

Abstraction Principles and the Classification of Second-Order Equivalence Relations

Sean C. Ebels-Duggan

Version submitted July 23, 2016.

Please see forthcoming article in
Notre Dame Journal of Formal Logic
for final version

Abstract

This paper improves two existing theorems of interest to neo-logicist philosophers of mathematics. The first is a classification theorem due to Fine for equivalence relations between concepts definable in a well-behaved second-order logic. The improved theorem states that if an equivalence relation E is defined without non-logical vocabulary, then the *bicardinal slice* of any equivalence class—those equinumerous elements of the equivalence class with equinumerous complements—can have one of only three profiles. The improvements to Fine’s theorem allow for an analysis of the well-behaved models had by an abstraction principle, and this in turn leads to an improvement of Walsh and Ebels-Duggan’s relative categoricity theorem.

1 Introduction

Neo-logicist philosophers of mathematics are impressed by the fact that some *abstraction principles* can interpret interesting fragments of mathematics. These principles are sentences of an enriched second-order language of the following form: for E an equivalence relation between second-order objects, and ∂ a functor taking second-order objects to first order objects, the abstraction principle $A_E[\partial]$ is the sentence

$$(\forall X, Y)(\partial X = \partial Y \leftrightarrow E(X, Y)) \quad (1)$$

In virtue of their form, say neo-logicists, abstraction principles are eligible to be “analytic truths”—meaning that they are epistemically near enough to logical truths.¹ And if abstraction principles count, in some sense, as near enough to logical, then so should any mathematics they interpret.

¹See most notably [26] and [10].

So goes the argument, and not without objections and replies. Much turns on what would count as a “(near enough to) logical” abstraction principle. But a plausible minimal requirement is this: an abstraction principle’s equivalence relation must itself be “logical”,² and so cannot require identification of particular objects, concepts, or relations for its definition. The thinking is that logic is indifferent to particulars, whether objects, concepts (monadic second-order objects), or relations (polyadic second-order objects). So for an abstraction principle to be logical, it is necessary that its equivalence relation be accordingly indifferent. A natural way to flesh out this notion of “indifference” is with the notion of *permutation invariance*: the status of the relation doesn’t vary, no matter how one exchanges the objects of its concern.³

Further complicating the neo-logicist’s hopes is the fact that not all abstraction principles—even those based on permutation invariant equivalence relations—*should* count as logical: so not only does said neo-logicist need an account of how an abstraction principle could be logical, but that account must also sort the good principles from the bad. So more needs to be said for what kinds of logical equivalence relations there are on concepts, and which are apt to yield suitably logical abstraction principles. It is thus of interest to classify logical equivalence relations on concepts.

This paper proves a classification theorem for such equivalence relations. Though discovered independently, it happens that the classification theorem here presented is a stronger version one already on the books. Kit Fine’s Theorem 4 of [7, p. 142] classifies infinite concepts in standard models. The theorem is then used to determine the finest abstraction principle satisfiable on all infinite standard models [7, Theorem 6, p. 144].

But Fine’s theorem in its given form hides its true power; hence our new presentation. In our version the result is put in a deductive setting in which cardinalities are well-behaved. This is relatively minor, but it allows us to apply the theorem to problems cast in just such deductive settings (about which more in a moment). More importantly, our version of the theorem sorts equivalence relations, and their abstraction principles, more usefully. In other words: there are abstraction principles well-discussed in the literature on neologicism, but as Fine states the theorem in [7], one needs to squint to see how their equivalence relations are classified. The version here given allows for more clear-eyed recognition of this sorting, and allows for a generalization of the classification to (Dedekind) finite concepts as well.

It is something of a journey from this restatement and expansion of Fine’s theorem to its payoff, but the midpoint is a worthy stop. The improved theorem enables an analysis of when a given abstraction principle has a well-behaved

²We put aside concerns, like those of Quine in [18], that second-order languages are themselves not “logical” in virtue of their quantifying over higher-order objects.

³This view is associated with Tarski’s identification of logical notions as those that are permutation invariant in this sense (see [22]), for a more recent defense see the work of Gila Sher (see [21]). Other relevant discussions can be found in [2] and [16]. The so-called Tarski-Sher thesis identifying logicity with permutation invariance is controversial; that permutation invariance is a necessary feature of logical notions is not (see [15]).

model (not just a standard model). The existence of such models, as we shall see, are determined by the classification of the principle’s equivalence relation on a certain sub-class of concepts—the *bicardinally equivalent* ones. This at least gives the neo-logicist a tool for analyzing the space of abstraction principles.

The terminal payoff comes in the form of a more direct result: the restated and expanded theorem allows an improvement of the relative categoricity theorem given by Walsh and Ebels-Duggan in [24]. Walsh and Ebels-Duggan showed that abstraction principles are *naturally relatively categorical* when and only when their equivalence relation is coarser than that of the neo-logicist’s favored abstraction principle, HP. Using the improved version of Fine’s theorem, we show that this result obtains even when we remove the qualifier “naturally”. This improvement allows for a plausible case to be made that HP and its ilk pass a minimal threshold for logicity, since (unqualified) relative categoricity is arguably the correct notion of permutation invariance for abstraction principles.

The remainder of this paper is organized as follows. Sections 2–4 are introductory, motivating the project and setting in place the formal machinery for the theorems. Section 2 explains why Fine’s theorem could use improvement, and gives an informal characterization of the improved theorem, which we call the “**Main Theorem**”. Section 3 describes the second-order language in which we will work, and its structures; as well as listing our use of standard abbreviations and giving formal clarifications of the terms we have loosely defined above and in section 2. Here also we describe the cardinality assumptions we adopt with slight modification from [24], explaining the sense in which our background logic admits only “well-behaved cardinalities”. The next section 4 explains why restricting our attention to bicardinally equivalent concepts is apt to our purposes.

With preliminaries done, section 5 provides a rigorous statement and proof of the **Main Theorem**. The remaining sections discuss the relevance of the **Main Theorem** to the neo-logicist project. Our aim in these sections will not be to argue for or against a particular version of neo-logicism, but instead to let the **Main Theorem** shed light on some technical questions of interest. Section 6 will address the sorting of so-called “bad companions” by the **Main Theorem**; while section 7 proves and discusses the extension of the relative categoricity theorem of [24]. For ease of exposition, we put off some of the more tedious or repetitive proofs to appendices, we conclude with these.

2 Why bother a sleeping theorem?

The restatement and expansion of Fine’s theorem is more than just a curious exercise, for we believe there is a *better* statement of the theorem than the original. We begin by giving Fine’s version.

To state Fine’s theorem efficiently, we need a bit of notation which we will use in the rest of the paper. Letting “ \cup ”, “ \cap ”, and “ $-$ ” denote the usual boolean functions on sets of union, intersection, and difference, we will use “ $X \triangle Y$ ” to mean the *symmetric difference* between X and Y , namely the set $(X - Y) \cup (Y -$

X). And we'll denote the complement of X in a universe M by $M - X$. Finally, we'll say that X and Y are *bicardinally equivalent* to mean that $|X| = |Y|$ and $|M - X| = |M - Y|$; that is, that X and Y are equinumerous (in a given structure \mathcal{M}) and that their complements are also equinumerous (also in the given structure). We'll abbreviate bicardinal equivalence by writing " $X \sqsubseteq Y$ ". Note that \sqsubseteq is (provably) an equivalence relation in second-order logic.

Now, to Fine's theorem: Let \mathcal{M} be a standard model of second-order logic, X and Y be infinite concepts such that $X \sqsubseteq Y$, and E be a permutation invariant equivalence relation. The concepts X and Y are *representative* just if

$$\mathcal{M} \models (\forall Z, W)(Z, W \sqsubseteq X \wedge E(X, Y) \rightarrow E(Z, W)) \quad (2)$$

Fine's Classification Theorem. Given a standard model \mathcal{M} , infinite concepts X and Y of \mathcal{M} are representative if and only if they meet exactly one of the following conditions.

1. $|X \triangle Y| = |X| < |M|$. In Fine's words, X and Y are "small but very different".
2. $\omega \leq |M - X|, |M - Y| < |X| = |M|$ and $|(M - X) \triangle (M - Y)| = |(M - X) \cup (M - Y)|$. In Fine's words this is that X and Y "are almost universal but with infinite very different complements."
3. $|X| = |M - X| = |M| = |M - Y| = |Y|$, and $|X \triangle Y| = |X \triangle (M - Y)| = |X|$ or vice versa. In Fine's words, X and Y "are bifurcatory with one very different from the other and from its complement."

This statement of Fine's theorem is somewhat dizzying; Fine's statement is more succinct since he uses definitions we have spelled out. But the complexity of the theorem is not by itself a count against it. The reason it can be improved, however, is that its classes do not obviously organize the variations of equivalence relations at play in the investigation of abstraction principles. The above statement of Fine's theorem also obscures a generalization of the theorem: in fact there is a general version of Fine's theorem that applies to finite, as well as infinite, concepts. It is hard to see how this could be given the classification Fine offers.

Restating Fine's theorem will make it easier to use, and highlights the connections between the resultant classification and abstraction principles of particular interest to neo-logicians. (Unfortunately, restating and expanding the theorem to the finite case requires proving it anew. One would hope to rely on Fine's proof, but in fact once the work is done for the extension, the initial theorem is all but proved.)

To improve on Fine's classification theorem, we start by thinking about three types of abstraction principles and their similarities.

We noted in the Introduction [1](#) that Frege's logicist project was to show that arithmetic is in some sense "really" logic by showing that arithmetic laws are in fact logical laws. Frege executes his program in two steps: first by proving in

second-order logic (without our cardinal assumptions) that BLV,

$$(\forall X, Y)(\varepsilon X = \varepsilon Y \leftrightarrow X = Y) \quad (3)$$

implies HP:

$$(\forall X, Y)(\#X = \#Y \leftrightarrow X \approx Y) \quad (4)$$

Here “=” indicates co-extensiveness of concepts, and “ \approx ” indicates equinumerosity, the existence of a bijection between concepts. Note that in the notation of (1), BLV is $A_=_$ and HP is A_{\approx} .

Frege then shows that in second-order logic (with full comprehension), HP interprets (what is now called) second-order Peano Arithmetic. In fact, relatives of HP will do the same thing,⁴ including the *bicardinality principle*, BP, which is A_{\sqsubseteq} . So if BLV is logical, then so are HP and BP, and thus so is arithmetic.

But BLV is no principle of logic, since it is quite famously inconsistent in second-order logic, deriving the Russell paradox. Neo-logicism proceeds against Frege’s objections in [8, §63ff], asserting arithmetic is logical from the second part of the program alone—that is, that as HP is plausibly logical, so then is arithmetic.⁵

At a minimum, abstraction principles must be consistent to count as logical. But that this minimum is not enough is the “bad company” problem: there are consistent abstraction principles that are otherwise unacceptable, at least to the neo-logicist. The prototypical bad companion is NP:⁶

$$(\forall X, Y)(\eta X = \eta Y \leftrightarrow |X \triangle Y| < \omega) \quad (5)$$

The principle NP is unwelcome to the neo-logicist because in contexts of well-behaved cardinalities, it implies that the universe is Dedekind finite.⁷

Another bad companion, discussed in [24, § 5.5], is the Complementation Principle, CP,

$$\begin{aligned} (\forall X, Y)(\odot X = \odot Y \leftrightarrow \\ [(|X| = |Y| = |M - X| \wedge (X = Y \vee X = M - Y)) \vee \\ (|X| \neq |M - X| \wedge |Y| \neq |M - Y|)]) \end{aligned}$$

This principle sorts only concepts the same size as their complements. Such concepts are grouped together with only their complements, and all other concepts are grouped in a “junk” equivalence class. As we will show below, this principle counts as “bad company” as well (see section 6.3).

⁴Interpreting Peano Arithmetic does not require the full power of HP, but only that an abstraction principle match HP on all finite concepts. See [12], and [14] for further discussion.

⁵Wright’s [26] marks the starting point of neo-logicism, though the observation that Frege’s program proceeds in two parts was made in [17, p. 183, p. 194].

⁶The first bad companions appeared in [3]; one of these was simplified into NP by Wright in [27].

⁷And this is unwelcome to neo-logicists because they claim an abstraction principle shouldn’t imply anything about sortal concepts unrelated to the abstracts; see [9, p. 415], [27, pp. 295–7] and [28, pp. 314–5], as well as [6].

Taking stock, we have so far two good abstraction principles in HP and BP, and three bad ones (BLV, NP, and CP). Can we sort these into useful classes? We can if, like Fine, we look at their behavior just on bicardinally equivalent concepts. Clearly, when restricted to just bicardinally equivalent concepts, the equivalence relations of HP (equinumerosity) and BP are *trivial*: they include all such concepts.

Second, BLV and NP are, in a sense, of the same type: the equivalence relations for both are concerned with the size of the symmetric difference between the two related concepts. The principle BLV discriminates concepts if their symmetric difference has non-zero size. Likewise NP discriminates if the symmetric difference is Dedekind infinite. Say such equivalence relations that sort concepts according to the size of their symmetric difference, are *separations*: two concepts with “few” objects falling under one but not both of the two concepts are equivalent; two concepts with “many” such objects are not equivalent.

Lastly, we have CP. Such an equivalence relation is neither trivial, nor a separation, but it has many of the drawbacks of separations. And such an equivalence relation can be generalized, as NP is a generalization of BLV: two bicardinally equivalent concepts can be grouped together if their symmetric difference is “small”, or the complement of their symmetric difference is “small”. So let us say that on bicardinally equivalent concepts, such relations are *complementations*: they sort concepts by the size of their symmetric difference, or the size of the complements of their symmetric difference.

Finally, say that an equivalence relation E is a *refinement* of an equivalence relation E' if whenever two concepts are E -equivalent, they are E' equivalent.⁸

We can now state the informal (though somewhat inexact) version of our main theorem:

(Informal) Main Theorem. Let E be a purely logical equivalence relation, and let our background logic be a strong but natural version of second-order logic. If we look only at E on bicardinally equivalent concepts, then E is either the trivial equivalence relation, or it is a refinement of a separation, or it is a refinement of a complementation.

This version of Fine’s result makes more obvious the relationship between its classification scheme and the kinds of abstraction principles that have been seen in the neo-logicist’s laboratory. In restating the theorem we observe a striking alignment: the problematic abstraction principles arise from *non-trivial* equivalence relations. This is no accident; as we will see in section 6, abstraction principles involving non-trivial equivalence relations limit the size of their models in just the way that is typical of bad company.

⁸See [7] and [1] for discussion of “finer” and “coarser” equivalence relations and their relevance to neo-logicism. Cook’s [5] standardizes the terminology and indicates relations between different kinds of invariance. It was Cook’s paper that made me see the connection to Fine’s work.

3 The language L_0 , comprehension, and cardinality assumptions

Our background logic is a basic, though robust, version of second-order logic, adopted with slight weakening from §2 of [24]. In short, the language L_0 uses lower-case letters to vary over first-order objects, and upper-case letters to vary over second-order objects of all finite arities (arity will be clear from context in our presentation)—such objects of singular arity are called “concepts” or “sets”, others are called “relations”. All terms of L_0 are variables; we exclude constants of either type. Formulae of L_0 and their interpretation are as usual.

Thus models of this language are of the form

$$\mathcal{M} = (M, S_1[M], S_2[M], \dots) \quad (6)$$

where M is non-empty and the members of $S_i[M]$ are subsets of M^i for $i \geq 1$. We do not require $S_i[M]$ to be the full power-set of M , for compatibility of our results with those of [24]; that is, we work in the non-standard semantics. Likewise we require our models to satisfy comprehension axioms for all formulae in their signature; these axioms are of the form

$$(\exists X)(\forall y)(Xy \leftrightarrow \Phi(y, \bar{p}, \bar{R})) \quad (7)$$

where \bar{p} is a sequence of parameters from M , and \bar{R} is a sequence of relation parameters from $\bigcup_{i \in \mathbb{N}} S_i[M]$. This ensures that all finite concepts, and many more besides, are included in $S_1[M]$. The remaining axioms of our logic, which we will call “second-order logic” are those of D2 found in [19, pp. 65–7], but excluding the axiom of choice. We will soon be augmenting this logic with choice principles to ensure well-behaved cardinalities.

Most symbols used are standard; we rehearse a few: The symbol \sqcup , which appears in Proposition 8, means the disjoint union; we will at times abuse notation using it to mean the union of two disjoint sets. As usual we use “ \preceq ” to assert the existence of an injection, and “ \approx ” the existence of a bijection. We will also use expressions like “ $|X| \leq |Y|$ ” and “ $|X| = |Y|$ ”, and we will use the convention of writing, e.g., $X \overset{f}{\preceq} Y$ to mean that f is an injection from X into Y ; likewise with the other expressions. As above the expression “ \boxminus ” indicates the relation of *bicardinality*; the expression “ $X \overset{f}{\boxminus} Y$ ” means that f is a bijection from M to itself and $f(X) = Y$ and $f(M - X) = M - Y$. Both \approx and \boxminus are provably equivalence relations in second-order logic; we will say in the obvious circumstances that concepts are cardinally, or bicardinally, equivalent.

