THE REVIEW OF SYMBOLIC LOGIC
Volume 8, Number 3, September 2015

RELATIVE CATEGORICITY AND ABSTRACTION PRINCIPLES

SEAN WALSH
Department of Logic and Philosophy of Science, University of California, Irvine
and
SEAN EBELS-DUGGAN
Department of Philosophy, Northwestern University

Abstract. Many recent writers in the philosophy of mathematics have put great weight on the relative categoricity of the traditional axiomatizations of our foundational theories of arithmetic and set theory (Parsons, 1990; Parsons, 2008, sec. 49; McGee, 1997; Lavine, 1999; Väänänen & Wang, 2014). Another great enterprise in contemporary philosophy of mathematics has been Wright's and Hale's project of founding mathematics on abstraction principles (Hale & Wright, 2001; Cook, 2007). In Walsh (2012), it was noted that one traditional abstraction principle, namely Hume's Principle, had a certain relative categoricity property, which here we term *natural relative categoricity*. In this paper, we show that most other abstraction principles are *not* naturally relatively categorical, so that there is in fact a large amount of incompatibility between these two recent trends in contemporary philosophy of mathematics. To better understand the precise demands of relative categoricity in the context of abstraction principles, we compare and contrast these constraints to (i) stability-like acceptability criteria on abstraction principles (cf. Cook, 2012), (ii) the Tarski-Sher logicality requirements on abstraction principles studied by Antonelli (2010b) and Fine (2002), and (iii) supervaluational ideas coming out of the work of Hodes (1984, 1990, 1991).

§1. Introduction. Our topic in this paper is the compatibility of abstraction principles and relative categoricity. The most famous example of an abstraction principle is Hume's Principle, which asserts that the number of X's is the same as the number of Y's if and only if X and Y are equinumerous, where this just means that there is a bijection between the X's and Y's. In this, the "number of" operator is understood to be a type-lowering operator which takes second-order entities and returns first-order entities. Much of the interest in Hume's Principle stems from the fact that it, in conjunction with certain axioms for second-order logic, recovers the now standard Peano axiomatization for arithmetic (Wright, 1983, chap. 4, cf. Walsh, 2012, sec. 2.2, pp. 1688 ff). Wright (1998) vivified this by asking us to consider the perspective of "Hero," who using Hume's Principle recovers all the laws of pure and applied arithmetic using Frege's definitions for zero, successor, and natural number.

The motivation for relative categoricity is likewise often conveyed by considering the perspective of agents. Parsons in his seminal essay (Parsons, 1990) and later book (Parsons, 2008, sec. 49) asks us to consider interlocutors h and c, each having access to her own structure satisfying the axioms for arithmetic, but whose access to the other's number structure comes primarily from the other's literal utterances. Initially it might appear that the incompleteness of arithmetic could allow for intractable disagreement: perhaps the structure which agent h has in mind witnesses the arithmetized version of the consistency

Received: September 11, 2014.

© Association for Symbolic Logic, 2015

doi:10.1017/S1755020315000052

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

of some theory, while the structure agent c has in mind does not. However, Parsons notes that this can't happen so long as each interlocutor can in addition perform mathematical induction on concepts defined in terms of the other's natural number structure. If this is granted, then the map Γ defined by $\Gamma(0_h) = 0_c$ and $\Gamma(s_h(n)) = s_c(\Gamma(n))$ is an isomorphism, wherein 0_i and s_i denote the zero and successor of agent i's number structure. Since isomorphisms preserve truth-values, the two agents will agree on all sentences of the pure arithmetical vocabulary in which zero and successor are taken as primitive.

So it's natural to ask whether there is similar agreement when it comes to abstraction principles. To expand upon Wright's example, we might envision Hero as well as an interlocutor Claudio, and ask whether their agreement upon Hume's Principle leads to an agreement on other numerical truths. Frege famously noted in the *Grundlagen* one potential source of disagreement: Hero and Claudio might disagree about whether everything is a number, or whether this-or-that object is a number (Frege, 1884; Frege, 1980, sec. 56 and 66). But it's still natural to ask what happens when we restrict attention to each agent's *pure numbers*, that is, to the range of each's "number of" operator. One important species of agreement has been stressed in the recent literature under the heading of the principle 'Nq'. In terms of our scenario, Hero and Claudio would both think that the number of Y's is equal to n, as defined in terms of their respective "number of" operators, if and only if $\exists^{=n}\ x\ Yx$, wherein this is the exact numerical quantifier defined in the usual first-order way (cf. Hale, 1987, pp. 223–224; Wright, 1999, p. 18; Hale & Wright, 2001, p. 322; Cook, 2007, p. 32;Walsh, 2014b, p. 92).

It turns out that this agreement extends to the truth-values of *all* pure numerical statements, where again we understand by "pure" the restriction to the range of each "number of" operator. For, our interlocutors, just like Parsons', have a natural way of translating between their individual pure number discourse. Designating Hero's "number of" with "$\#_h$" and Claudio's with "$\#_c$," we see that whenever Claudio utters a statement about his pure numbers, we may replace each instance of $\#_c$ by $\#_h$ and obtain a truth about Hero's pure numbers. This is due to the fact that the map $\Gamma(\#_h(X)) = \#_c(X)$ is an isomorphism between Hero's and Claudio's "pure number" structures, so that we can again appeal to the fact that isomorphisms preserve the relevant truth-values. So not only will Hero and Claudio agree about all the truths of number theory, they will agree also, for instance, that the number of evens is the same as the number of natural numbers.

