***Logische Syntax der Sprache***, Springer, Wien, 1934, translated by A. Smeaton as *The logical syntax of language*, Kegan Paul, London, 1937.

[8] Solomon Feferman, *Arithmetization of metamathematics in a general setting*, ***Fundamenta Mathematicae***, vol. 49 (1960), pp. 35–92.

[9] Qi Feng, Menacham Magidor, and W. Hugh Woodin, *Universally Baire sets of reals*, ***Set theory of the continuum*** (Haim Judah, Winfried Just, and W. Hugh Woodin, editors), Mathematical Sciences Research Institute, vol. 26, Springer-Verlag, Berlin, 1992, pp. 203–242.

[10] Leon Henkin, *Completeness in the theory of types*, ***The Journal of Symbolic Logic***, vol. 15 (1950), no. 2, pp. 81–91.

[11] Joost Johannes Joosten, ***Interpretability formalized***, Ph.D. thesis, Universiteit Utrecht, 2004.

[12] Akihiro Kanamori, ***The higher infinite: Large cardinals in set theory from their beginnings***, second ed., Springer Monographs in Mathematics, Springer, Berlin, 2003.

[13] Peter Koellner, *On the question of absolute undecidability*, ***Philosophia Mathematica***, vol. 14 (2006), no. 2, pp. 153–188.

[14] ———, *Truth in mathematics*: *The question of pluralism*, ***New waves in philosophy of mathematics*** (Otávio Bueno and Øystein Linnebo, editors), Palgrave Macmillan, 2009, forthcoming.

[15] Peter Koellner and W. Hugh Woodin, *Incompatible Ω-complete theories*, ***The Journal of Symbolic Logic***, (2009), forthcoming.

[16] ———, *Large cardinals from determinacy*, ***Handbook of set theory*** (Matthew Foreman and Akihiro Kanamori, editors), Springer, 2009, forthcoming.

[17] Paul B. Larson, ***The stationary tower: Notes on a course by W. Hugh Woodin***, University Lecture Series, vol. 32, American Mathematical Society, 2004.

[18] Richard Laver, *Certain very large cardinals are not created in small forcing extensions*, ***Annals of Pure and Applied Logic***, vol. 149 (2007), no. 1–3, pp. 1–6.

[19] Per Lindström, *A note on weak second order logic with variables for elementarily definable relations*, ***The proceedings of the Bertrand Russel memorial conference*** (***Uldum, 1971***), 1973, pp. 221–233.

[20] ———, ***Aspects of incompleteness***, second ed., Lecture Notes in Logic, vol. 10, Association for Symbolic Logic, 2003.

[21] Øystein Linnebo, *Plural quantification exposed*, ***Noûs***, vol. 37 (2003), no. 1, pp. 71–92.

[22] Donald A. Martin, *Multiple universes of sets and indeterminate truth values*, ***Topoi***, vol. 20 (2001), no. 1, pp. 5–16.

[23] Donald A. Martin and Robert M. Solovay, *A basis theorem for $\sum_3^1$ sets of reals*, ***Annals of Mathematics***, vol. 89 (1969), pp. 138–159.

[24] Franco Montagna and Antonella Mancini, *A minimal predicative set theory*, ***Notre Dame Journal of Formal Logic***, vol. 35 (1994), no. 2, pp. 186–203.

[25] Jan Mycielski, Pavel Pudlák, and Alan S. Stern, *A lattice of chapters of mathematics* (*interpretations between theorems*), ***Memoirs of the American Mathematical Society***, vol. 84 (1990), no. 426, iv+70.

[26] Edward Nelson, ***Predicative arithmetic***, Princeton Mathematical Notes, no. 32, Princeton University Press, 1986.

[27] Charles Parsons, *The uniqueness of the natural numbers*, ***Iyyun***, vol. 39 (1990), pp. 13–44.

[28] ———, ***Mathematical thought and its objects***, Cambridge University Press, 2008.

[29] Willard V. Quine, *Ontological relativity*, ***The Journal of Philosophy***, vol. 65 (1968), no. 7, pp. 185–212.

[30] ———, ***Philosophy of logic***, second ed., Harvard University Press, Cambridge, MA, 1986.

[31] Joseph R. Shoenfield, *The problem of predicativity*, ***Essays on the foundations of mathematics*** (Yehoshua Bar-Hillel, E. I. J. Poznanski, Michael O. Rabin, and Abraham Robinson, editors), Magnes Press, Hebrew University, Jerusalem, 1961, pp. 132–139.

[32] Wanda Szmielew and Alfred Tarski, *Mutual interpretability of some essentially undecidable theories*, ***Proceedings of the International Congress of Mathematicians***, vol. 1, Cambridge, 1950, p. 734.

[33] Jouko Väänänen, *Second-order logic and foundations of mathematics*, this Bulletin, vol. 7 (2001), no. 4, pp. 504–520.

[34] Albert Visser, *An overview of interpretability logic*, ***Advances in modal logic***, CSLI Lecture Notes, vol. 87, CSLI Publishing, Stanford, CA, 1998, vol. 1, pp. 307–359.

[35] ———, *Pairs, sets, and sequences in first-order theories*, ***Archive for Mathematical Logic***, vol. 47 (2008), pp. 299–326.

[36] W. Hugh Woodin, ***The axiom of determinacy, forcing axioms, and the nonstationary ideal***, de Gruyter Series in Logic and its Applications, vol. 1, de Gruyter, Berlin, 1999.

[37] ———, *The continuum hypothesis*, ***Logic Collouqium 2000*** (Rene Cori, Alexander Razborov, Steve Todorĉević, and Carol Wood, editors), Lecture Notes in Logic, vol. 19, Association for Symbolic Logic, 2005, pp. 143–197.

[38] ———, *The continuum hypothesis, the generic-multiverse of sets, and the* Ω*-conjecture*, preprint, 2006.

DEPARTMENT OF PHILOSOPHY
HARVARD UNIVERSITY
25 QUINCY STREET
CAMBRIDGE, MA 02138, USA
E-mail: koellner@fas.harvard.edu