We identify concepts as infinite *in* \mathcal{M} if they are Dedekind infinite, that is, that there is in \mathcal{M} an injection from the concept into a proper subconcept of itself, and we write this with the expression $|X| \geq \omega$. We say that X is finite in \mathcal{M} if it is not Dedekind infinite, writing it $|X| < \omega$. These expressions are used exclusively within the model \mathcal{M} . When we wish to say that a set X is finite or infinite in the metatheory, we will either say so explicitly, or use “ $|X| = n$ for some $n \in \mathbb{N}$ ” or “ $|X| < |\mathbb{N}|$ ”.

We will abuse notation and use “ M ” for both the first order domain of \mathcal{M} and the universal concept $\{x \mid x = x\}$; and we’ll use functional notation for relations that are functional in \mathcal{M} . In general we will write $f \in \mathcal{M}$ to mean that for some $i \in \mathbb{N}$, f is an i -ary function and $f \in S_{i+1}[M]$; similarly with concepts and relations. We write $f(X)$ to mean the image of X under f ; such a relation is always in \mathcal{M} by comprehension.

Though L_0 is expressively rich, it is not too rich, for (crucially for our results), any function $\pi : M \rightarrow M$ that extends to permute objects *of all types in* \mathcal{M} is an automorphism of \mathcal{M} . This can be seen by the fact that, for a given \mathcal{M} and $\pi : M \rightarrow M$, the “push model of \mathcal{M} under π ” is always isomorphic to \mathcal{M} , with π the witnessing isomorphism.⁹ If π is also such as to make \mathcal{M} the same structure as its “push model”, then such a π is an automorphism. By comprehension, if $\pi : M \rightarrow M$ is a bijection in \mathcal{M} , then π meets the needed criteria. Thus, we may state the relevant result in the following form for our use:

Permutation Invariance PI. If $\pi : M \rightarrow M \in \mathcal{M}$ is bijection and $\Phi(\overline{x}, \overline{X})$ is an L_0 -formula, then

$$\mathcal{M} \models (\forall \overline{x}, \overline{X})(\Phi(\overline{x}, \overline{X}) \leftrightarrow \Phi(\overline{\pi(x)}, \overline{\pi(X)})) \quad (8)$$

Under the same hypotheses about π , if $E(X, Y)$ is an L_0 -definable equivalence relation on concepts of M , then

$$\mathcal{M} \models E(X, Y) \leftrightarrow E(\pi(X), \pi(Y)) \quad (9)$$

Here, of course, E being L_0 -definable means definable *without* parameters: the formula defining E has only the second-order variables X and Y free. Definability via an L_0 -formula is the technical correlate of the informal “purely logical” used in Section 1. Our main results depend on **PI**, in the sense that they obtain for any equivalence relation that is permutation invariant, not just those L_0 -definable.

The second way our background logic is robust is that we require that the relation of equinumerosity behave as it does in more familiar contexts, e.g., ZFC. In particular, we require the following to obtain in all models \mathcal{M} under consideration:

Cardinal Comparability CC.

$$\mathcal{M} \models (\forall X, Y)(X \leq Y \vee Y \leq X) \quad (10)$$

Infinite Sums are Maxima ISM.

$$\mathcal{M} \models (\forall X, Y)(|X|, |Y| \geq \omega \rightarrow |X \sqcup Y| = \max(|X|, |Y|)) \quad (11)$$

Infinite Products are Maxima IPM.

$$\mathcal{M} \models (\forall X, Y)(|X|, |Y| \geq \omega \rightarrow |X \times Y| = \max(|X|, |Y|)) \quad (12)$$

⁹See [4, pp. 225–31] for more on this method.

Cardinalities are Well-Founded CWF. For any L_0 -formula $\Phi(X)$,

$$\mathcal{M} \models (\exists X)(\Phi(X)) \rightarrow (\exists X)(\forall Y)(\Phi(Y) \rightarrow X \preceq Y) \quad (13)$$

We take over these principles nearly directly from [24]; both for their handiness and as we will apply the **Main Theorem** to offer a solution to some open questions from that paper (see section 7).¹⁰ Thus, in what follows and unless otherwise noted, all structures satisfy full comprehension and these cardinality assumptions.

The principles **CC** and **ISM** are ubiquitous in the proof of our first version of the **Main Theorem**, Theorem 6. The schematic principle **CWF** is used in moving from Theorem 6 to the second formal version of the **Main Theorem**, Theorem 7. The principle **IPM** is deployed mainly in the form of a pairing function in section 7.

We will need specific notation suited to dealing with the divergence, in non-standard semantics, between Dedekind finitude and finitude in the metatheory. As this notation pertains only to the proof of the formal version of the **Main Theorem**, we introduce it in section 5.

4 Bicardinal equivalence and permutation invariance

Our **Main Theorem** says that an L_0 -definable equivalence relation, when restricted to bicardinally equivalent concepts, can have one of only three profiles. But why is it interesting to look at *this* particular restriction of such equivalence relations? This portion of our paper will motivate this restriction.

Our interest in L_0 -definable equivalence relations is that, provided that second-order languages are a part of logic, these relations meet at least one criterion of logicity: they are permutation invariant. But one interesting fact about second-order logic is that there are L_0 -definable properties that can *distinguish* second-order objects. To see what this means, consider the formula

$$Singleton(X) : (\exists x)(\forall y)(Xy \leftrightarrow y = x) \quad (14)$$

The formula $Singleton(X)$ can distinguish between second-order objects in sufficiently rich structures for second-order languages: if a structure includes distinct first-order objects a and b , then in that structure $Singleton(\{a\})$ while $\neg Singleton(\{a, b\})$. These second-order objects are thus distinguished by the formula “ $Singleton(X)$ ”. Thus, while first-order objects are indistinguishable using only L_0 -definable notions, second-order objects are not. It makes sense, then, that we would want to pay special attention to collections of second-order

¹⁰In [24] the principles **CC**, **ISM**, and **IPM** were deployed as consequences of the principle **GC**, though **GC**’s well-ordering was in the main results only to show that the restrictions to small concepts and to abstracts were required (see [24, equations 4.22–4.23 and following]). In this paper we will use only the cardinal consequences of **GC** listed above. Our logic is also weaker than that deployed in [24] in that we make no use of the principle **AC**.

objects that are *not* distinguishable in this way. And we need not look far for concepts that are so indistinguishable. As can be seen from **PI**, concepts are bicardinally equivalent if and only if they cannot be distinguished using an L_0 -formula. It is thus of interest to see how purely defined equivalence relations behave on these concepts in particular.

Relatedly, bicardinal equivalence is *indicative* of permutations. Let E be an equivalence relation on concepts. We will say that E is an *indicator of permutations* just if, for any second-order structure \mathcal{M} ,¹¹ and any function $f : M \rightarrow M$ such that if $X \in \mathcal{M}$ then $f(X) \in \mathcal{M}$, if

$$(\forall X, Y)(E(f(X), f(Y)) \rightarrow E(X, Y)) \quad (15)$$

in that structure, then f is a permutation (it is a bijection from the first-order domain M to itself).

Theorem 1. Bicardinality is an indicator of permutations.

Proof. Observe that if f is not injective, then (15) fails: let x, y be such that $f(x) = f(y)$, and note that $\neg(\{x, y\} \sqsubseteq \{f(x)\}) \wedge f(\{x, y\}) = \{f(x)\} \sqsubseteq \{ff(x)\}$. So it remains only to show that if f is injective and (15) holds in \mathcal{M} , then f is surjective.

Further, if $|M|$ is finite in \mathcal{M} , then every injection on M is a permutation. So we may further assume that M is not finite in \mathcal{M} .

Working in \mathcal{M} , we show first that

$$\text{if } |X| = |M| \text{ then } X - f(X) \sqsubseteq f(X - f(X)) \quad (16)$$

That the concepts are equinumerous follows from the fact the f is injective. To see that their complements are equinumerous, we have that:

$$|M| \geq |M - (X - f(X))| \quad (17)$$

$$= |(M - X) \cup f(X)| \geq |f(X)| = |M| \quad (18)$$

where the last equality follows again by the injectivity of f . By the Schröder-Bernstein theorem, $|M - (X - f(X))| = |M|$.¹² An identical argument shows that $|M - f(X - f(X))| = |M|$.

Thus by (16) we have

$$M - f(M) \sqsubseteq f(M - f(M)) \quad (19)$$

Since \sqsubseteq is preserved under complementation,

$$f(M) \sqsubseteq M - f(M - f(M)) \quad (20)$$

which implies that

$$|M| = |f(M)| = |M - f(M - f(M))| \quad (21)$$

¹¹Not just those with well-behaved cardinalities.

¹²Note this is provable in unaugmented second-order logic, see [19, Theorem 5.2, p. 102ff].

So by (16) together with (20) and (21) we obtain that

$$f(M) \sqsubseteq M - f(M - f(M)) \sqsubseteq f(M - f(M - f(M))) \quad (22)$$

By assumption we have (15), and so $M \sqsubseteq M - f(M - f(M))$ from (22). Complementation then gives that $\emptyset \sqsubseteq f(M - f(M))$, which means that $M - f(M) = \emptyset$, and so $f(M) = M$. Thus f is surjective, and so a permutation. \square

Bicardinally equivalent concepts are thus interesting in their own right. So it makes sense to observe the behavior of purely defined equivalence relations on just concepts that are \sqsubseteq -equivalent.

For this reason we introduce the following notation. As is customary, given an equivalence relation E and a concept X we denote the E -equivalence class containing X by $[X]_E$. Looking at how E behaves on the sets \sqsubseteq -equivalent to X , we say that the *bicardinal slice* of E is the set of equivalence classes of \sqsubseteq -equivalent sets. More formally,

Definition 2. Let \mathcal{M} be a model and E an equivalence relation on sets of \mathcal{M} . Given a set X in \mathcal{M} , we let

$$E(\sqsubseteq)(X, Y) \Leftrightarrow X \sqsubseteq Y \wedge E(X, Y) \quad (23)$$

so that

$$[X]_{E(\sqsubseteq)} = [X]_{\sqsubseteq} \cap [X]_E \quad (24)$$

Thinking of E as a collection of concepts, we can then write

$$E(\sqsubseteq) = \{[X]_{E(\sqsubseteq)} \mid X \in \mathcal{M}\} \quad (25)$$

We can then think of E as consisting in *bicardinal slices* of the form

$$E(\sqsubseteq)_X = \{[Y]_{E(\sqsubseteq)} \mid X \sqsubseteq Y\} \quad (26)$$

since

$$E(\sqsubseteq) = \bigcup_{X \in \mathcal{M}} E(\sqsubseteq)_X \quad (27)$$

These notions will be used in stating Theorem 7, the formal version of the **Main Theorem**.

5 The bicardinal classification of equivalence relations

We are now nearly in a position to state our main results formally. Our theorem will classify equivalence relations at bicardinal slices. We first must formally state the classes into which any relation can be sorted.

Theorem 7 is the formal version of the **Main Theorem**, which we prove in two stages. The intermediate stop is Theorem 6, the remaining step is deploying the cardinality assumption **CWF** to obtain the final result.

Proving these theorems is onerous because non-standard models allow that concepts finite according to \mathcal{M} might not be finite in the metatheory. This distinction is important because the proofs of our main results depend on Lemmata 22 and 24, which deploy, in the metatheory, finitely many permutations that “move” one concept onto another while preserving equivalence classes. Thus, we will at times wish to indicate for given sets X and Y that finitely many copies of Y are enough to cover X . Thus we write $n \times |Y| \geq |X|$ to mean that for some set Z with $|Z| = n \in \mathbb{N}$, $\mathcal{M} \models |Z \times Y| \geq |X|$. Similar expressions will be used in an associated way so that their meaning is obvious.

However, since finiteness in \mathcal{M} is not the same as finiteness in the metatheory, we cannot count on expected relations of cardinality obtaining. In particular, in \mathcal{M} we have the expected Archimedean property in that

$$(\forall X, Y)(|X|, |Y| < \omega \rightarrow (\exists Z)(|Z| < \omega \wedge |Z \times Y| \geq |X|)) \quad (28)$$

is a theorem of second-order logic. But in a non-standard model it is not in general true that for any $X, Y \in S_1[M]$ with $\mathcal{M} \models |X|, |Y| < \omega$, there is an $n \in \mathbb{N}$ such that $\mathcal{M} \models n \times |Y| \geq |X|$. For this reason we introduce the following definition:

Definition 3. Given $X, Y \in S_i[M]$, we write $|X| \trianglelefteq |Y|$ to mean there is an $n \in \mathbb{N}$ such that

$$\mathcal{M} \models n \times |Y| \geq |X| \quad (29)$$

Further, we write $|X| \triangleleft |Y|$ to mean $|X| \trianglelefteq |Y|$ and $\neg(|Y| \trianglelefteq |X|)$. Note that \trianglelefteq and \triangleleft are relations *in the metatheory*; they are not, in general, expressible in L_0 .

In what follows, all concepts are understood to be concepts in given structures \mathcal{M} . In accordance with the two steps towards our main result, we give two definitions for the purposes of classification.

Definition 4. Let E be an L_0 -definable equivalence relation over concepts of a given structure \mathcal{M} . For any concept X , we say that $E(\boxminus)_X$

1. is *trivial* if $[X]_{E(\boxminus)} = [X]_{\boxminus}$.
2. is *separative*, or *refines a separation*, if for all $Y, Z \in [X]_{\boxminus}$,

$$E(Y, Z) \Rightarrow |Y \triangle Z| \triangleleft |X| \quad (30)$$

3. is *complementative*, or *refines a complementation*, if for all $Y, Z \in [X]_{\boxminus}$,

$$E(Y, Z) \Rightarrow (|Y \triangle Z| \triangleleft |X| \text{ or } |M - (Y \triangle Z)| \triangleleft |X|) \quad (31)$$

Obviously if $E(\boxminus)_X$ is separative then it is complementative; therefore we say that $E(\boxminus)_X$ is *properly separative* or *refines a separation* if it is separative and non-trivial, and *properly complementative* or *refines a complementation* if it is complementative, non-separative, and non-trivial.

Remark 5. If X is properly complementative and $Y \in [X]_{E(\boxminus)}$, then $M - Y \in [X]_{E(\boxminus)}$.

We now state the intermediate and final results as:

Theorem 6. Let E be an L_0 -definable equivalence relation over a given structure \mathcal{M} . Then for any concept X , $E(\boxminus)_X$ is either trivial, separative, or complementative.

If $|M| > 2$ then exactly one of these options holds.

If X is finite in \mathcal{M} and $E(\boxminus)_X$ is nontrivial, then either

$$\mathcal{M} \models E(Y, Z) \Leftrightarrow |Y \triangle Z| \triangleleft |X| \quad (32)$$

or

$$\mathcal{M} \models E(Y, Z) \Leftrightarrow [|Y \triangle Z| \triangleleft |X| \text{ or } |M - (Y \triangle Z)| \triangleleft |X|] \quad (33)$$

Theorem 7 (Formal Main Theorem). Let E be an L_0 -definable equivalence relation over a given structure \mathcal{M} . Then for any concept X , $E(\boxminus)_X$ is either trivial, separative, or complementative.

If $|M| > 2$ then exactly one of these options holds.

Further, if $\mathcal{M} \models |X| < \omega$ then $[X]_{E(\boxminus)}$ is either $[X]_{\boxminus}$, or $\{X\}$, or $\{X, M - X\}$.

As we have said above, this is a strong version of Fine’s theorem—as one can see with the aid of a few pages of Venn Diagrams.

5.1 Non-Archimedean arithmetic

The proof of Theorems 6 and 7 will require, for a given X , the use of finitely (in the metatheory) many permutations, each of which fixes some member of $[X]_{E(\boxminus)}$. These permutations will “shuttle” portions of a given set Z onto its image under a bijection $f : Z \rightarrow X$.

We will deploy the relations \trianglelefteq and \triangleleft in the next section to prove Lemma 22, which is crucial to proving Theorem 7 in its full generality. For this we will need to show that certain expected “arithmetic” relations hold. Though tedious to state, we provide them as follows.