This result about Hume's Principle was proven in the earlier paper (Walsh, 2012, proposition 14, p. 1687). However, this earlier work left open the question of whether this phenomena persists when one considers other abstraction principles. For, Hale & Wright (2001) have emphasized that Hume's Principle is just one abstraction principle amongst many. Other principles can be obtained by replacing the equinumerosity relation with another equivalence relation on second-order entities and by introducing a new type-lowering operator for each such equivalence relation. So an *abstraction principle* is a principle of the following form:

$$A[E]: \qquad \forall\ X, Y\ (\partial(X) = \partial(Y) \leftrightarrow E(X, Y)) \tag{1.1}$$

In this, the quantifiers range over second-order entities and the operator ∂ takes second-order entities and returns first-order entities. Moreover, the type-lowering operator ∂ is understood to depend on the equivalence relation E. The case of the two agents is thus well-formalized by the following principle:

$$A^2[E]: \qquad \forall\ X, Y\ [(\partial_1(X) = \partial_1(Y) \leftrightarrow E(X, Y)) \wedge (\partial_2(X) = \partial_2(Y) \leftrightarrow E(X, Y))] \tag{1.2}$$

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Let's call the objects in the range of the ∂_i-operator *the abstracts* of interlocutor i, and let's denote this by $\mathrm{rng}(\partial_i)$. Then we say that the abstraction principle $A[E]$ is *naturally relatively categorical* if it can be proved from this principle $A^2[E]$ (and the axioms governing the second-order logic) that the map $\Gamma(\partial_1(X)) = \partial_2(X)$ is an isomorphism between the abstracts of the two interlocutors. For a more formal statement of natural relative categoricity, see Definition 3.3, which we provide after carefully setting up the particulars of the background second-order logic.

The present paper answers the question of what abstraction principles are naturally relatively categorical by presenting various equivalent characterizations in terms of sameness of cardinality and invariance under injections. These characterizations then allow us to ascertain easily whether a given abstraction principle is naturally relatively categorical (see Section 5 for examples). Let us build up to the statement of these equivalent characterizations by introducing some key definitions we deploy in this paper. First, let us say that the abstraction principle $A[E]$ is *cardinality coarsening on abstracts* if the following is provable from the abstraction principle and the associated principles of the second-order logic:

$$\forall\, X, Y\ ((Y \approx X\ \&\ X \subseteq \mathrm{rng}(\partial)) \rightarrow E(X, Y))] \tag{1.3}$$

In this, $Y \approx X$ is an abbreviation for the notion of equinumerosity operative in Hume's Principle, while $\mathrm{rng}(\partial)$ is the collection of objects in the range of the type-lowering operator ∂; as above, we sometimes refer to these as the *abstracts*. Note that in (1.3), the concept Y is not required to be subconcept of $\mathrm{rng}(\partial)$. Further, let us say that the abstraction principle $A[E]$ is *injection invariant on abstracts* if the following is provable from the abstraction principle and the associated principles of the ambient second-order logic:

$$\forall\ \text{injection}\ \iota : V \rightarrow V\ \forall\ X \subseteq \mathrm{rng}(\partial)\ E(X, \bar{\iota}(X)) \tag{1.4}$$

wherein $V = \{x : x = x\}$ is an abbreviation for the concept of all objects and $\bar{\iota}(X)$ means $\{\iota(x) : x \in X\}$. With this terminology in place, our primary characterization of natural relative categoricity can be stated as follows:

THEOREM 1.1. *The following are equivalent:*

1. *The abstraction principle* $A[E]$ *is naturally relatively categorical.*
2. *The abstraction principle* $A[E]$ *is injection invariant on abstracts.*
3. *The abstraction principle* $A[E]$ *is cardinality coarsening on abstracts.*

This theorem is proven in Section 4. Prior to establishing this theorem, in Section 2 we define the particulars of the second-order logic which we're employing– in short, we're assuming full impredicative comprehension and strong forms of the axiom of choice. Then in Section 3 we present the official definition of natural relative categoricity in Definition 3.3.

As mentioned above, Frege observed that abstraction principles like Hume's Principle don't determine whether or not everything is an abstract. The idea behind the natural relative categoricity of Hume's Principle is that this is the only kind of statement– expressible in the language of Hume's Principle– whose truth-value is left undetermined by Hume's Principle. So in natural relative categoricity we restrict attention down to the abstracts. A complementary idea is to restrict attention to the case where the abstraction operator is assumed to be a surjective map from concepts to objects. To this end, let us call an abstraction principle $A[E]$ *surjectively relatively categorical* if it can be proved (in the background logic) from $A^2[E]$, and the claim that ∂_1 and ∂_2 are surjective, that the

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

map $\Gamma(\partial_1(X)) = \partial_2(X)$ is an isomorphism between the abstracts of the two interlocutors. In analogy with our first main Theorem 1.1, our second main theorem establishes the following equivalent characterization of surjective relatively categorical abstraction principles:

THEOREM 1.2. *The following are equivalent:*

1. *The abstraction principle $A[E]$ is surjectively relatively categorical.*
2. *The abstraction principle $A[E]$ is permutation invariant, under the assumption that the abstraction operator is surjective.*
3. *The abstraction principle $A[E]$ is bicardinality coarsening, under the assumption that the abstraction operator is surjective.*

In analogy to injection invariance on abstracts (1.4), we say that $A[E]$ is *permutation invariant* if

$$\forall \text{ bijection } \pi : V \to V \;\; \forall X \; E(X, \overline{\pi}(X)) \tag{1.5}$$

Further, in analogue to cardinality coarsening on abstracts (1.3), we say that $A[E]$ is *bicardinality coarsening* if

$$\forall X, Y \; (Y \approx X \;\&\; V \setminus X \approx V \setminus Y) \rightarrow E(X, Y) \tag{1.6}$$

In this, $V \setminus X$ denotes the concept of all objects which are not in X. It's worth emphasizing that in the statement of Theorem 1.2, the last two conditions occur under the hypothesis that the abstraction operator ∂ is surjective: for every object b there is a concept F such that $\partial(F) = b$. Hence there is no restriction to abstracts in the formulations of permutation invariance (1.5) and bicardinality coarsening (1.6). As with our earlier theorem, Theorem 1.2 is proven in Section 4; the formal definition of surjective relative categoricity is given in Definition 3.6, subsequent to our treatment of the background second-order logic in the next section.