Proposition 8. The following hold for all $X, Y, Z, W \in \mathcal{M}$:

$$|X| \leq |Y| \text{ or } |Y| \leq |X| \quad (34)$$

$$|X| \leq |Y| \text{ and } |Z| \leq |W| \Rightarrow |X \sqcup Z| \leq |Y \sqcup W| \quad (35)$$

$$(|Z| = |W| \text{ and } |X \sqcup Z| \triangleleft |Y \sqcup W| \Rightarrow |X| \triangleleft |Y|) \quad (36)$$

$$|X| \leq |Y| \leq |Z| \Rightarrow |X| \leq |Z| \quad (37)$$

$$|X| \leq |Y \sqcup Z| \Rightarrow (|X| \leq |Y| \text{ or } |X| \leq |Z|) \quad (38)$$

$$|X| = |Y \sqcup Z| \Rightarrow |X| \leq |Y| \text{ or } |X| \leq |Z| \quad (39)$$

$$|X| \triangleleft |Y \sqcup Z| \Rightarrow (|X| \triangleleft |Y| \text{ or } |X| \triangleleft |Z|) \quad (40)$$

$$|X| \triangleleft |X \sqcup Y| \Rightarrow |X| \triangleleft |Y| \quad (41)$$

$$|Y \sqcup Z| \leq |X| \Rightarrow |Y| \leq |X| \text{ and } |Z| \leq |X| \quad (42)$$

$$|Y \sqcup Z| \triangleleft |X| \Rightarrow |Y| \triangleleft |X| \text{ and } |Z| \triangleleft |X| \quad (43)$$

$$|X| \triangleleft |Y| \Rightarrow |X| \triangleleft |Y \sqcup Z| \quad (44)$$

$$|X \sqcup Y| \leq |Z| \text{ and } |W| = |Y| \Rightarrow |X \sqcup W| \leq |Z| \quad (45)$$

$$|X| \leq |Y \sqcup Z| \text{ and } |W| = |Y| \Rightarrow |X| \leq |W \sqcup Z| \quad (46)$$

$$|X \sqcup Y| \triangleleft |Z| \text{ and } |W| = |Y| \Rightarrow |X \sqcup W| \triangleleft |Z| \quad (47)$$

$$|X| \triangleleft |Y \sqcup Z| \text{ and } |W| = |Y| \Rightarrow |X| \triangleleft |W \sqcup Z| \quad (48)$$

Their proofs are even more tedious, so we relegate them to Appendix 9.

The following two propositions relate infinity, in both its meta- and intra-theoretic senses, to the relation \triangleleft , and are easy to prove using (34) and **IPM**.

Proposition 9. For all $X, Y \in \mathcal{M}$,

$$|X| \triangleleft |Y| \Leftrightarrow \text{for all } n \in \omega, \mathcal{M} \models n \times |X| < |Y| \quad (49)$$

Proposition 10. If $\mathcal{M} \models |X| \geq \omega$, then

$$|Y| \triangleleft |X| \Leftrightarrow \mathcal{M} \models |Y| < |X|$$

5.2 Preparatory results

Our route to proving Theorem 7 is via Lemma 22, which is proved in the next section. This lemma specifies sufficient conditions for when $E(\boxminus)_X$ is trivial. The main way that we prove Lemma 22 is by exploiting the following consequence of **PI**:

Proposition 11. Let $\pi \in \mathcal{M}$ be a permutation of M such that π fixes X . Then for any $Y \in [X]_E$, $\pi(Y) \in [X]_E$.

For by **PI**, $Y \in [X]_E \Leftrightarrow \pi(Y) \in [\pi(X)]_E = [X]_E$, since π fixes X . Intuitively, the idea is that we can show sets X and Y to be E -equivalent if X is equivalent to some Z , and there is a permutation fixing X but sending Z to Y .

We introduce the following notation to shorten our exposition.

Definition 12. Let $f : X \rightarrow Y$ be a bijection with X and Y disjoint. Then we say that the *permutation induced by f* , $\iota(f) : M \rightarrow M$, is the function defined by

$$\iota(f)(x) = \begin{cases} f(x) & x \in X \\ f^{-1}(x) & x \in Y \\ x & \text{otherwise} \end{cases}$$

Notice that if $f \in \mathcal{M}$ then $\iota(f) \in \mathcal{M}$ is a well-defined bijection.

Our goal in this subsection is to prove Lemmata 13, 14, 15, and 16, the proofs of which exploit Proposition 11. Lemmata 13 and 14 are those with the most involved proofs. In the remainder of this section, where context makes subscripts unnecessary we will use “[X]” to indicate $[X]_E$.

Lemma 13. Suppose there is a $Y \neq X$ with $Y \in [X]$ and let $Z \subset Y - X$ with $\mathcal{M} \models Z \stackrel{f}{\preceq} X - Y$. Suppose further that at least one of the following obtains:

$$|Z| \leq |X \cap Y| \quad (50)$$

$$|Z| \leq |M - (X \cup Y)| \quad (51)$$

Then there is a permutation $\pi \in \mathcal{M}$ such that $[\pi X] = [X]$ and $Z \subset \pi(X - Y) \subset \pi(X)$, and

$$\pi(x) \neq x \Leftrightarrow x \in Z \cup f(Z) \quad (52)$$

Proof. The proof is essentially the same whether (50) or (51) obtains. We set W to be $X \cap Y$ if the former obtains, and $M - (X \cup Y)$ if not. We reason in \mathcal{M} .

As in the proposition, let $Z \stackrel{f}{\preceq} X - Y$, and let $Z \stackrel{g}{\preceq} n \times W$. The function g induces for each $j < n$ a partial function $h_j : W \rightarrow Z$ defined by $h_j(w) = g^{-1}(j, w)$. Note that each h_j is injective from its domain to its range, which are disjoint; so $\iota(h_j)$ is well-defined for each $j < n$.

We define for each j three permutations, as follows:

$$p_j = \iota(h_j) \quad (53)$$

$$q_j = \iota(f \upharpoonright \text{rng}(h_j)) \quad (54)$$

$$r_j(x) = \begin{cases} q_j(p_j(x)) & x \in \text{dom}(h_j) \\ p_j(q_j(x)) & x \in f(\text{rng}(h_j)) \\ x & \text{otherwise} \end{cases} \quad (55)$$

So p_j switches $\text{rng}(h_j)$ and $\text{dom}(h_j)$, q_j switches $\text{rng}(h_j)$ and $f(\text{rng}(h_j))$, and r_j switches $f(\text{rng}(h_j))$ and $\text{dom}(h_j)$.

Note that p_j fixes Y , q_j fixes $p_j(X)$, and r_j fixes $q_j(p_j(Y))$; thus three applications of Proposition 11 yield that for $s_j = p_j \circ q_j \circ r_j$,

$$s_j(X), s_j(Y) \in [s_j(X)] = [X] \quad (56)$$

Note further that for $w \in \text{dom}(h_j)$,

$$s_j(w) = r_j(q_j(p_j(w))) = q_j(p_j(q_j(p_j(w)))) = p_j(p_j(w)) = w \quad (57)$$

where the third equality is due to the fact that $q_j \upharpoonright \text{dom}(h_j)$ is the identity function. It follows from this and the construction of p_j, q_j , and r_j that if $s_j(x) \neq x$, then $x \in \text{rng}(h_j) \cup f(\text{rng}(h_j))$. Conversely, if $x \in \text{rng}(h_j) \subset Y - X$, then $s_j(x) \in X - Y$, and if $x \in f(\text{rng}(h_j)) \subset X - Y$, then $s_j(x) \in Y - X$. Thus if $x \in \text{rng}(h_j) \cup f(\text{rng}(h_j))$ then $s_j(x) \neq x$. So we have established

$$s_j(x) \neq x \Leftrightarrow x \in \text{rng}(h_j) \cup f(\text{rng}(h_j)) \quad (58)$$

Set $\pi = s_0 \circ \dots \circ s_{n-1}$, and note that $Z = \bigcup_{j < n} \text{rng}(h_j)$. So we have

$$\pi(X - Y) = s_0 \circ \dots \circ s_{n-1}(X - Y) \quad (59)$$

$$= s_0 \circ \dots \circ s_{n-1}(((X - Y) - f(Z)) \cup f(Z)) \quad (60)$$

$$= ((X - Y) - f(Z)) \cup s_0 \circ \dots \circ s_{n-1}(f(Z)) \quad (61)$$

$$= ((X - Y) - f(Z)) \cup s_0 \circ \dots \circ s_{n-1} \left(f \left(\bigcup_{j < n} \text{rng}(h_j) \right) \right) \quad (62)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} s_j(f(\text{rng}(h_j))) \quad (63)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} r_j(q_j(p_j(f(\text{rng}(h_j))))) \quad (64)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} r_j(q_j(f(\text{rng}(h_j)))) \quad (65)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} r_j(f^{-1}(f(\text{rng}(h_j)))) \quad (66)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} r_j(\text{rng}(h_j)) \quad (67)$$

$$= ((X - Y) - f(Z)) \cup \bigcup_{j < n} \text{rng}(h_j) \quad (68)$$

$$= ((X - Y) - f(Z)) \cup Z \quad (69)$$

Here (61) follows by (58), as does (63). Equations (65, 66, and 67) follow respectively from (53, 54, and 55).

Thus $Z \subset \pi(X - Y) \subset \pi(X) \in [X]$ by (56); moreover, (58) implies (52). \square

Lemma 14. Suppose there is a $Y \neq X$ with $Y \in [X]$, and let $Z \subset M - (X \cup Y)$ with $Z \stackrel{f}{\preceq} X \cap Y$. Suppose further that one of the following obtains:

$$|Z| \trianglelefteq |X - Y| \quad (70)$$

$$|Z| \trianglelefteq |Y - X| \quad (71)$$

Then there is a permutation $\pi \in \mathcal{M}$ such that $[\pi X] = [X]$, $Z \subset \pi(X \cap Y)$, and satisfying (52).

As the idea for the proof of Lemma 14 is essentially the same as that of Lemma 13, we omit it.

Now for two easy lemmata:

Lemma 15. Suppose $Z \subset M - (X \cup Y)$, $Z \stackrel{f}{\preceq} X - Y$ with $Y \in [X]$ and $Y \neq X$. Then there is a permutation $\pi \in \mathcal{M}$ such that $[\pi X] = [X]$, $Z \subset \pi(X - Y)$, satisfying (52).

Lemma 16. Suppose $Z \subset X - Y$, $Z \stackrel{f}{\preceq} X \cap Y$ with $Y \in [X]$ and $Y \neq X$. Then there is a permutation $\pi \in \mathcal{M}$ such that $[\pi X] = [X]$, $Z \subset \pi(X \cap Y)$, and satisfying (52).

Proof of Lemmata 15 and 16. For Proposition 15, note that $\pi = \iota(f \upharpoonright Z)$ fixes Y , so $[X] = [\pi(X)]$ by Proposition 11. The rest is obvious and routine.

For Lemma 16, proceed similarly but set $\pi = \iota(f \upharpoonright Z)$. \square

5.3 Almost complementarity and symmetry

The results of the previous section showed what functions are needed in order to, speaking loosely, “transform one set into another”, while staying in the same equivalence class. In this section we define properties that determine sufficient conditions for the application of those results. In particular, we establish Lemma 22, which says that as long as these properties obtain for a set of a given cardinality, then all sets of that cardinality are E -equivalent.

We begin, then, by defining these properties.

Definition 17. We say that sets X, Y are *almost complementary in \mathcal{M}* if $|X| = |Y|$ and *none* of the following is satisfied in \mathcal{M} :

$$|Y - X| \leq |X \cap Y| \quad (72)$$

$$|X - Y| \leq |X \cap Y| \quad (73)$$

$$|X - Y| \leq |M - (X \cup Y)| \quad (74)$$

$$|Y - X| \leq |M - (X \cup Y)| \quad (75)$$

Similarly, we say that X, Y are *symmetric in \mathcal{M}* if $|X| = |Y|$ and for some $Z \in \mathcal{M}$ distinct from X and Y , $|Z| = |X|$ and neither of the following is satisfied for \mathcal{M} :

$$|Z - (X \cup Y)| \leq |X - Y| \quad (76)$$

$$|Z - (X \cup Y)| \leq |Y - X| \quad (77)$$

In many cases, we will establish that either (76) or (77) holds in \mathcal{M} by establishing that one of the following holds for \mathcal{M} :

$$|M - (X \cup Y)| \leq |X - Y| \quad (78)$$

$$|M - (X \cup Y)| \leq |Y - X| \quad (79)$$

This is licensed by (37).

The following two propositions follow by (47) and (48) and CC.

Proposition 18. Suppose $\pi : M \rightarrow M$ is a permutation. Then:

1. X, Y are almost complementary if and only if $\pi(X), \pi(Y)$ are almost complementary.
2. X, Y are symmetric if and only if $\pi(X), \pi(Y)$ are symmetric.

Proposition 19. Distinct sets X and Y cannot be both almost complementary and symmetric.

Definition 20. In what follows, distinct, E -equivalent sets that satisfy one of (72–75) and one of (76–77) will play a special rôle. Thus we say that sets X, Y are *opportune* if they are distinct, $Y \in [X]$, and they are neither almost complementary nor symmetric. We moreover will say that a single set X is opportune if there is a Y such that the pair X, Y is opportune.

Definition 21. Two sets X, Y are *relatively finite in \mathcal{M}* just if $|X - Y| = |Y - X|$ in \mathcal{M} . (Note that if X, Y are relatively finite, then $X \boxminus Y$.) We set $RF(X)$ to be the set of sets Y that are relatively finite in \mathcal{M} to X .

Lemma 22. If X, Y are opportune then $RF(X) \subseteq [X]_{E(\boxminus)}$. That is, every set relatively finite to X in \mathcal{M} is E -equivalent to X .

Proof. Let $Z \in RF(X)$ with $Z \stackrel{f}{\approx} X$ in \mathcal{M} ; we may assume that f fixes $Z \cap X$ pointwise. We show that there is a permutation π of M such that $\pi(X) = Z$ and $[\pi(X)] = [X]$.

First we partition Z :

$$\begin{aligned} Z_1 &= f^{-1}(X - Y) \cap Y - X & Z_2 &= f^{-1}(X - Y) - Y \\ Z_3 &= f^{-1}(X \cap Y) \cap Y - X & Z_4 &= f^{-1}(X \cap Y) - Y \\ Z_5 &= f^{-1}(X) \cap X = X \cap Z \end{aligned}$$

where the equality in Z_5 is due to the choice of f . It is easy to see that these sets are pairwise disjoint and that their union is Z .

Note that $Z_1 \stackrel{f|Z_1}{\preceq} X - Y$. Since X and Y are not almost complementary, at least one these holds:

$$|Z_1| \leq |X \cap Y| \quad |Z_1| \leq |M - (X \cup Y)| \quad (80)$$

Either way Lemma 13 provides a permutation $\pi_1 \in \mathcal{M}$ such that $[\pi_1 X] = [X]$ and $Z_1 \subset \pi_1(X - Y) \subset \pi_1(X)$ and satisfying (52). Note that by (52), $Z_5 \subset \pi_1(X)$, so

$$Z_1 \cup Z_5 \subset \pi(X) \quad (81)$$

Set $X_1 = \pi_1(X), Y_1 = \pi_1(Y)$.

Then since $Z_2 \subset M - (X_1 \cup Y_1)$ (by (52)), and since (by (52) again) $Z_2 \stackrel{f|Z_2}{\preceq} X_1 - Y_1$, by Proposition 15 there is a permutation $\pi_2 \in \mathcal{M}$ with $Z_2 \subset \pi_2(X_1 - Y_1) \subset \pi_2(X_1) \in [X]$ and satisfying (52). By (81) and (52), we have

$$Z_1 \cup Z_2 \cup Z_5 \subset \pi_2(X_1) \in [X] \quad (82)$$

Set $X_2 = \pi_2(X_1)$, $Y_2 = \pi_2(Y_1)$. Since $Z_3 \subset Y_2 - X_2$ and $Z_3 \stackrel{f \upharpoonright Z_3}{\preceq} X_2 \cap Y_2$ (by (52) twice), there is by Lemma 16 a permutation $\pi_3 \in \mathcal{M}$ satisfying (52) and with $Z \subset \pi_3(X_2 - Y_2) \subset \pi_3(X_2) \in [X]$. By (82) and (52) we have that

$$Z_1 \cup Z_2 \cup Z_3 \cup Z_5 \subset \pi_3(X_2) \quad (83)$$

Finally set $X_3 = \pi_3(X_2)$, $Y_3 = \pi_3(Y_2)$. Note that $Z_4 \subset M - (X_3 \cup Y_3)$ and that $Z_4 \stackrel{f \upharpoonright Z_4}{\preceq} X_3 \cap Y_3$ (by (52) twice). Notice now that since π_1, π_2 , and $\pi_3 \in \mathcal{M}$ are permutations and since X, Y are not symmetric, X_3, Y_3 are not symmetric by Proposition 18. Thus we have at least one of

$$|Z_4| \leq |X_3 - Y_3| \quad |Z_4| \leq |Y_3 - X_3| \quad (84)$$

obtains. So by Lemma 14, there is a permutation $\pi_4 \in \mathcal{M}$ such that

$$Z_4 \subset \pi_4(X_3 \cap Y_3) \subset \pi_4(X_3) \in [X] \quad (85)$$

By (83) and again (52) we have that

$$Z = Z_1 \cup Z_2 \cup Z_3 \cup Z_4 \cup Z_5 = \pi_4(X_3) = \pi_4(\pi_3(\pi_2(\pi_1(X)))) = \pi(X) \quad (86)$$

for $\pi = \pi_1 \circ \pi_2 \circ \pi_3 \circ \pi_4$; which was to be established. \square

Dealing with opportune sets via Lemma 22 will enable us to deal with cases in which X is finite, both within and outside of \mathcal{M} . We use a special class of opportune sets to deal with X infinite in \mathcal{M} .