The aim of our Theorems 1.1 and 1.2 is to answer the question of what abstraction principles are relatively categorical in the specified senses, and in the coarsest of terms these theorems indicate that such abstraction principles look a lot like Hume's Principle, so that relative categoricity amongst abstraction principles is the exception rather than the rule. This obviously doesn't directly imply that there's a problem with appeals to either abstraction principles or relative categoricity considerations. But it indicates that a choice must be made: the advocate of relative categoricity arguments will find her preferred route to determinacy of truth value blocked in the case of most abstraction principles, and the advocate of abstraction principles might be pressed to find some other means by which to secure determinacy of truth value. That said, it's obviously non-trivial to spell out precisely what philosophical concern or question is intended to be assuaged by securing determinacy of truth-value (cf. Button & Walsh, 2015), and there are similarly problems with making out the case that abstraction principles can secure knowledge of our foundational theories of arithmetic and set theory (cf. Walsh, 2014b). This is not the place to adjudicate these larger philosophical issues. Rather, the aim of this paper is limited to showing that the tools which one segment of the philosophy of mathematics community have been using are largely incompatible with the tools employed by another part of the community. And this despite the fact that both relative categoricity and abstraction principles can be seen as latter-day descendants of the idea that the subject-matter of mathematics is given by implicit definitions of its fundamental concepts (cf. Hale & Wright, 2000; Shapiro, 2005, p. 13, pp. 168–169; Shapiro, 1991, p. 190; Shapiro, 2000, pp. 132 ff).

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

The present paper is organized as follows. In Section 2 we set out the particulars of the background second-order logic which we're employing. Then in Section 3 we make a preliminary study of the map $\Gamma(\partial_1(X)) = \partial_2(X)$, which we there call *the natural bijection*. In the subsequent Section 4 we prove Theorems 1.1 and 1.2, as well as note some related results on the underlying equivalence relations. In Section 5, we use these theorems to determine quickly whether some well-known abstraction principles are relatively categorical in the senses we have adumbrated. In Sections 6–9 we contrast the nature of the requirement of relative categoricity to other constraints on abstraction principles related to invariance and determinacy of truth-value studied by authors such as Cook, Antonelli, Fine, and Hodes. In particular, in Section 6 we indicate where relatively categorical abstraction principles fit into the stability hierarchy that has arisen in response to the Bad Company problem. In Section 7, we distinguish our notion of permutation invariance (1.5) from notions studied by Antonelli and Fine and related to the Tarski-Sher thesis on logicality. Finally, in Section 8 we note that the determinacy of truth value ideas coming out of our notions of relative categoricity might be orthogonal to the determinacy ideas coming out of Hodes' supervaluationism (cf. Question 8.5).

§2. Background second-order logic. We work in a background second-order signature which contains a sort for objects as well as a sort for n-ary relations for each $n \geq 1$. Objects are written with lower-case roman letters $a, b, c, d, x, y, z, \ldots$. The unary relations are called *concepts* and written with upper-case roman letters A, B, C, X, Y, Z, while n-ary relations for $n > 1$ are written with upper-case roman letters R, S. The predication relation is written Xa or $a \in X$ for objects a and concepts X; and it is typically written $R(a_1, \ldots, a_n)$ for n-ary relations. For the sake of definiteness, let us then stipulate:

DEFINITION 2.1. *The background second-order signature L_0 is the many-sorted signature which consists merely of (i) sorts for objects and n-ary relations for each $n \geq 1$, and (ii) for each $n \geq 1$, the $(n+1)$-ary predication relations $R(x_1, \ldots, x_n)$ wherein R is an n-ary relation and $x_1, \ldots, x_n$ are objects.*

The models of L_0 thus have the following form:

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots) \tag{2.1}$$

wherein M is a non-empty set and $S_1[M] \subseteq P(M^n)$, and wherein the predication relations are interpreted with the membership relation from the ambient set-theoretic metatheory. Often in what follows we will be discussing isomorphisms of L_0-structures and related expansions. In this connection, it's useful to explicitly note that L_0 itself does not contain any constant or function symbols and does not contain any relation symbols besides the predication relation symbols.

Suppose that L is an expansion of L_0. An L-structure $\mathcal{M}$ whose L_0-reduct is written as in (2.1) is called *standard* if $S_n[M] = P(M^n)$; we do *not* assume here that all structures are standard. The *full comprehension schema* for concepts in L is the collection of all the following axioms:

$$\exists\, X\, \forall\, x\, (Xx \leftrightarrow \Phi(x)) \tag{2.2}$$

wherein $\Phi(x)$ is an L-formula and X does not appear free in $\Phi(x)$, but where $\Phi(x)$ may contain other free variables which are reserved for parameters. There are similar comprehension schemas for the n-ary relations.

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

In this paper, it is assumed that all theories contain the full comprehension schema in their signature for n-ary relations for all $n \geq 1$. Our primary motivation for working with the full comprehension schema in this paper is that it is presupposed by the statement of natural relative categoricity, as we will make clear in the formal presentation of this notion below. The comprehension schema allows us to use usual boolean connectives $A \cap B$, $A \cup B$, $A \setminus B$, $A \subseteq B$ with their usual meanings on both concepts and n-ary relations for $n > 1$. Likewise, we use $A \times B$ to denote the binary concept consisting of pairs (a, b) where a is from A and b is from B. Sometimes in what follows we use the disjoint union notation $A = B \sqcup C$, which of course just means that $A = B \cup C$ and $B \cap C = \emptyset$. In this paper, $\emptyset$ and V are reserved for the concept of no objects and the concept of all objects, respectively:

$$\emptyset = \{x : x \neq x\}, \qquad V = \{x : x = x\} \tag{2.3}$$

In what follows, we often employ various abbreviations for formulas in second-order logic. In particular, we use bracket notation $\{x : \Phi(x)\}$ as short-hand for the unique concept X determined by $\Phi(x)$ from the comprehension axiom in equation (2.2), and likewise we write $\{(x_1, \ldots, x_n) : \Phi(x_1, \ldots, x_n)\}$ in the case of n-ary relations. Functions are identified with their graphs, so that $G : A \to B$ is an abbreviation for the claim that the binary relation G is such that for all a from A there is unique b from B with $G(a, b)$. Likewise we have the following abbreviations for cardinality-related notions:

$X \approx Y$ or $|X| = |Y|$ is an abbreviation for there being a bijection $F : X \to Y$. (2.4)

$|X| \leq |Y|$ is an abbreviation for there being an injection $F : X \to Y$. (2.5)

$|X| < \omega$ is an abbreviation for the claim that $X$is Dedekind-finite, i.e. any injection $F : X \to X$ is also a surjection; and we abbreviate $|X| \geq \omega$ for its negation. (2.6)

Another abbreviation which we shall employ repeatedly in what follows pertains to images of concepts under maps on objects. Suppose that $f : X \to Y$ is a map and suppose $X_0 \subseteq X$. Then we define the image of X_0 under f as $\overline{f}(X_0) = \{f(x) : x \in X_0\}$, which is a concept by full comprehension when $f \in S_2[M]$. Typically we apply this in the case where $f : V \to V$, so that $\overline{f}(X)$ is defined for any concept X.

The other definitions that we need in order to state our results are two forms of the axiom of choice. The first form is designated as AC, and it is the following schema:

$$[\forall\, \overline{x}\, \exists\, R'\, \varphi(R', \overline{x})] \to \exists\, R\, [\forall\, \overline{x}\, \varphi(R[\overline{x}], \overline{x})] \tag{2.7}$$

wherein $R[\overline{x}] = \{\overline{y} : R\overline{x}\overline{y}\}$, which exists by full comprehension. Intuitively this says that if for every n-tuple of objects $\overline{x}$ there is an m-ary concept R' witnessing the condition $\varphi(R', \overline{x})$, then there is an $(n+m)$-ary concept R such that for all n-tuples $\overline{x}$ the m-ary concept $R[\overline{x}]$ is a witness. More intuitively still, this version of the axiom of choice says that if for each object there is a concept satisfying a certain condition, then there is a uniform way to select these concepts. This version of the axiom of choice was used frequently in the earlier papers (Walsh, 2012, definition 5, p. 1683;Walsh, 2014a) since in the setting of limited comprehension it is a natural component of a sufficient condition for the so-called Δ^1_1-comprehension schema.

The other form of the axiom of choice that we employ is a form of global choice. Suppose that T is a theory in one of our signatures. Then we let T+GC be the expansion of T by a new binary relation symbol $<$ on objects in the signature, with axioms saying that $<$ is a linear order of the first-order objects, and we additionally have a schema in the expanded signature saying that any instantiated formula $\varphi(x)$ in the expanded signature,

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

perhaps containing parameters, that holds of some first-order object x will hold of a $<$-least element:

$$[\exists\, x\ \varphi(x)] \rightarrow [\exists\, x\ \varphi(x)\ \&\ \forall\ y\ <\ x\ \neg\varphi(y)] \tag{2.8}$$

Since all our theories T contain full comprehension (2.2), we have that the graph of $<$ forms a binary concept in $T + \mathtt{GC}$. Of course the postulated binary relation $<$ does not necessarily have anything to do with the usual "less than" relation on the natural numbers. This form of global choice was also defined and employed in the paper Walsh (2014a) where it was likewise designated as GC. So in contrast to equation (2.1), models of our global choice principle GC have the form:

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <) \tag{2.9}$$

where $<$ is a linear order on M such that any non-empty $\mathcal{M}$-definable subset has a least element.

It's worth stressing that L_0 does *not* include the global well-order (cf. Definition 2.1). For, the equivalence relations that we will consider will all be L_0-formulas, and in Section 7, we will note that this implies that the equivalence relations are logical in the sense of Tarski-Sher, and this would be not be true if they included the global well-order. However, as far as theories and structures go, in what follows, it is assumed that all theories and structures contain AC and GC. This of course may be taken to hold for standard structures by recourse to the axiom of choice in the metatheory. Since we are additionally assuming full comprehension, our models in effect look and act a lot like standard models. However, the advantage of working with arbitrary models of these axioms– instead of restricting ourselves to the standard models– is that we have the benefits of the completeness theorem (cf. Enderton, 2001, chap. 4; Manzano, 1996, chap. VII.2). So even though we are working model-theoretically, everything can in principle be turned into a concrete deduction by recourse to this theorem. Because the semantics for second-order logic are a contentious affair (cf. Linnebo, 2011, chap. 6 for overview), and because relative categoricity arguments have been traditionally motivated by concerns with the standard semantics for second-order logic (Parsons, 2008, p. 270; McGee, 1997, pp. 45–47; Lavine, 1999, p. 5), it is useful to adopt a framework in which our results do not depend on the choice of semantics for second-order logic.

Our use of AC (2.7) is rather limited in this paper: we appeal to it to establish the implication recorded in Figure 1 in Section 4 which allows us to go from conditions on an equivalence relation to conditions on the associated abstraction principle; and we appeal to it a final time to treat notions of finiteness in our discussion of the examples in Section 5. As for global choice GC (2.8), the reason why we assume it in this paper is that it permits us to transfer the usual properties of cardinal arithmetic to our deductive setting. In particular, in what follows we make use of the following three properties:

$$\textit{Cardinal Comparability} : \forall\, X, Y\ (|X| \leq |Y| \vee |Y| \leq |X|) \tag{2.10}$$

$$\textit{Infinite Sums are Maxs} : \tag{2.11}$$
$$\forall\, X\ \ |X| \geq \omega \rightarrow [\forall\, Y, Z\ (X = Y \sqcup Z \rightarrow |X| = \max\{|Y|\,, |Z|\})]$$

$$\textit{Infinite Products are Maxs} : \tag{2.12}$$
$$\forall\, Y, Z\ (|Y| \geq \omega \vee |Z| \geq \omega) \rightarrow |Y \times Z| = \max\{|Y|\,, |Z|\}$$

Of course, by Cardinality Comparability (2.10), the maximum expressions make good sense. For instance, the clause $|X| = \max\{|Y|\,, |Z|\}$ is just an abbreviation for the following

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

conjunction of conditionals, and Cardinality Comparability (2.10) implies that one of the two antecedents is satisfied:

$$(|Y| \geq |Z| \rightarrow |X| = |Y|) \wedge (|Z| \geq |Y| \rightarrow |X| = |Z|) \tag{2.13}$$

Note that Cardinality Comparatibility (2.10) follows from global choice GC: for, the global well-order restricted to any two concepts yields two well-orders, and we can then use the traditional proof that well-orders are either order-isomorphic or one is isomorphic to an initial segment of the other (cf. Hrbacek & Jech, 1999, p. 105; Kunen, 1980, p. 15); and this result trivially implies that each is comparable to the other in terms of cardinality as well. Similarly, we can transfer the usual proof of Infinite Sums are Maxs and Infinite Products are Maxs (2.11)–(2.12) as expressed in the language of set theory to our framework for second-order logic, since this proof proceeds by transfinite induction, which we can emulate with our global well-order (cf. Hrbacek & Jech, 1999, pp. 134 ff; Kunen, 1980, p. 29; Kunen, 2011, p. 73).