Definition 23. We will say that X, Y are *ideally opportune* if they are opportune and either $|X| \leq |M - X|$ or $|Y| \leq |M - Y|$. We will also say that the single concept X is ideally opportune if there is a Y such that X, Y are opportune and $|X| \leq |M - X|$.

We now turn to proving another lemma that will address the more general case, in which X is not finite in the metatheory (though it may be finite in \mathcal{M}). We begin with another preparatory lemma.

Lemma 24. Suppose X is ideally opportune, and $X \boxminus Z$. Then $Z \in [X]$.

Proof. Note that if $|X| < \omega$ then X and Z are relatively finite, so $Z \in [X]_{E(\boxminus)}$ by Lemma 22. So we assume $|X| \geq \omega$.

Thus by Proposition 10, and since X is ideally opportune, we may assume that $|Z| = |X| \leq |M - X| = |M - Z|$.

Now if $|X| = |M - X|$, then

$$|M - Z| \geq |Z| = |X| = |M - X| = |M|$$

If $|Z - X| < |X - Z|$, then as $M - X = (M - (X \cup Z)) \cup (Z - X)$, by ISM, we obtain

$$|X| \leq |M - (X \cup Z)| \quad (87)$$

Similarly for $|X - Z| < |Z - X|$.

On the other hand if $|X| < |M - X|$, then as $|X| = |Z|$, $|X \cup Z| < |M - X|$ by **ISM**. Then by **ISM** again, we have that

$$|M - (X \cup Z)| = |M| \geq |M - X| > |X|$$

which again yields (87).

Given (87) with f a witnessing injection, observe that $f(X)$ and X are disjoint and so relatively finite. By Lemma 22, since X is one of a pair of opportune sets, $f(X) \in [X]_{E(\Theta)}$. But then $f(X), X$ are opportune as well. Moreover, as $Z, f(X)$ are disjoint and equinumerous, they also are relatively finite. So by Lemma 22, $Z \in [f(X)]_{E(\Theta)} = [X]_{E(\Theta)}$. \square

5.4 The intermediate Theorem 6

We are now nearly ready to prove Theorem 6. We first deal with degenerate cases, namely those in $E(\Theta)_X$ that may be trivial *and also* separative. The following propositions follow from **IPM** and (40) and (44) of Proposition 8.

Proposition 25. If

$$\mathcal{M} \models |M| \leq 1 \vee |X| = 0 \vee |X| = |M| < \omega \quad (88)$$

then $E(\Theta)_X$ is both a separation and trivial.

Further, if $|M| = 2$ and $|X| = 1$, then $E(\Theta)_X$ is either both trivial and a complementation, or it is a separation.

Proposition 26. If $Y, Z \in [X]_{\Theta}$ are symmetric, then $|Y \triangle Z| \triangleleft |X|$.

Proposition 27. If $Y, Z \in [X]_{\Theta}$ are almost complementary, then $|Y \cap Z|, |M - (Y \cup Z)| \triangleleft |X|$.

Proposition 28. If $\mathcal{M} \models \omega \leq |X| \leq |M - X|$ and $E(\Theta)_X$ is either separative or complementative, then it is not trivial.

The crucial move in proving Theorem 6 is in the following Lemma:

Lemma 29. Suppose $\mathcal{M} \models |M| > 2 \wedge 1 < |X| \leq |M - X|$ and E is an L_0 definable equivalence relation over \mathcal{M} . Then $E(\Theta)_X$ is exactly one of: trivial, properly separative, or properly complementative.

Proof. We first establish the lemma for the case where $|X| < \omega$ in \mathcal{M} . Obviously no more than one can obtain; we show at least one must. Since all equinumerous sets finite in \mathcal{M} are relatively finite in \mathcal{M} , by Lemma 22, if there are opportune sets in $[X]_{E(\Theta)}$, then

$$[X]_{\Theta} = [X]_{\Theta} \cap RF(X) \subseteq [X]_{E(\Theta)}$$

and so $E(\Theta)_X$ is trivial. So we may assume there are no opportune sets $Y, Z \in [X]_{E(\Theta)}$, and thus that for any $Y, Z \in [X]_{E(\Theta)}$, either Y, Z are symmetric, or Y, Z are almost complementary by Lemma 22.

Suppose now that all pairs of set in $[X]_{E(\boxminus)}$ are symmetric. By Proposition 26, all such pairs satisfy (30). Now by Proposition 18 this holds for every concept bicardinally equivalent to X ; thus $E(\boxminus)_X$ is properly separative.

On the other hand, if there are $Y, Z \in [X]_{E(\boxminus)}$ that are not symmetric, then these are almost complementary. By Proposition 27, the non-symmetric concepts in $[X]_{E(\boxminus)}$ satisfy (31). Clearly then (30) fails of them, so $E(\boxminus)_X$ is not separative. By Proposition 26 all the symmetric concepts satisfy (30), and so (31). Proposition 18 ensures then that $E(\boxminus)_X$ is complementative, and since (as just shown) not separative, it is properly complementative. So the result holds for all $|X|$ finite in \mathcal{M} .

The case in which X is infinite in \mathcal{M} is similar. By Proposition 28, at most one of the options can hold; we show at least one must. By Lemma 24 if $E(\boxminus)_X$ is non-trivial then there are no ideally opportune concepts in $[X]_{E(\boxminus)}$. Since by assumption $|X| \leq |M - X|$ it follows that there are no opportune concepts in $[X]_{E(\boxminus)}$. The above argument establishes again that $E(\boxminus)_X$ is either properly separative or properly complementative. \square

Corollary 30. Suppose $\mathcal{M} \models |M| > 2 \wedge 1 < |M - X| < |X|$ and E is an L_0 definable equivalence relation over \mathcal{M} . Then $E(\boxminus)_X$ is exactly one of: trivial, properly separative, or properly complementative.

Proof. It is easy to see that if E is an equivalence relation, then so is the L_0 definable

$$E^c(X, Y) \Leftrightarrow E(M - X, M - Y)$$

Thus, by Lemma 29, $E^c(\boxminus)_{M-X}$ is exactly one of trivial, properly separative, or properly complementative. But by the definition of E^c , $E^c(\boxminus)_{M-X}$ has the same profile as $E(\boxminus)_X$. \square

Lemma 29 and Corollary 30 suffice to prove the first part of Theorem 6. The remainder follows from the following lemma.

Lemma 31. Suppose $\mathcal{M} \models |X| < \omega$ and there is a $W \neq X$ with $W \in [X]_{E(\boxminus)}$, and $E(\boxminus)_X$ non-trivial. Then for any $Y, Z \in [X]_{\boxminus}$, either (32) or (33) holds.

Proof. By the first part of Theorem 6, $E(\boxminus)_X$ is either properly separative or properly complementative. By Lemma 22, we know that though X, W are relatively finite (since they are finite), they are not opportune, so X, W are either symmetric or almost complementary.

We will first show that if X, W are symmetric, then for any $Z \in [X]_{\boxminus}$,

$$|X \triangle Z| < |X| \Rightarrow \mathcal{M} \models E(X, Z) \tag{89}$$

Assume X, W symmetric, so for some H with $|X| = |W| = |H|$,

$$|H - (X \cup W)| \triangleright |X - W|, |W - X| \tag{90}$$

Assuming

$$|X \triangle Z| < |X| \tag{91}$$

we have that $|X| = |H| \triangleright |H - (X \cup W)|$; it then follows from (41) of Proposition 8 that

$$|X \cap W| \triangleright |X - W|, |W - X| \quad (92)$$

So by (38) of Proposition 8, we have

$$|Z - X| \leq |X \triangle Z| \triangleleft |X \cap W| \quad (93)$$

Thus by Lemma 13, $Z - X \subseteq \pi(X)$ with $\pi(X \cap Z) = X \cap Z$ and $E(\pi(X), X)$. Thus $Z \in [X]_E$, and this establishes (32).

Now we will show that if X, W are almost complementary, then for any $Z \in [X]_\Box$,

$$[|X \triangle Z| \triangleleft |X| \text{ or } |X \cap Z|, |M - (X \cup Z)| \triangleleft |X|] \Rightarrow \mathcal{M} \models E(X, Z) \quad (94)$$

So assume X, W are almost complementary.

If (91) obtains, then we have

$$|(Z - X) \cap W| \leq |Z - X| \leq |Z \triangle X| \triangleleft |X \cap W| \quad (95)$$

by (41) of Proposition 8, and so $Z \in [X]_E$ by Lemma 13.

On the other hand suppose

$$|X \cap Z| \triangleleft |X| \quad (96)$$

$$|M - (X \cup Z)| \triangleleft |X| \quad (97)$$

By (96) and (41), $|X \cap Z| \triangleleft |X - Z|$. But then by (97) and (41) again,¹³

$$|(W - Z) \cap X| \leq |X \cap Z| \triangleleft |W \cap Z| \quad \text{and} \quad (98)$$

$$|W - (X \cup Z)| \leq |M - (X \cup Z)| \triangleleft |W \cap Z|, \quad \text{so} \quad (99)$$

$$|W - Z| \triangleleft |W \cap Z| \quad (100)$$

by (40). As similar argument shows that

$$|Z - W| \triangleleft |W \cap Z| \quad (101)$$

and thus that $|W \triangle Z| \triangleleft |W|$ by (38). From here we can use the argument from earlier in this proof establishing (32), and the fact that $[W]_E = [X]_E$, and we have that $Z \in [X]_E$. \square

From Theorem 6 we obtain the following corollary, which we will use in proving Theorem 7.

Corollary 32. If X is finite in the metatheory, then $[X]_{E(\Box)}$ is either $[X]_\Box$, or $\{X\}$, or $\{X, M - X\}$.

Proof. If X is finite in the metatheory then $|X \triangle Y| \triangleleft |X|$ if and only if $|X \triangle Y| = 0$. Likewise for $|X \cap Y|, |M - (X \cup Y)| \triangleleft |X|$. \square

¹³The reader may wish to aid her reasoning in these following arguments using Venn Diagrams; label a set A “S” for *relatively small*, and B “L” for *relatively large*, if $A \triangleleft B$. Proposition 8 basically asserts that the expected reasoning will obtain (e.g., if A is relatively large but $A \cap B$ is relatively small, it follows that $A - B$ is relatively large).

5.5 The Classification Theorem 7

To prove Theorem 7 we need to use **CWF** to address some complications lingering due to our use of non-standard semantics. We show that under **CWF** all proper separations and proper complementations on finite concepts are the finest possible, even if “finite” only means “finite in \mathcal{M} ”. Theorem 7 follows from the theorem proved in this section.

Definition 33. Suppose that E is properly separative on $[X]_{\boxminus}$. Then $J \in \mathcal{M}$ is called a *measure* of E on $[X]_{\boxminus}$ just if for all $Y, Z \in [X]_{\boxminus}$,

$$E(Y, Z) \Leftrightarrow |Y \triangle Z| \triangleleft |J|$$

Suppose that E is properly complementative on $[X]_{\boxminus}$. Then $J \in \mathcal{M}$ is called a *measure* of E on $[X]_{\boxminus}$ just if for all $Y, Z \in [X]_{\boxminus}$,

$$E(Y, Z) \Leftrightarrow |Y \triangle Z| \triangleleft |J| \text{ or } |M - (Y \triangle Z)| \triangleleft |J|$$

Theorem 34. Let $\mathcal{M} \models |X| < \omega$. If E is nontrivial at $[X]_{\boxminus}$ then its measure is \emptyset .

Proof. If E is nontrivial on $[X]_{\boxminus}$ then by Theorem 6 it is either a proper separation or a proper complementation. Suppose J is a non-zero measure of E on $[X]_{\boxminus}$.

Let $L(\xi)$ abbreviate the formula

$$(\forall U, V \in [X]_{\boxminus})(E(U, V) \rightarrow (|U \triangle V| < |\xi| \vee |M - (U \triangle V)| < |\xi|))$$

Clearly $L(J)$, so by **CWF**, there is a smallest J' such that $L(J')$. If $J' \neq \emptyset$, then observe that by Corollary 32 that $|J'| \neq n$ for any natural number n —that is, J is not finite in the metatheory. Thus, letting $j \in J'$ we have that $\neg L(J' - \{j\})$, and so there are $U, V \in [X]_{E(\boxminus)}$ such that

$$E(U, V) \text{ but } |U \triangle V| \not\triangleleft |J' - \{j\}| \quad (102)$$

(If the witnessing U, V to $\neg L(J')$ satisfy the second disjunct of the consequent, just choose $U, M - V$ so that they satisfy (102)). As $L(J)$ we have that $|U \triangle V| = |J' - \{j\}|$.

By assumption $E(\boxminus)_X$ is non-trivial, so regardless of whether it is a complementation or a separation, there is a Z with

$$U \neq V \quad |U \triangle V| \triangleleft |J| \quad |Z - (U \cup V)| \leq |J| \quad (103)$$

As $|X| \geq |J|$, $|U| = |V| = |X|$, and $|U \triangle V| \triangleleft |J|$, we have that $|U \cap V| \geq |X|$ by Proposition 8. Thus, with (103) we have that

$$|U \cap V|, |Z - (U \cup V)| > 2 \quad (104)$$

Thus, let $a \in U \cap V$ and $b, c \in Z - (U \cup V)$, and set

$$U' = (U - \{a\}) \cup \{b\} \quad V' = (V - \{a\}) \cup \{c\} \quad (105)$$

Note now that as

$$|U \triangle U'| = |V \triangle V'| = 2 \triangleleft |J| \quad (106)$$

we have $E(U, U')$, $E(V, V')$, and so $E(U', V')$. Thus, since $L(J)$, $|U' \triangle V'| < |J|$, but it is easy to see that the construction of U' and V' ensure that $|U' \triangle V'| = |J \cup \{a\}| > |J|$ \square

Theorem 34 proves Theorem 7 for X finite in \mathcal{M} . For the rest one needs only consider X such that $\mathcal{M} \models |X| \geq \omega$. By Proposition 10, if $E(\boxminus)_X$ is separative, then

$$\mathcal{M} \models (\exists X' \subseteq X)(\forall Y, Z \in [X]_{\boxminus})(E(Y, Z) \rightarrow |Y \triangle Z| < |X'|) \quad (107)$$

and likewise if $E(\boxminus)_X$ is complementative. This completes the proof of Theorem 7.

6 The classification of “Bad Companions”

The “bad company problem” faced by neo-logicists is the two part challenge of finding a plausible criterion for the “logicality” of abstraction principles, and showing that the good abstraction principles like HP are logical in this sense, while the bad ones like BLV, NP, and CP are not. It is thus relevant to note how Theorem 7 and its apparatus bear on the joint consistency of abstraction principles, since at the very least any abstraction principles qualifying as “logical” should be jointly consistent.

Consistency results in the absence of well-behaved cardinalities are difficult to obtain. As such our discussion will be directed towards the neo-logicist who thinks “logical” abstraction principles must be jointly consistent *in the presence of well-behaved cardinalities*. We’ll thus use “jointly consistent” as if it implicitly has the qualification on the behavior of cardinalities.

Such a neo-logicist is committed to thinking that joint consistency with HP is a necessary condition for the logicality of abstraction principles. In the remainder of this section we apply Theorem 7 and its apparatus to discern which abstraction principles are consistent with a Dedekind infinite universe. These obviously determine which abstraction principles are consistent with HP, and thus which must be ruled out from being “logical”.

One reason why the bad company problem can appear vexing is that any “logical” equivalence relation can give rise to an abstraction principle. Facing an untamed menagerie of such equivalence relations, the neo-logicist would face an equally wild zoo of abstraction principles. The import of Theorem 7 is in taming the menagerie—or better, classifying the species. At each bicardinal slice of \mathcal{M} , there are only three “classes” of equivalence relations on that slice. This enables us to see clearly why certain abstraction principles have the types of models they do: at certain (finite or infinite!) bicardinal classes, *non-trivial* equivalence relations distinguish more equivalence classes than there are objects.

We have said that BLV, NP, and CP all count as bad companions. We thus have so far two bad companions deploying refinements of separations, and one deploying a refinement of a complementation. Of the former, BLV involves the finer (in fact finest) separation; as we will see in contexts with **IPM** it is in a sense the paradigm case of a separative abstraction principle. We will also discuss the paradigm case of a complementative abstraction principle; it is not CP but a close relative, the *Liberated Complementation Principle* (LCP):

$$(\forall X, Y)(\ell X = \ell Y \leftrightarrow X = Y \vee X = M - Y)$$

Remark 35. The equivalence relation for BLV is the finest equivalence relation that refines a separation for every $X \in \mathcal{M}$ (it is also the finest equivalence relation full stop, see [1]).

The equivalence relation for NP refines a separation at all bicardinal slices of \mathcal{M} for Dedekind infinite sets.

Finally, the equivalence relation for LCP refines a complementation at $X \in \mathcal{M}$ if $|X| = |M - X|$, and in fact it is the finest equivalence relation refining a complementation at bicardinal slices for such X .