Another use of the global well-order $(V, <)$ that we employ is in giving an equivalent characterization of Dedekind-finiteness (2.6). First, let's introduce the following notation for the initial segments: $I_a = \{b : b < a\}$ and $\overline{I}_a = \{b : b \leq a\}$. Further, since $(V, <)$ is a non-empty well-order, it has a least element, which we designate as zero or 0. There's also a natural partial successor function s defined as follows:

$$s(a) = \min_{<}(V \setminus \overline{I}_a) = \min_{<}\{b : b > a\} \tag{2.14}$$

This function s might be partial because there might be a greatest element in well-order $(V, <)$. Finally, let's say that a *limit point* in $(V, <)$ is a point $a > 0$ such that $b < a$ implies $s(b) < a$. As in the theory of ordinals, the well-order $(V, <)$ splits into zero, successors, and limits. Finally, let's say that a is *finite* if a is strictly below the all the limit points, and let's say that X is *finite* if X is bijective with I_a or $\overline{I}_a$ for some finite a. Then one can show using induction that:

PROPOSITION 2.2. *(i) X is finite if and only if $|X| < \omega$. (ii) If there is a least limit, and the least limit point is designated as ω, then X is infinite iff there is an injection $\iota : I_\omega \rightarrow X$, which happens iff $|X| \geq \omega$.*

Hence, $|X| \geq \omega$ (as defined in (2.6)) aligns extensionally with $|I_\omega| \leq |X|$ as defined in (2.5).

Sometimes in what follows we use some standard terminology for describing equivalence relations on a set. So suppose that E is an equivalence relation on a set P. Usually in what follows P will be the power set $P(M)$ of some set M. Then we use $[X]_E = \{Y \in P : E(X, Y)\}$ as an abbreviation for the E-equivalence classes of an element X of P. Further, we use $\frac{P}{E} = \{[X]_E : X \in P\}$ for the set of all equivalence classes. Finally, representatives for the equivalence classes will be given by any injection $\iota : {}^{P}/_{E} \rightarrow P$ such that $E(X, \iota([X]_E))$. More generally, often in what follows we shall be interested in the related situation of injections $\iota : \frac{P(M)}{E} \rightarrow M$. As we'll see at the outset of the next section, this is sometimes useful for construction models of abstraction principles.

§3. The natural bijection. With these preliminary definitions pertaining to the background second-order logic in place, we can now proceed to define what a model of an abstraction principle is. Suppose that $E(X, Y)$ is an L_0-formula with exactly two free concept variables. Let $L_0[\partial]$ be the expansion of L_0 by a function symbol ∂ which takes unary concepts as inputs and outputs objects. Then the *abstraction principle* $A[E]$ *associated*

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

to E is as in equation (1.1) from the previous section. By abuse of notation, we also use $A[E]$ to refer to the theory consisting of this abstraction principle, the full comprehension schema (2.2), the axiom of choice AC (2.7), and the global choice schema GC (2.8). Then in contrast to equation (2.9), models of $A[E]$ have the form:

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <, \partial) \tag{3.1}$$

wherein $\partial : S_1[M] \to M$ and wherein $\mathcal{M}$ models the full comprehension schema (2.2), the axiom of choice (2.7), and the global choice schema GC (2.8). In the case where the structure $\mathcal{M}$ from (2.9) is standard, if there is an injection $\iota : \frac{P(M)}{E} \to M$, then one can build a model as in equation (3.1) by setting $\partial(X) = \iota([X]_E)$. Of course, there is such an injection if and only if $\left|\frac{P(M)}{E}\right| \leq |M|$, which is a non-trivial assumption.

Now consider the case in which there are, within a single model, two abstraction operators which satisfy a given abstraction principle. Again suppose that $E(X, Y)$ is an L_0-formula with exactly two free concept variables. Let $L_0[\partial_1, \partial_2]$ be the expansion of L_0 by two function symbols ∂_1, ∂_2 which takes unary concepts as inputs and outputs objects. Then the theory $A^2[E]$ consists of the axiom (1.2) from Section 1, as well as the full comprehension schema (2.2), the axiom of choice AC (2.7), and the global choice schema GC (2.8). So models of $A^2[E]$ have the following form:

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <, \partial_1, \partial_2) \tag{3.2}$$

wherein $\partial_i : S_1[M] \to M$ and $\mathcal{M}$ models the full comprehension schema (2.2), the axiom of choice AC (2.7), and the global choice schema GC (2.8).