6.1 Translation and restricted abstraction principles

In the remainder of this section we will need the following definitions:

Definition 36. For $\mathcal{M} \models |X|, \omega \leq |M| = |M - X|$, we will write $Y \leq [X]_{\Box}$ as shorthand for $|Y| < |X| \vee Y \in [X]_{\Box}$.

Let E be an equivalence relation. Let $\varphi(Y, Z)$ be a formula. We say that φ is *functional below X in \mathcal{M}* if

$$\mathcal{M} \models |X| \leq |M - X| \wedge (\forall Y)(Y \leq [X]_{\Box} \rightarrow (\exists! Z)\varphi(Y, Z)) \quad (108)$$

We say that φ is *functional at X* if

$$\mathcal{M} \models (\forall Y)(Y \in [X]_{\Box} \rightarrow (\exists! Z)\varphi(Y, Z)) \quad (109)$$

Often in the presence of such a φ we will use functional notation, using $F_{\varphi}(Y)$ for the unique object Z such that $\varphi(Y, Z)$, dropping the subscript where it is unneeded for understanding.

If $A_E[\partial]$ is an abstraction principle, and φ a formula functional below X in \mathcal{M} , then we say that θ *translates $A_{E'}(\leq X)$ via φ* in \mathcal{M} to mean that

$$\mathcal{M} \models (\forall Y, Z)(Y, Z \leq [X]_{\Box} \rightarrow (\partial(F_{\varphi}(Y)) = (\partial(F_{\varphi}(Z)) \leftrightarrow E'(Y, Z))) \quad (110)$$

If φ is functional at X , then we say that $A_E[\partial]$ *translates $A_{E'}(= X)$ via φ* in \mathcal{M} to mean that

$$\mathcal{M} \models (\forall Y, Z)(Y, Z \in [X]_{\Box} \rightarrow (\partial(F_{\varphi}(Y)) = \partial(F_{\varphi}(Z)) \leftrightarrow E'(Y, Z))) \quad (111)$$

Finally we may say that $A_E[\partial]$ *translates $A_{E'}$ in \mathcal{M} via φ* just if

$$\mathcal{M} \models (\forall Y, Z)(\partial(F_{\varphi}(Y)) = \partial(F_{\varphi}(Z)) \leftrightarrow E'(Y, Z)) \quad (112)$$

These definitions afford us the capacity to talk of the *restricted abstraction principles* $A_E[\partial](\leq X)$ and $A_E[\partial](= X)$ in the sense that \mathcal{M} satisfies one of these just if “ $(\forall x)(x = x)$ ” translates that principle in \mathcal{M} via “ $Y = Z$ ”.

In what follows of this section we show that if $2 < |X| = |M - X|$ and $E(\Xi)_X$ is non-trivial, then (in the presence of **IPM**) A_E is inconsistent. The engine of these results will be the following two propositions.

Proposition 37. Let ϖ be an L_0 -sentence. Suppose that for all $\mathcal{M} \models A_E[\partial]$, $A_E[\partial] \wedge \varpi$ translates $A_{E'}(= M)$ in \mathcal{M} via φ , and that $A_{E'}(= M)$ and ϖ are jointly inconsistent (in our background logic). Then A_E and ϖ are jointly inconsistent (in our background logic) as well.

Proof. If $A_E \wedge \varpi$ is consistent then by the completeness of second-order logic for the non-standard semantics, there is a structure $\mathcal{M} = (M, S_1[M], \dots, \partial) \models A_E \wedge \varpi$. As $\varphi(Y, Z)$ is functional at M , define $\partial'Y = \partial(F_\varphi(Y))$; ∂' is then a function in \mathcal{M} from concepts to objects. We now augment \mathcal{M} to $\mathcal{M}' = (M, S_1[M], \dots, \partial, \partial')$, and note that since A_E translates $A_{E'}(= M)$, $\mathcal{M}' \models A_{E'}[\partial'](= M)$. Since ϖ is an L_0 -sentence (with parameters from \mathcal{M}), $\mathcal{M}' \models \varpi$. A routine induction on the complexity of formulas shows that \mathcal{M}' satisfies the comprehension axioms in the expanded language $L_0[\partial']$. Thus if $A_E[\partial] \wedge \varpi$ has a model then so does $A_{E'}[\partial'](= M) \wedge \varpi$. \square

For the next proposition, observe that if $\mathcal{M} \models |M| \geq \omega$, then **IPM** ensures that there is a bijection $\langle \cdot, \cdot \rangle : M \times M \rightarrow M$. We will use this notation in what follows, as well as writing $\langle X, Y \rangle$ to mean $\{\langle x, y \rangle \mid x \in X, y \in Y\}$.¹⁴ We also obtain:

Proposition 38. For $|M| \geq \omega$, let $\langle \cdot, \cdot \rangle : M \times M \rightarrow M$ be a bijection, whose existence is assured by **IPM**. For X, Y , if either

$$|Y| = 1 \leq |X| \leq |M - X| \quad (113)$$

or

$$|Y|, \omega \leq |X| \leq |M - X| \quad (114)$$

then $\langle Y, X \rangle \in [X]_\Xi$.

Lemma 39. Let $\mathcal{M} \models |X| \geq \omega$. If $\mathcal{M} \models |X| \leq |M - X|$ and $E(\Xi)_X$ refines a separation, then A_E translates $\text{BLV}(\leq X)$ via

$$\varrho(Y, Z) := Z = \langle Y, X \rangle \quad (115)$$

Under the same conditions, if $E(\Xi)_X$ refines a complementation (and does not refine a separation), then A_E translates LCP via

$$\begin{aligned} \varphi(Y, U) := & (\exists V)((\partial\emptyset \in Y \wedge V = \langle Y, X \rangle) \vee (\partial\emptyset \in M - Y \wedge V = \langle M - Y, X \rangle)) \\ & \wedge ((V \neq \langle M, X \rangle \wedge V = U) \vee (V = \langle M, X \rangle \wedge U = \emptyset)) \end{aligned} \quad (116)$$

¹⁴The arguments of this section elaborate on the one given in [6].

Proof. For both assertions, note: By hypothesis, $\mathcal{M} \models |M| \geq |X| \geq \omega$, so by **IPM** the bijection $\langle \cdot, \cdot \rangle$ exists. Notice that since $E(\boxminus)_X$ is non-trivial, $\partial\emptyset \neq \partial Y$ for any $Y \in [X]_{\boxminus}$, since otherwise by **PI** $E(\boxminus)_X$ would be trivial, as $E(\emptyset, X)$ holds whenever $E(\emptyset, f(X))$ holds for $|X| = |f(X)|$ and $|M - X| = |f(M - X)|$. Further if this $Y \neq \emptyset$ then $\langle Y, X \rangle \in [X]_{\boxminus}$ by Proposition **38** (under the conditions of the second assertion this follows since $|X| = |M - X| = |M|$.)

We prove the first assertion in brief: It is easy to verify that ϱ is functional (and so is functional below X). Let $|Y|, |Z| \leq |X|$.

For distinct $Y, Z \neq \emptyset$, $|\langle Y, X \rangle \triangle \langle Z, X \rangle| = |X|$ by **IPM**, so if $Y \neq Z$ then since $E(\boxminus)_X$ is a separation, $\neg E(\langle Y, X \rangle, \langle Z, X \rangle)$, and so by A_E , $\partial\langle Y, X \rangle \neq \partial\langle Z, X \rangle$. Conversely if $Y = Z$ then $\langle Y, X \rangle = \langle Z, X \rangle$. Since E is an equivalence relation, it follows that $E(\langle Y, X \rangle, \langle Z, X \rangle)$, and so, by A_E , that $\partial\langle Y, X \rangle = \partial\langle Z, X \rangle$.

For the second assertion, it is easy to verify that φ is functional (and so is functional below X). Towards verifying the consequent of (110) and working in \mathcal{M} , we need only show that

$$\partial U_\varphi(Y) = \partial U_\varphi(Z) \leftrightarrow Y = Z \vee Y = M - Z \quad (117)$$

for all concepts $Y, Z \in \mathcal{M}$.

By the definition of U_φ , $U_\varphi(\emptyset) = U_\varphi(M) = \emptyset$. So clearly if X and Y are chosen from \emptyset, M , then (117) obtains.

Now, if $Y = \emptyset$ then $U_\varphi(Y) = \emptyset$, and if $Y = M$ then $U_\varphi(Y) = \emptyset$. On the other hand if $Y \neq \emptyset, M$, then $U_\varphi(Y) \in [X]_{\boxminus}$. Thus by $A_E[\partial]$, $\partial(U_\varphi(Y)) \neq \partial(U_\varphi(\emptyset)) = \partial(U_\varphi(M))$. Thus:

$$\emptyset \neq Z \rightarrow \partial(U_\varphi \emptyset) \neq \partial(U_\varphi Z) \quad (118)$$

$$M \neq Z \rightarrow \partial(U_\varphi M) \neq \partial(U_\varphi Z) \quad (119)$$

Now if $Y, Z \neq \emptyset, M$ then we have

$$Y \neq Z \rightarrow |U_\varphi(Y) \triangle U_\varphi(Z)| = |X| \quad (120)$$

$$Y \neq M - Z \rightarrow |M - (U_\varphi(Y) \triangle U_\varphi(Z))| = |X| \quad (121)$$

So, since $E(\boxminus)_X$ refines a complementation we have that

$$\partial(U_\varphi(Y)) = \partial(U_\varphi(Z)) \rightarrow Y = Z \vee Y = M - Z \quad (122)$$

by $A_E[\partial]$.

Clearly by the construction of U_φ :

$$Y = Z \rightarrow E(U_\varphi(Y), U_\varphi(Z)) \quad (123)$$

$$Z \neq \emptyset \neq Y = M - Z \rightarrow E(U_\varphi(Y), U_\varphi(Z)) \quad (124)$$

So we have, again by $A_E[\partial]$ and the $E(\boxminus)_X$ refining a complementation:

$$Y = Z \vee Y = M - Z \rightarrow \partial(U_\varphi(Y)) = \partial(U_\varphi(Z)) \quad (125)$$

Which completes the proof of the translation of LCP. \square

We know now some of which abstraction principles translate restrictions of BLV and LCP. We now show when, in the presence of **IPM**, these restrictions are inconsistent for each such principle. Each of the following two subsections will conclude by applying Lemma 39; the final subsection applies Theorem 7 to unify those results.

6.2 Refinements of separations and $\text{BLV}(= X)$

The utility of Lemma 39 is in relating equivalence relations that are non-trivial on bicardinal slices to what might be called their prime examples. In this section we treat BLV and its restrictions as the prime examples of separations, and show that in infinite structures, refinements of separations translate certain restrictions of BLV. The next section proves analogous results for LCP and its restrictions.

From the Lemma we obtain very quickly:

Corollary 40. Suppose $\mathcal{M} \models |X| = |M - X| = |M|$. Then $\mathcal{M} \not\models \text{BLV}(= X)$.

Proof. Given X as in the hypothesis, if $M \models \text{BLV}(= X)$ then $E(\boxminus)_X$ is a separation, and so refines a separation. Thus $\text{BLV}(= X)$ translates both $\text{BLV}[\varepsilon_1](\leq X)$ and $\text{BLV}[\varepsilon_2](\leq M - X)$. With $M \overset{f}{\approx} X$ and $M \overset{g}{\approx} M - X$, set

$$\varepsilon Y = \begin{cases} f(\varepsilon_1 Y) & |Y| \leq X \\ g(\varepsilon_2 Y) & |M - Y| < |M - X| \end{cases}$$

Clearly then $\text{BLV}(= X)$ and $|X| = |M - X| = |M|$ translate BLV, which is inconsistent. Thus BLV and $|X| = |M - X| = |M|$ are jointly inconsistent, and by Proposition 37, $\text{BLV}(= X)$ and $|X| = |M - X| = |M|$ are jointly inconsistent, which was to be demonstrated. \square

The situation is no better for $\text{BLV}(= X)$ if M is finite (in \mathcal{M}):

Lemma 41. Suppose $\mathcal{M} \models 0 < |X| < |M| < \omega$. Then $\mathcal{M} \not\models \text{BLV}(= X)$.

Proof. Let X be given, and since $|M| > 1$, we establish the following abbreviations: for $x \in X$ and $a \notin X$,

$$X_x = X - \{x\} \tag{126}$$

$$a + X_x = X_x \cup \{a\} \tag{127}$$

Notice that for all $x \in X$ and $a \notin X$, $|a + X_x| \in [X]_{\boxminus}$ but $a + X_x \neq X$. Thus if $\mathcal{M} \models \text{BLV}(= X)$, then $\varepsilon X \neq \varepsilon(a + X_x)$. Moreover, for distinct $x, y \in X$, $a + X_x$ and $a + X_y$ are distinct (thus so are their abstracts), and for distinct $a, b \notin X$, $a + X_x$ and $b + X_x$ are distinct (thus so are their abstracts).

Now $|\{a + X_x \mid x \in X\}| = |X|$ for each $a \notin X$, and $|\{a + X_x \mid a \notin X\}| = |M - X|$ for each $x \in X$. Further for fixed $x_0 \in X, a_0 \notin X$,

$$\{a_0 + X_x \mid x \in X\} \cap \{a + X_{x_0} \mid a \notin X\} = \{a_0 + X_{x_0}\}$$

Lastly note that by Permutation Invariance, $\varepsilon\emptyset \neq \varepsilon X, \varepsilon(a + X_x)$ for any $a \notin X, x \in X$. Thus, and since $\mathcal{M} \models |M| < \omega$, and letting

$$A = \{\varepsilon(a_0 + X_x) \mid x \in X\} \quad B = \{\varepsilon(a + X_{x_0}) \mid a \notin X\} \quad (128)$$

we have

$$|M| = |M - X| + |X| \quad (129)$$

$$= |A| + |(B - \{\varepsilon(a_0 + X_{x_0})\}) \cup \{\varepsilon X\}| \quad (130)$$

$$= |A \cup B \cup \{\varepsilon X\}| \quad (131)$$

$$< |A \cup B \cup \{\varepsilon X\}| + |\{\varepsilon\emptyset\}| \quad (132)$$

$$\leq |\text{rng}(\varepsilon)| \leq |M| \quad (133)$$

which means that $|M| < |M|$, a contradiction. \square

The preceding results on refinements of separations and on restrictions of BLV yield:

Corollary 42. If $E(\boxminus)_X$ refines a separation and $\mathcal{M} \models |X| = |M - X|$, then $\mathcal{M} \not\models A_E$.

Proof. Corollary 40 and Lemma 41 imply that (in the presence of **IPM**)

$$\mathcal{M} \models |X| = |M - X| \rightarrow \neg \text{BLV}(= X) \quad (134)$$

If $E(\boxminus)_X$ is a separation and $\mathcal{M} \models |X| < \omega$ then by Theorem 7 $\mathcal{M} \models \text{BLV}(= X)$ if $\mathcal{M} \models A_E$. But this contradicts (134).

On the other hand, if $\mathcal{M} \models |X| \geq \omega$, then by Lemma 39, A_E translates $\text{BLV}(= X)$. So A_E cannot hold in \mathcal{M} as again this contradicts (134). \square

6.3 Refinements of complementations and $\text{LCP}(= X)$

The situation is quite similar for LCP. Here we handle the finite case first:

Lemma 43.

$$\text{LCP}(= X) \models |X| = |M - X| < \omega \rightarrow |M| \leq 4 \quad (135)$$

$$\text{LCP} \models |M| < \omega \rightarrow |M| \leq 2 \quad (136)$$

Here, abusing notation, “ \models ” indicates the derivability relation in second-order logic with well-behaved cardinalities.

We begin with the following:

Proposition 44. The following is a theorem of second-order logic:

$$(\forall X, Y)(|X| > 2, |Y| \geq 2 \rightarrow |X \cup Y| \leq |X \times Y|) \quad (137)$$

with a strict inequality in the consequent if $X \cup Y$ is Dedekind finite.

Proof. We assume without loss of generality that X and Y are disjoint, let $a, b, c \in X$ and $d, e \in Y$ be pairwise distinct. Then define $f : X \cup Y \rightarrow X \times Y$ by

$$f(x) = \begin{cases} (a, x) & x \in Y, x \neq d \\ (x, d) & x \in X \\ (b, e) & x = d \end{cases}$$

Clearly f is injective; it is not surjective since $(c, e) \notin \text{rng}(f)$. Further, if $g : X \times Y \rightarrow X \cup Y$ is a bijection, then $g \circ f$ witnesses that $X \cup Y$ is Dedekind infinite. \square

Proof of Lemma 43. For the first assertion, suppose the antecedent. Reasoning deductively, suppose now that $|M| > 4$, then $|X|, |M - X| > 2$. Set $X = A$ and $M - X = B$, then for each $(a, b) \in A \times B$, set $A_b = (A - \{a\}) \cup \{b\}$ and $B_a = (B - \{b\}) \cup \{a\}$. Now for $(a, b), (c, d) \in A \times B$ distinct (as a pair), we have

$$\textcircled{C}A_b = \textcircled{C}B_a \quad \textcircled{C}A_d = \textcircled{C}B_c \quad \textcircled{C}A_b \neq \textcircled{C}A_c \quad (138)$$

So $f(x, y) = \textcircled{C}A_y = \textcircled{C}B_x$ defines an injection from $A \times B$ into $M = A \cup B$, contradicting Proposition 44. Thus, $|M| \leq 4$.

For the second assertion, suppose $|M| > 2$. Then $f(x) = \ell\{x\}$ shows that for $\text{rng}(\ell)_1 = \{\ell X \mid |X| = 1\}$, $|\text{rng}(\ell)_1| = |M|$ and so by finiteness $\text{rng}(\ell)_1 = M$. However for any $m \in M$, $\ell\{m\} \neq \ell\emptyset$ by LCP so $\ell\emptyset \notin \text{rng}(\ell)_1 = M$, a contradiction. \square

Lemma 43 would make it appear that $\text{LCP}(= X)$ must have mostly structures of infinite size. But it has *no* structures of infinite size:

Proposition 45. In the presence of **IPM**, $\text{LCP}(= X)$ implies the universe is Dedekind finite.

Proof. Given **IPM**, if $|M| \geq \omega$, there is a concept X such that $|X| = |M - X| = |M|$. We show that $\text{LCP}(= X) \wedge |M| \geq \omega$ translates $\text{BLV}(= X)$, so then by Corollary 40 and Proposition 37, the former is inconsistent.

Let $f : M \rightarrow X$ and $g : M \rightarrow M - X$ be bijections, and define ε by

$$\varepsilon Y = \begin{cases} f(\ell Y) & \ell\emptyset \in Y \\ g(\ell Y) & \ell\emptyset \notin Y \end{cases}$$

Clearly ε is functional at X . Now if $Y, Z \in [X]_{\Box}$, $\ell\emptyset \in Y, Z$ and $Y \neq Z$, then by $\text{LCP}(= X)$ and the injectivity of f , $\varepsilon Y \neq \varepsilon Z$. If $\ell\emptyset \in Y - Z$, then as $f(M) \cap g(M) = \emptyset$, $\varepsilon Y \neq \varepsilon Z$. Lastly if $\ell\emptyset \notin Y, Z$, then by $\text{LCP}(= X)$ and the injectivity of g , $\varepsilon Y \neq \varepsilon Z$. \square

Together with Lemmata 39 and 43, from Proposition 45 we obtain

Corollary 46. If $\mathcal{M} \models |X| = |M - X|$ and $E(\Box)_X$ refines a complementation, then $\mathcal{M} \models A_E \rightarrow |M| \leq 4$.

Proof. Propositions 43 and 45 together give that

$$\mathcal{M} \models |X| = |M - X| \geq 4 \rightarrow \neg \text{LCP}(= X) \quad (139)$$

If $\mathcal{M} \models |X| < \omega$ and $E(\boxminus)_X$ refines a complementation, then by Theorem 7 $\mathcal{M} \models \text{LCP}(= X)$ if $\mathcal{M} \models A_E$. But then (139) gives that $|X| = |M - X| < 4$.

On the contrary if $\mathcal{M} \models |X| \geq \omega$, then as $E(\boxminus)_X$ refines a complementation, \square

6.4 Applying the bicardinal classification

Having dealt individually with each kind of non-trivial equivalence relation, we now apply Theorem 7, for putting together Corollaries 42 and 46 with Theorem 7 gives the following:

Theorem 47. If $\mathcal{M} \models 2 < |X| = |M - X|$ and $E(\boxminus)_X$ is nontrivial, then $\mathcal{M} \not\models A_E$.

Proof. By Theorem 7, if $E(\boxminus)_X$ is not trivial it either refines a separation or refines a complementation. In the former case A_E is inconsistent by Corollary 42, and in the latter case A_E is inconsistent by Corollary 46 and by our assumption that $|M| > 4$. \square

Remark 48. The converse of Theorem 47 fails, for let A_E be the abstraction principle **NewV**:

$$(\forall X, Y)(\varsigma X = \varsigma Y \leftrightarrow (|X| = |Y| = |M| \vee (|X|, |Y| < |M| \wedge X = Y)))$$

But by König's theorem, **NewV** has no standard models whose first-order domain is a singular limit cardinal.¹⁵

To put Theorem 47 in the terms used earlier: If M is large enough (having greater than four elements), then \mathcal{M} satisfies no abstraction principle whose equivalence relation is non-trivial on concepts which evenly divide the universe. So Theorem 7 gives a mark by which to identify (at least) many bad companions: these insufferable principles limit the size of the universe by non-trivially carving concepts of maximally large bicardinality.¹⁶

¹⁵There is in fact a small error on just this point at [24, p. 596] in stating the results of [20, p. 315]: “For instance, the claim that the abstraction principle New V is strongly stable is equivalent to the generalized continuum hypothesis.” The correct statement is this: **New V** is not strongly stable, and whether it is stable depends on whether $2^\kappa = \kappa^+$ for unboundedly many κ .

¹⁶Theorem 47 is in some ways a companion to Fine's *Characterization* Theorem, mentioned earlier: that what he calls the *basal* abstraction principle is the finest satisfiable in all infinite domains. As Theorem 7 is a deductive analogue of Fine's Classification Theorem, a deductive analogue of his Characterization Theorem can also be obtained from Theorem 7, though we leave this for Appendix 10.

7 Classification and relative categoricity

In [24] Walsh and Ebels-Duggan introduce a criterion that distinguishes the abstraction principle HP. The so-called “Julius Caesar problem”¹⁷ arises from the following fact. There are abstraction principles, among them HP, with the following unfortunate property: for a given base L_0 -structure, it is possible to interpret ∂ in two ways, both satisfying the abstraction principle, but such that the two expanded structures are neither isomorphic nor elementarily equivalent. That this can be done with the equivalence relation \approx and HP was noted by Frege himself [8, § 56, §§66ff]. Walsh and Ebels-Duggan showed that though these abstraction principles lack *these* properties, some of them have weaker but still noteworthy equivalence properties.

Abusing notation somewhat, let $A_E[\partial]$ be not only the abstraction principle correlated with E , but the theory containing as axioms that abstraction principle and the axioms of our background logic (including those for well-behaved cardinalities). With similar abuse, we’ll use $A_E^2[\partial_1, \partial_2]$ be a theory containing axioms for our background logic and the abstraction principles $A_E[\partial_1]$ and $A_E[\partial_2]$. In other words, $A_E^2[\partial_1, \partial_2]$ is a theory with two copies of the abstraction principle A_E , one for each abstraction operator. To introduce Walsh and Ebels-Duggan’s weakened equivalence notions we first introduce the relevant notion of an isomorphism between induced models:

Definition 49. Given a structure $\mathcal{M} = (M, S_1[M], S_2[M], \dots, \partial_1, \partial_2)$ satisfying A_E^2 , for $i = 1, 2$, let

$$\mathcal{M}_i = (\text{rng}(\partial_i), S_1[M] \cap P(\text{rng}(\partial_i)), S_2[M] \cap P(\text{rng}(\partial_i) \times \text{rng}(\partial_i)) \dots, \partial_i) \quad (140)$$

We say that \mathcal{M}_i is the model *induced* by ∂_i .

As noted in [24], for Γ to be an isomorphism between \mathcal{M}_1 and \mathcal{M}_2 it is sufficient for Γ to satisfy the following condition: For all $X \in S_1[M] \cap P(\text{rng}(\partial_1))$,

$$\Gamma \partial_1 X = \partial_2(\Gamma X) \quad (141)$$

Definition 50. The theory A_E is *naturally relatively categorical* just if for any model \mathcal{M} , and any ∂_1, ∂_2 such that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$, the *natural bijection* $\Gamma : \text{rng}(\partial_1) \rightarrow \text{rng}(\partial_2)$, defined by

$$\Gamma \partial_1 X = \partial_2 X \quad (142)$$

is an isomorphism between the induced models \mathcal{M}_1 and \mathcal{M}_2 .

The theory A_E is *relatively categorical* just if for any model \mathcal{M} , and any ∂_1, ∂_2 such that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$, $\mathcal{M}_1 \cong \mathcal{M}_2$.

The theory A_E is *relatively elementarily equivalent* just if for any model \mathcal{M} , and any ∂_1, ∂_2 such that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$,

$$\mathcal{M}_1 \models \varphi_{\partial_1} \Leftrightarrow \mathcal{M}_2 \models \varphi_{\partial_2}$$

for every *sentence* (i.e., closed formula) φ in the language of the theory $A_E[\partial]$, and φ_{∂_i} is the result of replacing every occurrence of ∂ in φ with ∂_i .

¹⁷See [8, § 55], and for example [13] and [11, Chapter 14].

Thus natural relative categoricity is distinguished from relative categoricity in that the former requires that the natural bijection be an isomorphism, the latter only requires that an isomorphism exist.

One of the main theorems of [24] established necessary and sufficient conditions for the *natural* relative categoricity of an abstraction principle; some of which we state as:

Natural Relative Categoricity Theorem NRCT ([24]). Let A_E be an abstraction principle. The following are equivalent:

1. A_E is naturally relatively categorical.
2. $A_E[\partial] \models (\forall X, Y)(|Y| = |X| \leq |\text{rng}(\partial)| \rightarrow E(X, Y))$.

where again “ \models ” means the deductive consequence relation for our strong background logic. Abstraction principles satisfying this second condition are said to be *cardinality coarsening on abstracts*.

As can be seen plainly, \approx is the finest equivalence relation such that A_E is cardinality coarsening on abstracts; thus HP is the finest naturally relatively categorical abstraction principle. This gives the neo-logician a criterion by which to distinguish HP from other abstraction principles; notably, the abstraction principle **NewV**, which is equivalent to $\text{BLV}(<|M|)$, is *not* naturally relatively categorical (see [24, § 5]).

But there is more than can be asked. In [23, Proposition 14, p. 1687] Walsh shows that HP is relatively categorical in the stronger, unqualified sense. Moreover the foregoing establishes the following implication relations:

$$\begin{array}{ccc}
\text{Natural relative categoricity} & \Leftrightarrow & \text{Cardinality coarsening on abstracts} \\
\downarrow & & \\
\text{Relative categoricity} & & \\
\downarrow & & \\
\text{Relative elementary equivalence} & &
\end{array} \tag{143}$$

Walsh and Ebels-Duggan thus raised two questions in [24]:

Question 51. Are the conditions for relative categoricity the same as for natural relative categoricity? Can we drop the specification of the natural bijection in the **NRCT**? More precisely is (2) equivalent to the relative categoricity of A_E ?

Question 52. What is the relation between relative categoricity, natural relative categoricity, and relative elementary equivalence?

It is a consequence of Theorem 7 that Question 51 can be answered in the affirmative, and that, in answer to Question 52, all the implication arrows of (143) can be reversed.

From here onward it will be helpful to recall the definitions of restricted abstraction principles from page 26.

Lemma 53. Suppose A_E is an abstraction principle, and that $A_E[\partial]$ translates $\text{BLV}(\leq 1)$ via $\varphi(Y, Z)$. Then $A_E[\partial]$ is not relatively elementarily equivalent, and so not relatively categorical.

Proof. Let ∂_1 be given such that $\mathcal{M} \models A_E[\partial_1]$. Note first that as A_E translates $\text{BLV}(\leq 1)$, $|\text{rng}(\partial)| = |M| \geq \omega$. We may thus assume that $\text{rng}(\partial_1) = M$. The ∂_2 we will construct will have the same range; this allow us to avoid worries that the translation φ of $\text{BLV}(\leq 1)$ in \mathcal{M} , which may contain nested quantifiers, may change its behavior on the induced models \mathcal{M}_1 and \mathcal{M}_2 .

For φ translating as above, let $x \in U(X)$ if and only if $\varphi(X, Y)$ and $x \in Y$. Consider now

$$\psi_i = (\exists a)(a = \partial_i U(\{a\})) \quad (144)$$

On the one hand, if $\mathcal{M} \models \neg\psi_1$, then select $a \in M$, and note that there are $b, c \in M$ with $b = \partial_1 U(\{a\})$ and $a = \partial_1 U(\{c\})$. Let $f(b) = a$, $f(a) = b$, and f the identity map on $M - \{a, b\}$. As $f \in \mathcal{M}$ since it is definable, setting $\partial_2 X = f(\partial_1 X)$ we have that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$. And yet $\mathcal{M}_2 \models \psi_2$.

On the other hand, if $\mathcal{M} \models \psi_1$, let

$$A = \{\partial_1 X \mid (\exists a)(a = \partial_1 X = \partial_1 U(\{a\}))\} \quad (145)$$

Suppose first that $|A| = |\text{rng}(\partial_1)|$. As $|A| \geq \omega$, there is a $B \subset A$ such that $|B| = |A - B| = |A|$ by **IPM**; let $g : B \rightarrow A - B$ be a bijection. Then let

$$\partial_2 X = \begin{cases} g(\partial_1 X) & \partial_1 X \in B \\ g^{-1}(\partial_1 X) & \partial_1 X \in g(B) \\ \partial_1 X & \partial_1 X \notin A \end{cases} \quad (146)$$

Again we have that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$, and $\mathcal{M}_2 \models \neg\psi_2$: for if $\partial_1 U(\{a\}) = a \in B$ then $\partial_2 U(\{a\}) \in A - B$, and so $\partial_2 U(\{a\}) \neq a$. And likewise if $\partial_1 U(\{a\}) = a \in A - B$.

Lastly suppose $|A| < |\text{rng}(\partial_1)|$, then as $|\text{rng}(\partial_1)| \geq \omega$, then by **ISM** there is an injection $g : A \rightarrow \text{rng}(\partial_1) - A$. Setting $B = A$, the construction as in (146) again delivers the verdict that $\mathcal{M}_2 \models \neg\psi_2$. \square

Lastly, we prove a parallel to Lemma 39.

Lemma 54. Suppose $\mathcal{M} \models 0 < |X| < \omega \leq |M|$ and $E(\boxminus)_X$ refines a separation. Then A_E translates $\text{BLV}(\leq 1)$.

Proof. By Proposition 38, for any singleton Y , $\langle Y, X \rangle \in [X]_{\boxminus}$. Further, if $Y \neq Z$ and both are singletons, then $\langle Y, X \rangle$ and $\langle Z, X \rangle$ are disjoint, so $|\langle Y, X \rangle \triangle \langle Z, X \rangle| \not\leq |X|$. Thus, setting $\varepsilon Y = \partial \langle Y, X \rangle$, we have for singletons Y and Z :

$$\varepsilon Y = \varepsilon Z \Leftrightarrow E(\langle Y, X \rangle, \langle Z, X \rangle) \Leftrightarrow \langle Y, X \rangle = \langle Z, X \rangle \Leftrightarrow Y = Z$$

\square

We now answer Questions 51 and 52:

Theorem 55. An abstraction principle A_E is relatively categorical if and only if it is cardinality coarsening on abstracts.

Further, an abstraction principle is relatively elementarily equivalent if and only if it is relatively categorical.

Proof. The right-to-left direction of the second assertion is trivial; for the right-to-left direction of the first, suppose A_E is cardinality coarsening on abstracts, and let $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$. Let Γ be the natural bijection defined by $\Gamma\partial_1 X = \partial_2 X$; note that as $|\Gamma X| = |X|$, since A_E is cardinality coarsening on abstracts, in $A_E^2[\partial_1, \partial_2]$ it follows that $\partial_i X = \partial_i \Gamma X$. Thus

$$\Gamma\partial_1 X = \partial_2 X = \partial_2 \Gamma X \quad (147)$$

so Γ is an isomorphism.¹⁸

For the left-to-right direction of both assertions, observe that the following exhaust the possibilities for the cardinal placement of any two equinumerous concepts:

$$\mathcal{M} \models |M| = |Y| = |X| = |M - X| = |M - Y| \quad (148)$$

$$\mathcal{M} \models |M| = |Y| = |X| = |M - X| > |M - Y| \quad (149)$$

$$\mathcal{M} \models |M| = |Y| = |X| > |M - X| \geq |M - Y| \quad (150)$$

$$\mathcal{M} \models \omega \leq |X| = |Y| < |M| \quad (151)$$

$$\mathcal{M} \models |X| = |Y| < \omega \leq |M| \quad (152)$$

$$\mathcal{M} \models |X| = |Y| \leq |M| < \omega \quad (153)$$

To prove the second assertion, then it suffices to show that on each of these possibilities, if $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$ then either (i) \mathcal{M}_1 and \mathcal{M}_2 are isomorphic (and so elementarily equivalent), or (ii) there are ∂_1, ∂_2 such that \mathcal{M}_1 and \mathcal{M}_2 are not elementarily equivalent (and so not isomorphic). From the right-to-left direction of the first assertion, we know that if E is cardinality coarsening on abstracts, then (i) holds for all possibilities. Thus to prove the second assertion, it suffices to prove that if E is not cardinality coarsening on abstracts, then either $\mathcal{M} \not\models A_E$, or there are ∂_1, ∂_2 with $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$ but \mathcal{M}_1 and \mathcal{M}_2 are not elementarily equivalent. This will prove the first assertion as well.