In the description of natural relative categoricity from the earlier section, one of the key ideas is that we restrict down to the ranges of the individual abstraction operators. Formally, we make this precise by taking a model $\mathcal{M}$ of $A^2[E]$ as in equation (3.2), and defining the following induced $L_0[\partial_i]$-structure for $i = 1, 2$:

$$\mathcal{M}_i = (\mathrm{rng}(\partial_i), S_1[M] \cap P(\mathrm{rng}(\partial_i)), S_2[M] \cap P(\mathrm{rng}(\partial_i)^2), \ldots, \partial_i \restriction (S_1[M] \cap P(\mathrm{rng}(\partial_i)))) \tag{3.3}$$

Hence $\mathcal{M} \mapsto \mathcal{M}_1$ and $\mathcal{M} \mapsto \mathcal{M}_2$ are maps from an $L_0[\partial_1, \partial_2]$-structure $\mathcal{M}$ to an $L_0[\partial_i]$-structure $\mathcal{M}_i$. So notationally, $\mathcal{M}_i$ is a structure induced from $\mathcal{M}$, and not simply yet another structure indexed by a subscript. Note that we do *not* include the global-well order $<$ in the signature of the induced structures $\mathcal{M}_i$. This is because our natural relative categoricity concerns isomorphisms between these structures, and we do not want to insist that isomorphisms preserve this global well-order. This is because the global well-order is an artifact employed to make various second-order notions like cardinality more like classical metatheoretic notions. But of course since we're reasoning about the induced structures $\mathcal{M}_i$ as defined within the larger structure $\mathcal{M}$, we can use global choice in that setting to reason about the induced structures if we like. This disparity between the induced structures $\mathcal{M}_i$ and the structure $\mathcal{M}$ as regards global choice does not extend to the issue of comprehension. For, the induced structures $\mathcal{M}_i$ models the full comprehension schema (2.2) in its signature simply because they are definable within the structure $\mathcal{M}$ which is assumed to satisfy comprehension in its signature. However, note that in general there is no reason that the induced structure $\mathcal{M}_i$ need model the abstraction principle $A[E]$.

The notion of natural relative categoricity was defined in Section 1 by the condition that the map $\Gamma(\partial_1(X)) = \partial_2(X)$ was an isomorphism. Before further examining the condition that this map is an isomorphism, let us take a first and preliminary step of examining the

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

properties of the map itself, which we call the natural bijection. So given any model $\mathcal{M}$ of $A^2[E]$ as in equation (3.2), *the natural bijection* Γ is the map $\Gamma : \mathrm{rng}(\partial_1) \to \mathrm{rng}(\partial_2)$ defined by

$$\Gamma(\partial_1(X)) = \partial_2(X) \tag{3.4}$$

wherein X ranges over elements of $S_1[M]$. It then follows from the axiom (1.2) of $A^2[E]$ that this map is well-defined and injective:

$$\partial_1(X) = \partial_1(Y) \iff E(X, Y) \iff \partial_2(X) = \partial_2(Y) \tag{3.5}$$

Trivially, by definition, $\Gamma : \mathrm{rng}(\partial_1) \to \mathrm{rng}(\partial_2)$ is surjective and so it is indeed a bijection. Further, the natural bijection is definable in $\mathcal{M}$ by the following formula:

$$\Gamma(x) = y \iff \mathcal{M} \models [\exists\, X\ (\partial_1(X) = x\ \&\ \partial_2(X) = y)] \tag{3.6}$$

So by full comprehension (2.2), the graph of the natural bijection Γ is a member of $S_2[M]$. Likewise, by full comprehension (2.2), if $X \in S_1[M] \cap P(\mathrm{rng}(\partial_1))$ then the following is an element of $S_1[M] \cap P(\mathrm{rng}(\partial_2))$:

$$\overline{\Gamma}(X) = \{\Gamma(x) : x \in X\} \tag{3.7}$$

and similarly for n-ary relations. By abuse of notation, we also use the symbol $\overline{\Gamma}$ to refer to the map $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ given by Γ on the objects and $\overline{\Gamma}$ as in equation (3.7) on the n-ary relations for all $n \geq 1$; and we sometimes also refer to the map $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ as the natural bijection. This map $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is also trivially an injection since Γ is. As is easily verified, it is a surjection as well.

So, indeed $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is also a bijection. For ease of future reference, let's record this in the following definition:

DEFINITION 3.1. *Suppose that $\mathcal{M}$ is a model of $A^2[E]$ as in equation (3.2), and that $\mathcal{M}_1, \mathcal{M}_2$ are the induced structures as in equation (3.3). Then the* natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ *is given by the bijection $\Gamma : \mathrm{rng}(\partial_1) \to \mathrm{rng}(\partial_2)$ defined by $\Gamma(\partial_1(X)) = \partial_2(X)$ for each concept X from the ambient structure $\mathcal{M}$. Further, $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is defined on concepts $X \subseteq \mathrm{rng}(\partial_1)$ from the ambient structure by $\overline{\Gamma}(X) = \{\Gamma(x) : x \in X\}$, and similarly for n-ary relations.*

The various appeals to the full comprehension schema (2.2) that we made in the previous paragraph underscore the apparent necessity of the adoption of this schema in the context of the present discussion. For instance, to show that the graph of Γ exists as a binary concept, we appealed to its definition in equation (3.6), which is Σ^1_1. This is precisely the amount of comprehension that one needs to show that Basic Law V, the abstraction principle of Frege's *Grundgesetze*, is inconsistent (cf. Walsh, 2012, proposition 4, p. 1682, proposition 29, p. 1692). Hence, it seems that studying natural relative categoricity in the context of limited comprehension would not be feasible. Before moving on, it's worth recording one final point in regards to the natural bijection: namely, that a routine argument establishes the following.

PROPOSITION 3.2. *If $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is the natural bijection, then its inverse $\Delta = \Gamma^{-1}$ is the natural bijection $\overline{\Delta} : \mathcal{M}_2 \to \mathcal{M}_1$.*

So having defined the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$, let's now examine carefully what it would mean for this to be a isomorphism between the induced structures $\mathcal{M}_1$ and $\mathcal{M}_2$. Recall that if L is an arbitrary signature, then two L-structures $\mathcal{N}_1$ and $\mathcal{N}_2$ are

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

isomorphic if there is a bijection $\gamma : \mathcal{N}_1 \to \mathcal{N}_2$ such that for all L-formulas $\theta(x_1, \ldots, x_n)$ and $a_1, \ldots, a_n$ from $\mathcal{N}_1$, it is the case that

$$\mathcal{N}_1 \models \theta(a_1, \ldots, a_n) \iff \mathcal{N}_2 \models \theta(\gamma(a_1), \ldots, \gamma(a_n)) \tag{3.8}$$

Of course, this condition is difficult to verify directly, so one usually works with the equivalent condition that equation (3.8) holds in the case of atomic formulas (cf. Marker, 2002, definition 1.1.3, pp. 8–9, the proof of Theorem 1.1.10, and p. 13, or Enderton, 2001, p. 94 , the Homomorphism theorem part (c), and p. 96). For instance, consider the atomic formula $\theta(x_1, \ldots, x_n, y) \equiv F(x_1, \ldots, x_n) = y$. Suppose that $a_1, \ldots, a_n, b$ are from $\mathcal{N}_1$ and that $F^{\mathcal{N}_1}(a_1, \ldots, a_n) = b$, so that $\mathcal{N}_1 \models \theta(a_1, \ldots, a_n, b)$. Then equation (3.8) implies that $\mathcal{N}_2 \models \theta(\gamma(a_1), \ldots, \gamma(a_n), \gamma(b))$, or that $F^{\mathcal{N}_2}(\gamma(a_1), \ldots, \gamma(a_n)) = \gamma(b)$, which of course implies that $F^{\mathcal{N}_2}(\gamma(a_1), \ldots, \gamma(a_n)) = \gamma(F^{\mathcal{N}_1}(a_1, \ldots, a_n))$. Elementary considerations such as these show that $\gamma : \mathcal{N}_1 \to \mathcal{N}_2$ is an isomorphism if and only if for all relations symbols R, constant symbols c, and function symbols F in the signature of the structures, and all $a_1, \ldots, a_n$ from $\mathcal{N}_1$, one has

$$R^{\mathcal{N}_1}(a_1, \ldots, a_n) \iff R^{\mathcal{N}_2}(\gamma(a_1), \ldots, \gamma(a_n)) \tag{3.9}$$

$$\gamma(c^{\mathcal{N}_1}) = c^{\mathcal{N}_2} \tag{3.10}$$

$$\gamma(F^{\mathcal{N}_1}(a_1, \ldots, a_n)) = F^{\mathcal{N}_2}(\gamma(a_1), \ldots, \gamma(a_n)) \tag{3.11}$$

While these considerations are admittedly elementary, it's worth underscoring them since they help to motivate the definition of natural relative categoricity (Definition 3.3), which we now build towards.

Now consider the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ and what it would mean for it to be an isomorphism. Since $\overline{\Gamma}(X) = \{\Gamma(x) : x \in X\}$, clearly one has that equation (3.9) always holds in the case of the predication relations, which per the definition of L_0 in Definition 2.1 are the *only* relations in the signature of L_0. So $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism of $L_0[\partial]$-structures if and only if equation (3.11) holds with respect to the operator ∂. That is, the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism of $L_0[\partial]$-structures if and only if

$$X \in (S_1[M] \cap P(\mathrm{rng}(\partial_1)) \Longrightarrow \Gamma(\partial_1(X)) = \partial_2(\overline{\Gamma}(X)) \tag{3.12}$$

Given the way that the natural bijection Γ was defined in equation (3.4), this happens if and only if

$$X \in (S_1[M] \cap P(\mathrm{rng}(\partial_1)) \Longrightarrow \partial_2(X) = \partial_2(\overline{\Gamma}(X)) \tag{3.13}$$

which, by the fact that the model $\mathcal{M}$ from equation (3.2) satisfies $A^2[E]$, holds if and only if

$$\mathcal{M} \models [\forall X\ (X \subseteq \mathrm{rng}(\partial_1)) \to E(X, \overline{\Gamma}(X))] \tag{3.14}$$

Now, the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism if and only if its inverse $\Delta = \Gamma^{-1}$ is an isomorphism, and by Proposition 3.2 its inverse is the natural bijection $\overline{\Delta} : \mathcal{M}_2 \to \mathcal{M}_1$. Hence, by parity of reasoning and the fact that E is an equivalence relation, one has that $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism if and only if

$$\mathcal{M} \models [\forall Y\ (Y \subseteq \mathrm{rng}(\partial_2)) \to E(Y, \overline{\Gamma}^{-1}(Y))] \tag{3.15}$$

For ease of future reference, let us summarize these results as follows. First let's record our official definition of natural relative categoricity:

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

DEFINITION 3.3. *An abstraction principle* $A[E]$ *is* naturally relatively categorical *if all models* $\mathcal{M}$ *of* $A^2[E]$ *from equation (3.2), the natural bijection* $\overline{\Gamma} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ *from Definition 3.1 is an isomorphism of the induced structures* $\mathcal{M}_1, \mathcal{M}_2$ *from equation (3.3).*

The only way in which this formalization of the notion is more precise than the descriptions of this notion given in Section 1 is that now we have formally defined the particulars of our background second-order logic and have likewise defined the natural bijection and indicated precisely what it takes for it to be a isomorphism.

The elementary considerations from the previous paragraphs give us a simple equivalent characterization of natural relative categoricity. In particular, we have:

PROPOSITION 3.4. *An abstraction principle* $A[E]$ *is naturally relatively categorical if and only if all models* $\mathcal{M}$ *of* $A^2[E]$ *from equation (3.2) satisfy one of the two following equivalent conditions:*

$$\mathcal{M} \models [\forall\, X\ (X \subseteq \mathrm{rng}(\partial_1)) \rightarrow E(X, \overline{\Gamma}(X))] \tag{3.14}$$

$$\mathcal{M} \models [\forall\, Y\ (Y \subseteq \mathrm{rng}(\partial_2)) \rightarrow E(Y, \overline{\Gamma}^{-1}(Y))] \tag{3.15}$$

wherein Γ *is the natural bijection (cf. Definition 3.1).*

Expressed in these terms, natural relative categoricity is patently a deductive property of the theory $A^2[E]$.