So suppose that A_E is not cardinality coarsening on abstracts; let $\mathcal{M} \models A_E$ witness this failure. So there are $X, Y \in \mathcal{M}$ such that

$$\mathcal{M} \models |Y| = |X| \leq |\text{rng}(\partial_1)| \wedge \neg E(X, Y) \quad (154)$$

Clearly then $\mathcal{M} \models 0 < |X|$.

The first case in which (148) holds can be ruled out since if $E(\equiv)_X$ is non-trivial, then by Corollaries 42 and 46, $\mathcal{M} \not\models A_E$. Thus for the second case (149) we must assume that $E(\equiv)_X$ is trivial. Without loss of generality assume $\mathcal{M} \models X \subset Y$. Since by assumption (154), $\mathcal{M} \models |M| = |X| = |Y| \leq |\text{rng}(\partial)|$, let $f, g \in \mathcal{M}$ be respective witnesses of $|X| = |\text{rng}(\partial)|$ and $|Y| = |\text{rng}(\partial)|$, then

¹⁸This direction of the proof generalizes the proof in [23, Proposition 14].

set $\partial_1 = f \circ \partial$ and $\partial_2 = g \circ \partial$. Since ∂_1, ∂_2 are both definable injections in \mathcal{M} , it follows that $\mathcal{M} \models A_E^2[\partial_1, \partial_2]$. Note, however, that for all $Z \subset X$, if $|Z| = |X|$ then $Z \sqsupseteq X$. Hence, since $E(\sqsupseteq)_X$ is trivial, $E(X, Z)$. Thus, as

$$\mathcal{M}_1 \models (\forall x)(Xx \leftrightarrow x = x)$$

it follows that

$$\mathcal{M}_1 \models (\forall U, W)((\forall x)(Ux \leftrightarrow x = x) \wedge |W| = |U| \rightarrow \partial_1 U = \partial_1 W) \quad (155)$$

However, notice that

$$\mathcal{M}_2 \models (\forall x)(Yx \leftrightarrow x = x)$$

and that $f \circ g \in S_2[M]$ and $f \circ g \subset Y \times Y$. Thus $f \circ g \in \mathcal{M}_2$, so

$$\mathcal{M}_2 \models |X| = |Y| \wedge \partial_2 X \neq \partial_2 Y$$

as by assumption $X \subset Y$ and $\mathcal{M} \models A_E[\partial_2] \wedge \neg E(X, Y)$. Consequently, we see that

$$\mathcal{M}_2 \not\models (\forall U, W)((\forall x)(Ux \leftrightarrow x = x) \wedge |W| = |U| \rightarrow \partial_2 U = \partial_2 W)$$

So by (155) A_E is not relatively elementarily equivalent, and so not relatively categorical.

For (150) observe that by **IPM** if $|M| = |X| \geq \omega$ then there is $Z \subset X$ such that $|Z| = |M - Z| = |M|$; by (148) $E(\sqsupseteq)_Z$ is trivial, and so by (149) both X and Y are in $[Z]_E$.

If (151) holds, then by Theorem 7, $E(\sqsupseteq)_X$ is non-trivial, but then by Lemma 39 A_E translates $\text{BLV}(\leq X)$ on \mathcal{M} . Likewise if (152), then by Lemma 54 we see again that A_E translates $\text{BLV}(\leq 1)$ on \mathcal{M} . But then by Lemma 53 A_E is not relatively elementarily equivalent, and so not relatively categorical.

Finally if (153) holds then by Theorem 7, as $0 < |X| < |M|$, by Theorem 47 and $\mathcal{M} \models A_E$, we have that $|M| \leq 4$ and $\mathcal{M} \models \text{LCP}(= X)$. If $|M| = 2$ then if $E(\sqsupseteq)_X$ refines a complementation then it is trivial (and so cardinality coarsening on abstracts), so $|M| = 4$. But then by the argument of [24, § 5.5, p. 594], A_E is not relatively elementarily equivalent, and so not relatively categorical. \square

Walsh and Ebels-Duggan prove another theorem [24, Theorem 1.2] for another weakened equivalence property: that if all objects are abstracts, then the induced models for A_E are isomorphic via the natural bijection if and only if bicardinally equivalent concepts are E -equivalent. They ask an analogous question as well: can one drop the assumption that the natural bijection is an isomorphism and still obtain the biconditional? The proof of Theorem 55 indicates an affirmative answer here also: to adapt the proof one must observe that (149) is irrelevant on the assumption that X and Y are bicardinally, but not E -, equivalent. The other cases remain the same.

Thus we have:

Theorem 56. Suppose that $\mathcal{M} \models A_E^2[\partial_1, \partial_2] \wedge \text{rng}(\partial_1) = \text{rng}(\partial_2) = M$. Then $\mathcal{M}_1 \cong \mathcal{M}_2$ if and only if A_E is bicardinality coarsening on abstracts.

It may seem surprising to see in Theorem 55 the equivalence of relative elementary equivalence with relative categoricity. But reflection suggests that this shouldn't be so striking after all: for Theorem 55 depends on Theorem 7, and so on the equivalence relations in play being permutation invariant because L_0 -definable. The equivalence is in a sense an artifact of this restriction. It is an open question as to whether other invariance conditions (see [7], [1]), [5]) would yield the same results.

However, reflection on this equivalence could be taken as philosophically relevant, and indeed good news for the neo-logicist. For it is trivial, by PI, that in an L_0 -structure \mathcal{M} , if π is a permutation in \mathcal{M} , then the structure resulting from that permutation is both isomorphic and elementarily equivalent to \mathcal{M} . This is trivial, of course, because the permuted structure just is the original structure. But it is noteworthy that this triviality hides the coinstantiation of isomorphism and elementary equivalence.

In light of this, one might ask: is this coinstantiation preserved under abstraction? In more detail: given two abstraction operators, ∂_1, ∂_2 for an abstraction principle A_E , the natural bijection $\Gamma(\partial_1 X) = \partial_2$ will always be in any model of $A_E[\partial_1, \partial_2]$. Therefore if ∂_1 and ∂_2 share the same range, then \mathcal{M}_2 can be regarded as the structure resulting from the permutation Γ of the structure \mathcal{M}_1 . Is it the case that under any such permutation, the resulting models will be isomorphic if and only if elementarily equivalent? Indeed, answers Theorem 55, the answer is affirmative: this property of coinstantiation is, in the given sense, preserved under permutations of abstractions. One might argue on behalf of the neo-logicist (though we will not) that this shows abstraction is in some sense a “logical” operation.

8 Conclusion

We began with the promise that strengthening Fine's Classification Theorem would advance our understanding of the demarcation of logical abstraction principles. The strengthened version of Fine's theorem yields a strengthened relative categoricity theorem; the relevant advancement thus comes in the option provided by the latter result.

Walsh and Ebels-Duggan note that relative categoricity as we have described it cannot rule out bad companions like NP: since NP has only Dedekind finite models, it (the *principle*) is cardinality coarsening on abstracts. They also note, however, that the equivalence relation underlying NP is not cardinality coarsening *on small concepts*—a notion that applies not to abstraction principles but to equivalence relations. (See [24, section 5.4].)

The confluence of results presented in this paper offers a stronger statement of this point. As we can see from the results in sections 6 and 7, when nontrivial equivalence relations are manifest in a structure, they generate failures of relative categoricity. It *is* curious that NP is relatively categorical as a principle. But it is more relevant that the equivalence relation deployed by NP is nontrivial on all infinite bicardinal slices—it has non-trivial manifestations. And wherever

nontriviality appears in a structure satisfying an associated abstraction principle, it generates a failure of relative categoricity. Thus, even though NP is by the letter relatively categorical, it deploys an equivalence relation that generates failures of relative categoricity. On this line of thinking, NP can thus be ruled out. The same goes for other bad companions.

This is a sketchy argument for several reasons, but for our purposes it needn't be more. As we said at the outset, our goal is to provide and clarify options. All things considered, this sort of account may be rejected. But the results here adduced, along with those in [24] and [23], indicate that the sort of answer just offered is one of the things that should be considered.

In light of this, it may be worth considering relative categoricity as kin to permutation invariance. Both are motivated by the idea that logic is indifferent to the particulars of objects. Permutations represent this indifference by exchanging objects arbitrarily, while relative categoricity represents it by treating of arbitrarily selected abstraction assignments. Here, more issues loom. Our discussion has been organized around permutation invariance, but there are stronger conditions claiming to be necessary for logicality. Permutation invariance is *intra-modular*: permutations simply re-order or re-organize the elements of a given structure. Recent work has focused on the *trans-modular* notion of *isomorphism between structures*. This is arguably a better mark of the logical than permutation invariance, and better matches the intuitive notion of indifference to objects. Though the language L_0 is logical in this stronger sense, we have not addressed how a notion like relative categoricity would apply in a

trans-modular setting.¹⁹

One last point is worth making. The results of section 6 can be recast. Say

¹⁹Jack Woods, in [25], has addressed the question of *indefinite* abstraction principles in a trans-modular setting. Given a domain D and an equivalence relation E on $P(D)$, let

$$F(D)_E = \{f : P(D) \rightarrow D \mid (\forall X, Y \subset D)(fX = fY \Leftrightarrow E(X, Y))\}$$

In this setting of [25], the members of $F(D)_E$ are called *abstraction functions* for E , while an *abstraction operator* for E is a function σ taking domains D as an argument, outputting a set $\sigma^D \subseteq F(D)_E$. If $\zeta : D \rightarrow D'$ is a bijection, then ζ can be extended to an isomorphism ζ^+ of all types over D and D' . As such, $\zeta^+(\sigma^D)$ will be $\{\zeta^+ \circ f \circ (\zeta^+)^{-1} \mid f \in \sigma^D\}$ (see [25, pp. 281ff]). The abstraction operator σ is then *isomorphism invariant* just if $\zeta^+(\sigma^D) = \sigma^{D'}$ for all D, D' , and bijections $\zeta : D \rightarrow D'$. An abstraction operator is said to be *full* just if for any D , $\sigma^D = F(D)_E$. Woods endorses the view that indifference to particular objects is the mark of the logical, and thus that isomorphism invariance is the mark of the logical.

Woods offers as evidence for this claim the proposition on [25, p. 298] that all and only abstraction operators logical in the sense of isomorphism invariance are both full and associated with an isomorphism invariant (in a broader sense) equivalence relation. However, this proposition, and the lemma used to prove it (presented on [25, p. 296]), are incorrect. The lemma in question claims that if σ is (non-empty and) invariant then it is full; but the proof makes two incorrect assumptions: the first is that if $f, g \in F(D)_E$ and $\text{rng}(f) \approx \text{rng}(g)$ then $M - \text{rng}(f) \approx M - \text{rng}(g)$ [25, p. 296]. The second is that if $\zeta(f(A)) = g(A)$ for $f, g \in F(D)_E$ and all $A \subseteq D$, then $\zeta^+(f) = g$. But this is not in general true, for the extension of ζ to ζ^+ has $\zeta^+(f) = \zeta^+ \circ f \circ (\zeta^+)^{-1}$, and while this implies that $\zeta^+(f)(A) = g((\zeta^+)^{-1}(A))$, the last $g((\zeta^+)^{-1}(A)) = g(A)$ holds only, in general, when E is coarser than \approx ; see [24, Theorems 1.1 and 1.2] and Theorem 55.

The lemma and resulting proposition can be shown false by the example of \approx for E and $\sigma^D = \{f \in F(D)_E \mid \text{rng}(f) = D\}$. Clearly this is not full, and it can be shown isomorphism invariant using the techniques of the lemma on [25, p. 297]. This would suggest suitable restrictions on the cardinality or identity of the range of the abstraction functions might repair the lemma and proposition. But more robust counterexamples can be generated by the results of the present paper. For these we will first need the following observation.

Suppose $\mathcal{M}[\partial_1, \partial_2]$ with first-order domain M is standard and witnesses the fact that A_E is not relatively categorical. Then if $\zeta^+ : \mathcal{M} \rightarrow \mathcal{M}'$ is an isomorphism, then $\mathcal{M}'[\zeta^+(\partial_1), \zeta^+(\partial_2)]$ also witnesses that A_E is not relatively categorical, for $\mathcal{M}_i \cong \mathcal{M}'_i$ for $i = 1, 2$.

The main idea of the counterexample is to choose a pair ∂_1, ∂_2 of surjective abstraction functions such that $\mathcal{M}_1 = \mathcal{M}[\partial_1] \not\cong \mathcal{M}_2 = \mathcal{M}[\partial_2]$. Then for M the domain of \mathcal{M} , we let σ^M be the set of all abstraction functions ∂ such that $\mathcal{M}[\partial]$ is isomorphic to $\mathcal{M}[\partial_1]$. Clearly ∂_2 will not be among these. Then, for every domain D , we set σ^D to be the set containing $\zeta^+(\partial)$ for each $\partial \in \sigma^M$ and each bijection $\zeta : M \rightarrow D$. This ensures σ will be isomorphism invariant, but also ensures, by the observation given above, that it will not be full.

More formally: it is a consequence of Theorem 56 that there are abstraction principles with surjective abstraction functions ∂_1, ∂_2 on a domain M such that for no bijection $\zeta : M \rightarrow D$ is there a bijection $\pi : D \rightarrow D$ such that

$$\pi(\zeta^+(\partial_1)(A)) = \zeta^+(\partial_2)(\pi(A))$$

(That is, the bijection π does not commute with $\zeta^+(\partial_1)$ and $\zeta^+(\partial_2)$.) This is true of **NewV**, as demonstrated by Theorem 56 in conjunction with the results of [24, Section 5.2].

Thus for each $D \approx M$, set

$$\sigma^D = \{\partial \in F(D)_E \mid \partial = \zeta^+ \circ \partial_1 \circ (\zeta^+)^{-1} \text{ for some bijection } \zeta : M \rightarrow D\}$$

The indefinite abstraction operator σ is then isomorphism invariant by construction, but $\zeta^+(\partial_2) \notin \sigma^{\zeta(D)}$ for any bijection ζ , so σ is not full. Applied to the cited case of **NewV** developed in [24], we might choose ∂_1 such that the natural membership relation derived from it is well-founded. Thus for any D and any $f \in \sigma^D$, f will foster such well-founded relations as well. But for each D there will always be $\partial_2 \in F(D)_E$ such that the natural membership relation is not well-founded, and thus such ∂_2 will be omitted from each σ^D .

that A_E proves the universe set-like—in symbols, $A_E \models \text{ZFC}(M)$ —just if **CC**, **ISM**, **IPM**, and every instance of **CWF** is provable in the theory A_E . The results of section 6, put into this context, say that if A_E proves the universe set-like, and if E is provably non-trivial at “large enough” bicardinalities, then A_E is inconsistent. For if $A_E \models \text{ZFC}(M)$ and E is non-trivial at a large enough cardinality, then since $\text{ZFC}(M)$, at that cardinal E is either separative or complementative. Then following the theorems of that section, **IPM** renders the principle inconsistent. This can be summed up by:

Corollary 57. If a theory A_E proves $|M| > 4$ then A_E is inconsistent (without the help of our cardinality principles) if and only if A_E proves that there is a well-ordering of M and $[X]_{E(\Box)}$ is non-trivial for some X with $|X| = |M - X| = |M|$.

The non-trivial direction follows because our cardinality principles follow from a well-ordering of the universe.

Similarly we obtain necessary and sufficient conditions for satisfiability (possession of a standard model):

Corollary 58. If a theory A_E proves $|M| > 4$ then A_E is unsatisfiable if and only if $[X]_{E(\Box)}$ is non-trivial for some X with $|X| = |M - X| = |M|$.

The non-trivial direction again follows by our results, since the cardinality principles are true in all standard infinite models.

These are not, however, the best results available, since a weaker condition on the right-hand side may also be sufficient for inconsistency (respectively, unsatisfiability). We can already obtain such improvements understanding “large enough” to mean not “universe-sized”, but “exponentially large” in the sense of Fine (see Appendix 10). That is, replace “for some X with $|X| = |M - X| = |M|$ ” on the right with “for some X such that $TOP(X)$ ”, and the corollaries still hold.

In any case, as an explanation of the consistency and satisfiability profiles of abstraction principles it is worth considering.

Outside of the discussion on neo-logicism, we note that the classification theorems proved herein, being completely general, may have applications of interest in other research areas related to second-order logic. Having answered some questions, it is thus apt to raise one in conclusion. We showed in Theorem 6 that for concepts finite in \mathcal{M} , the conditional arrows “ \Rightarrow ” reverse for both separative and complementative bicardinal slices. Is this true for more than just finite concepts? In other words, are there L_0 -definable equivalence relations in which the arrow of implication does not reverse?

Though the lemma and proposition are false, the above counterexamples do not obviously determine the correctness of Woods’s overall assertion: that isomorphism invariance (for abstraction operators) is the mark of logic’s characteristic indifference to the particularities of objects. The second counterexample *does* depend on the choice of an abstraction function, but *not* on the choice of any first-order object. This accords with Woods’s prediction.

9 Appendix: The proof of Proposition 8

Proof. Suppose $|X| \not\leq |Y|$, so by Cardinal Comparability, for all $n \in \mathbb{N}$, $\mathcal{M} \models n \times |Y| < |X|$. Thus $1 \times |X| \geq |Y|$, so $|Y| \leq |X|$, establishing (34).

For $X \cap Z = Y \cap W = \emptyset$ and $|Z| \stackrel{g}{\leq} |W|$, suppose $|X| \leq |Y|$. Choose n, U , and f such that $|U| = n$ (in the metatheory), and $|X| \stackrel{f}{\leq} |U \times Y|$. Then choose $u \in U$, and then let

$$h(v) = \begin{cases} f(v) & v \in X \\ (u, g(v)) & v \in Z \end{cases} \quad (156)$$

The function h is an injection establishing (35).

Toward establishing (36), assume $|Z| \leq |W|$ and $|X \sqcup Z| \leq |Y \sqcup W|$, but $\neg(|Y \sqcup W| \leq |X \sqcup Z|)$. For a contradiction, assume $\neg(|X| \triangleleft |Y|)$. By (34), $|Y| \leq |X|$. But then since $|W| \leq |Z|$, by (35) we have $|Y \sqcup W| \leq |X \sqcup Z|$, contradicting our assumption.

For (37), choose n, m such that $|X| \stackrel{f}{\leq} n \times |Y|$ and $|Y| \stackrel{g}{\leq} m \times |Z|$. Then define $h : X \rightarrow nm(n+1) \times Z$ by

$$h(x) = (nmf_1(x) + g_1(f_1(x)), g_2(f_2(x))) \quad (157)$$

where f_1 and g_1 output the left value of f and g respectively, and f_2 and g_2 output the right value of f and g , respectively. Verifying that h is injective is routine, $h \in \mathcal{M}$ by comprehension.

To establish (38), suppose $\mathcal{M} \models X \stackrel{f}{\leq} n \times (Y \sqcup Z)$. By (34), either $Y \leq Z$ or $Z \leq Y$; assume the former, letting g witness the injection in \mathcal{M} . Then

$$h(x) = \begin{cases} f(x) & f_2(x) \in Z \\ (n + f_1(x), g_2(f_2(x))) & f_2(x) \in Y \end{cases} \quad (158)$$

injects X into $2n \times Z$; $h \in \mathcal{M}$ by comprehension. So $X \leq Z$; a similar argument shows that if $Z \leq Y$ then $X \leq Y$. Note that (39) follows immediately.

For (40), suppose $\mathcal{M} \models |Y| \stackrel{f}{\leq} n \times |X|$ and $\mathcal{M} \models |Z| \stackrel{g}{\leq} m \times |X|$. Then

$$h(u) = \begin{cases} f(u) & u \in Y \\ (n + i, x) & u \in Z, g(u) = (i, x) \end{cases} \quad (159)$$

injects $|Y \sqcup Z|$ into $(n + m) \times |X|$ in \mathcal{M} . So if $|X| \triangleleft |Y \sqcup Z|$, $|Y \sqcup Z| \not\leq |X|$, and so $|Y| \not\leq |X|$ or $|Z| \not\leq |X|$. By (34) this implies that either $|X| \triangleleft |Y|$ or $|X| \triangleleft |Z|$ holds. Note that (41) follows immediately.

To establish (42), assume $|Y \sqcup Z| \stackrel{f}{\leq} |X|$, then $f \upharpoonright Y$ and $f \upharpoonright Z$ are the required injections. If $|X| \leq |Y|$ then by (35) and (37) $|X| \leq |Y \sqcup Z|$; so (43) follows from (42).

For (44), assume $|X| \triangleleft |Y|$, so by the definition of \triangleleft , we have that $|Y| \not\leq |X|$. Thus by (42), $|Y \sqcup Z| \not\leq |X|$. Then by (34), $|X| \triangleleft |Y \sqcup Z|$.

For (45), assume $|X \sqcup Y| \overset{f}{\triangleleft} |Z|$ and $|W| \overset{g}{=} |Y|$. Then

$$h(u) = \begin{cases} f(u) & u \in X \\ f(g(u)) & u \in W \end{cases} \quad (160)$$

witnesses $|X \sqcup W| \triangleleft |Z|$. A similar argument establishes (46). Then (47) follows from (45), (46) and (34), as does (48). \square

10 Appendix: Fine's Characterization Theorem

We have noted throughout that our Main Theorem is a version of Fine's Classification Theorem; additionally Theorem 47 bears a striking similarity to Fine's Characterization Theorem [7, Theorem 6, p. 144]. In this last appendix we show how to present Fine's Characterization Theorem in the deductive setting of our paper.

Following Fine [7, p. 143], say that a concept is *exponentially large* in \mathcal{M} just its subconcepts outnumber the objects of M . For short-hand, now say that a concept X is "Top" just if X and its complement are both exponentially large. The *basal* equivalence relation $E_0(X, Y)$ is the equivalence relation such that for any \mathcal{M} and $X, Y \in \mathcal{M}$, $E(X, Y)$ holds (in the metatheory)²⁰ just if either both X and Y are Top and $\mathcal{M} \models X \boxplus Y$, or neither X nor Y is Top and $\mathcal{M} \models X = Y$.

Fine's Characterization Theorem. The basal relation E_0 is the finest equivalence relation satisfying Permutation Invariance such that for any infinite *standard* model \mathcal{M} , $M \models A_{E_0}$.

The most noticable differences between Fine's Characterization Theorem and Theorem 47 are that the latter concerns only infinite standard models, while the former addresses all models with well-behaved cardinalities. Fine's proof of the Characterization theorem uses his Classification theorem. As we have stressed, this classification theorem is a version of what we have proved as our Main Theorem. But it is hard to see how, given that Fine's terminology doesn't neatly capture the array of possibilities. But some reflection, with Fine's suggested aid of a Venn Diagram, shows that the Fine's Classification theorem can be restated as follows.

Fine's Classification Theorem (Restated). For X infinite and \mathcal{M} standard:

- (1) If $|X| < |M|$ and $E(\boxplus)_X$ is does not refine a separation, then $E(\boxplus)_X$ is trivial.
- (2) If $|X| = |M| > |M - (X \cup Y)|$ and $E(\boxplus)_X$ refines neither a separation nor a complementation, then $E(\boxplus)_X$ is trivial.
- (3) If $|X| = |M - X|$ and $E(\boxplus)_X$ refines neither a separation nor a complementation, then $E(\boxplus)_X$ is trivial.

²⁰Fine's results are cast in terms of partitions of $P(M)$, so at this point we won't talk about \mathcal{M} satisfying $E_0(X, Y)$, since Fine's presentation doesn't ensure that E_0 is expressible. We will show how to express something like E_0 in L_0 below, see Definition 60.

And this yields as a corollary our Main Theorem, restricted to infinite standard models: that $E(\boxminus)_X$ must satisfy at least one of being trivial, refining a separation, or refining a complementation.

Our Theorem 7 can thus be regarded as sharpening and expanding Fine’s achievement in his Classification Theorem, except for one easily ironed wrinkle:

Remark 59. In the statement of our theorem: our main concern has been with identifying L_0 -definability as a necessary condition on an equivalence relation’s being logical, so our conditions on E have been that it be L_0 -definable. However, inspection of the proofs of our results will show that L_0 -definability is used only in invoking Permutation Invariance. Thus all of our results can be put as Fine’s are: about equivalence relations satisfying Permutation Invariance.

As we will show below, Fine’s Characterization Theorem can also be sharpened and expanded, with the help of Theorem 7. We prove the Characterization Theorem in the deductive setting of Theorem 7.

To do so, we first need to use L_0 , rather than the metatheory, to describe concepts being exponentially large, “Top”, and so to describe the basal equivalence relation E_0 . Given a relation $R(x, y)$, we let $R[x] = \{y \mid R(x, y)\}$. If $R \in \mathcal{M}$ then so is $R[x]$ by comprehension. Using this we can express, in L_0 , the claim that R “injects” from equivalence classes of subconcepts of a given concept W to objects with the following formula (see also [24, p. 588], [19, p. 105]):

$$(\forall U \subseteq W)(\exists x)(R[x] \subseteq U \wedge E(R[x], U)) \wedge (\forall x, y)(\neg E(R[x], R[y]) \rightarrow x \neq y) \quad (161)$$

We will use the expression “ $\left| \frac{S_1[M] \upharpoonright W}{E} \right| \stackrel{R}{\leq} |M|$ ” to abbreviate (161).²¹ We now have the material required for expressing the basal equivalence relation:

²¹It is worth noting that since we are allowing non-standard models, for some E there may be “false negatives” (where $W = M$ we omit the restriction “ $\upharpoonright M$ ”): There are models \mathcal{M} and equivalence relations E such that

$$\mathcal{M} \models \neg(\exists R) \left(\left| \frac{S_1[M]}{E} \right| \stackrel{R}{\leq} |M| \right)$$

even though there is, in the metatheory, an injection from the E -partition of $S_1[M]$ to M . A sketch of the proof is as follows:

The witnessing equivalence relation will be \approx . Let L be a language expanding L_0 by countably many constants c_i for first-order objects, and the T be the theory containing as axioms all comprehension axioms in the signature, as well as all sentences of the form:

$$\begin{array}{ll} c_i \neq c_j & \text{for each } i \neq j \\ \neg(\exists R)(\varphi(R) \wedge \left| \frac{S_1[M]}{\approx} \right| \stackrel{R}{\leq} |M|) & \text{for each formula } \varphi(R) \text{ of } L \end{array}$$

Every finite subset of T is consistent (and in fact satisfiable), so T is consistent, and has a model whose first-order domain M is infinite *in the metatheory*. The resulting model witnesses the truth of the theorem.

Definition 60. Given E an L_0 -definable equivalence relation, abbreviate as follows:

$$EXPL(X) := \neg \left| \frac{S_1[M] \upharpoonright X}{=} \right| \leq |M| \quad (162)$$

$$TOP(X) := EXPL(X) \wedge EXPL(M - X) \quad (163)$$

$$E_0(X, Y) := (TOP(X) \wedge TOP(Y) \wedge X \boxplus Y) \vee (\neg(TOP(X) \vee TOP(Y)) \wedge X = Y) \quad (164)$$

The abstraction principle A_{E_0} is the correlate of Fine's basal abstraction principle.

Theorem 61 (Generalization of Fine's Characterization Theorem). The abstraction principle A_{E_0} is consistent, and for any expressible equivalence relation E satisfying Permutation Invariance on all models, if A_E is consistent then E_0 is finer than E .

(Recall that E_0 is finer than E means that $\models (\forall X, Y)(E_0(X, Y) \rightarrow E(X, Y))$.)

Note that Theorem 61 implies Fine's Characterization Theorem.

Proof. We give only the portion of the proof not available elsewhere. To see that A_{E_0} is consistent, see [7, p. 144]. Towards establishing the second assertion, note that extensional equality of concepts is the finest Permutation Invariant equivalence relation ([1, Theorem 2, p. 281] also uses this fact). So let $\mathcal{M} \models A_E$; if E is strictly finer than E_0 , there are W, S such that $\mathcal{M} \models TOP(W) \wedge TOP(S)$, $W \boxplus S$ and $\neg E(W, S)$. Thus $E(\boxplus)_W$ is non-trivial, and by Theorem 7, it either refines a separation or refines a complementation.

If it refines a complementation then $|W| = |M - W|$. If $|M| \leq 4$ then

$$\mathcal{M} \models (\forall X, Y)(E_0(X, Y) \leftrightarrow X \approx Y)$$

and so A_{E_0} is HP. But HP has no models of finite size, and since E is finer than E_0 , A_E has no finite models either. Thus M is infinite. But then by Theorem 47, $\mathcal{M} \not\models A_E$.

If E refines a separation, then by Lemma 39, A_E and $|W| \leq |M - W|$ translate $BLV(\leq W)$. Set $R(x, y)$ to be $(\exists Y)(x = \varepsilon Y \wedge Y y)$; R exists by comprehension and $R[\varepsilon Y] = Y$. It is easy to verify using $BLV(\leq W)$ implies

$$\left| \frac{S_1[M] \upharpoonright W}{=} \right| \stackrel{R}{\leq} |M|$$

and this contradicts the assumption that $TOP(W)$.²²

□

²²Fine's Characterization results can in fact be generalized neatly to more types of invariance, as Fine does himself; see [7, Corollary 7, p. 146], and [5].

Acknowledgements

The author wishes to thank Sean Walsh, Roy Cook, Jack Woods, and Eileen Nutting for helpful discussions, comments, and encouragement on this paper. All mistakes, however, are my own. The work on this paper was done much in anticipation of showing it to the author's late friend and former teacher, Aldo Antonelli. That Aldo could not see the conclusion of his inspiration and support marks the completion of this project with great sadness.

References

- [1] G. Aldo Antonelli. Notions of Invariance for Abstraction Principles. *Philosophia Mathematica*, 18(3):276–292, 2010.
- [2] Denis Bonnay. Logicality and Invariance. *Bulletin of Symbolic Logic*, 14(1):29–68, 2008.
- [3] George Boolos. The Standard Equality of Numbers. In *Meaning and Method: Essays in Honor of Hilary Putnam*, pages 261–277. Cambridge University Press, Cambridge, 1990.
- [4] Tim Button. *The Limits of Realism*. Oxford University Press, Oxford, 2013.
- [5] Roy T. Cook. Abstraction and four kinds of invariance (or: What's so logical about counting). *Philosophia Mathematica*, 2016.
- [6] Sean C. Ebels-Duggan. The Nuisance Principle in Infinite Settings. *Thought: A Journal of Philosophy*, 4(4):263–268, December 2015.
- [7] Kit Fine. *The Limits of Abstraction*. The Clarendon Press, Oxford, 2002.
- [8] Gottlob Frege. *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Northwestern University Press, Evanston, second edition, 1980. Translated by John Langshaw Austin.
- [9] Bob Hale. Reals by Abstraction. *Philosophia Mathematica*, 8(3):100–123, 2000. Reprinted in [10].
- [10] Bob Hale and Crispin Wright. *The Reason's Proper Study*. Oxford University Press, Oxford, 2001.
- [11] Bob Hale and Crispin Wright. To Bury Caesar. In *The Reason's Proper Study*, pages 335–396. Clarendon Press, Oxford, 2001.
- [12] Richard G. Heck, Jr. Finitude and Hume's principle. *Journal of Philosophical Logic*, 26(6):589–617, 1997.

- [13] Richard G. Heck, Jr. The Julius Caesar Objection. In *Language, Thought, and Logic: Essays in Honour of Michael Dummett*, pages 273–308. Oxford University Press, Oxford, 1997. Edited by Richard G. Heck Jr.
- [14] Fraser MacBride. On Finite Hume. *Philosophia Mathematica*, 8:150–159, 2000.
- [15] John MacFarlane. Logical constants. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition, 2015.
- [16] Vann McGee. Logical operations. *Journal of Philosophical Logic*, 25:567–580, 1996.
- [17] Charles Parsons. Frege’s Theory of Number. In Max Black, editor, *Philosophy in America*, pages 180–203. Cornell University Press, Ithaca, 1965.
- [18] W.V. Quine. *Philosophy of Logic*. Harvard University Press, Cambridge, 1970.
- [19] Stewart Shapiro. *Foundations without Foundationalism: A Case for Second-Order Logic*, volume 17 of *Oxford Logic Guides*. The Clarendon Press, New York, 1991.
- [20] Stewart Shapiro and Alan Weir. New V, ZF and Abstraction. *Philosophia Mathematica*, 7:293–321, 1999.
- [21] Gila Sher. *The Bounds of Logic: A Generalized Viewpoint*. MIT Press, 1991.
- [22] Alfred Tarski. What are Logical Notions? *History and Philosophy of Logic*, 7(2):143–154, 1986.
- [23] Sean Walsh. Comparing Hume’s Principle, Basic Law V and Peano Arithmetic. *Annals of Pure and Applied Logic*, 163:1679–1709, 2012.
- [24] Sean Walsh and Sean Ebels-Duggan. Relative Categoricity and Abstraction Principles. *The Review of Symbolic Logic*, 8(3):572–606, September 2015.
- [25] Jack Woods. Logical indefinites. *Logique et Analyse*, 227:277–307, 2014.
- [26] Crispin Wright. *Frege’s Conception of Numbers as Objects*, volume 2 of *Scots Philosophical Monographs*. Aberdeen University Press, Aberdeen, 1983.
- [27] Crispin Wright. On the Philosophical Significance of Frege’s Theorem. In Richard G. Heck Jr., editor, *Language, Thought, and Logic: Essays in Honour of Michael Dummett*, pages 201–244. Oxford University Press, Oxford, 1997.
- [28] Crispin Wright. Is Hume’s Principle Analytic? *Notre Dame Journal of Formal Logic*, 40(1):6–30, 1999.