In the next section we'll prove Theorem 1.1 which gives a characterization of natural relative categoricity in terms of cardinality coarsening on abstracts (1.3) and injection invariance on abstracts (1.4). In the previous section we've formally defined our background second-order logic and so we can be a bit more precise now about the content of these conditions. Recall that $A[E]$ may be used as the abbreviation for the theory consisting of the abstraction principle (1.1) in addition to the full comprehension schema (2.2), the axiom of choice (2.7), and the global choice schema (2.8). Then officially, we say that $A[E]$ is *cardinality coarsening on abstracts* if the following is a theorem of $A[E]$:

$$\forall\, X, Y\ ((Y \approx X\ \&\ X \subseteq \mathrm{rng}(\partial)) \rightarrow E(X, Y))] \tag{1.3}$$

Likewise, officially $A[E]$ is *injection invariant on abstracts* if the following is a theorem of $A[E]$:

$$\forall\ \text{injection}\ \iota : V \rightarrow V\ \forall\, X \subseteq \mathrm{rng}(\partial)\ E(X, \bar{\iota}(X)) \tag{1.4}$$

So both cardinality coarsening on abstracts and injection invariance on abstracts are, by definition, deductive properties of the theory $A[E]$. By contrast, as was made clear by Proposition 3.4 of the previous paragraph, natural relative categoricity is a deductive property of the theory $A^2[E]$. So one of the implications of Theorem 1.1, which establishes the equivalence of these notions, is that we're able to further reduce natural relative categoricity to a deductive property of the theory $A[E]$ as opposed to $A^2[E]$.

Before setting up the particulars of our second main theorem, let's record for reference when an isomorphism $\overline{H} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is equal to the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$. It's natural to focus attention on those isomorphisms $\overline{H} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ whose restriction $H \restriction \mathrm{rng}(\partial_1)$ is an element of $S_2[M]$, since it is only with respect to these that we can define further elements of $\mathcal{M}$ in terms of $\overline{H}$ and H by recourse to the comprehension schema (2.2). By a routine argument, we can establish the following:

PROPOSITION 3.5. *Suppose that* $\mathcal{M}$ *is a model of* $A^2[E]$ *with induced structures* $\mathcal{M}_1$ *and* $\mathcal{M}_2$. *Suppose that* $\overline{H} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ *is a map whose restriction* $H \restriction \mathrm{rng}(\partial_1)$ *to* $\mathrm{rng}(\partial_1)$

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

is an element of $S_2[M]$. *Then* $\overline{H} : \mathcal{M}_1 \to \mathcal{M}_2$ *is isomorphism if and only if one has* $\mathcal{M} \models [\forall\, X\ (X \subseteq \mathrm{rng}(\partial_1) \to E(X, \overline{H}(X))]$.

As a corollary to Theorem 1.1, we will establish in the next section that such isomorphisms are *always* equal to the natural bijection in the setting of natural relative categoricity (cf. Corollary 4.1).

Finally, let's briefly say something about the content of our second main Theorem 1.2. For ease of future reference, let's record the official notion in the following definition:

DEFINITION 3.6. *An abstraction principle* $A[E]$ *is* surjectively relatively categorical *if all models* $\mathcal{M}$ *of* $A^2[E]$ *from (3.2) wherein the abstraction operators* $\partial_i : S_1[M] \to M$ *are surjective, the natural bijection* $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ *from Definition 3.1 is an isomorpism of the induced structures* $\mathcal{M}_1, \mathcal{M}_2$ *from equation (3.3).*

So clearly natural relative categoricity implies surjective relative categoricity. For an example of an abstraction principle which is surjectively relatively categorical but not naturally relatively categorical, see the example of the Bicardinality Principle in 5.3. In the statement of Theorem 1.2, the key notions were that of permutation invariance (1.5) and bicardinality coarsening (1.6). In the context of Theorem 1.2, it is understood that to say an abstraction principle $A[E]$ has one of these properties is to say that these properties are deducible from the supposition that (i) the abstraction operator is a surjection as well as from (ii) the abstraction principle itself, the full comprehension schema (2.2), the axiom of choice (2.7), and the global choice schema GC (2.8).

§4. The equivalent characterizations. The goal of this section is to establish Theorems 1.1 and 1.2. The first provides an equivalent characterization of natural relative categoricity in terms of cardinality coarsening on abstracts (1.3) and injective invariance on abstracts (1.4). The second provides an equivalent characterization of surjectively relatively categoricity in terms of bicardinality coarsening (1.6) and permutation invariance (1.5). These theorems gives us two qualitatively distinct means by which to identify and recognize our versions of relative categoricity. For, in and of themselves, natural relative categoricity and surjective relative categoricity (Definitions 3.3 and 3.6) are claims about determining a single structure. But the notions of injection invariance on abstracts (1.4) and permutation invariance (1.5) deal in a different currency: these conditions say say that a certain second-order relation should be invariant under certain mappings of the entire domain. Cardinality coarsening on abstracts (1.3) and bicardinality coarsening (1.6) are different still: they are more local in character and concern the comparative sizes of a concept (and its relative complement).

Here is then the proof of Theorem 1.1:

Proof. First suppose that $A[E]$ is naturally relatively categorical. Suppose that $\mathcal{M}^* = (M, S_1[M], S_2[M], \ldots, <, \partial_1)$ is an arbitrary model of $A[E]$. Note that $V = \{x : x = x\}$ (2.3) as interpreted on $\mathcal{M}^*$ is exactly M. So with an eye towards showing injection invariance on abstracts, suppose that $\iota : M \to M$ is an injection whose graph is in $S_2[M]$. Then define $\partial_2 : S_1[M] \to M$ by $\partial_2 = \iota \circ \partial_1$. Note that since $\iota : M \to M$ is an injection, we have that the following holds in $\mathcal{M}$:

$$\partial_2(X) = \partial_2(Y) \iff \partial_1(X) = \partial_1(Y) \iff E(X, Y) \tag{4.1}$$

Hence, since ∂_2 is $\mathcal{M}^*$-definable, the following structure is a model of $A^2[E]$:

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <, \partial_1, \partial_2) \tag{4.2}$$

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Since by hypothesis $A[E]$ is naturally relatively categorical, we have that the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism. Then by definition of Γ in equation (3.4), we have that $X \in S_1[M]$ implies $\Gamma(\partial_1(X)) = \partial_2(X) = \iota(\partial_1(X))$. Hence $\Gamma = \iota \restriction (\mathrm{rng}(\partial_1))$. Now we may finally finish verifying injection invariance on abstracts. Suppose that $X \in S_1[M] \cap P(\mathrm{rng}(\partial_1))$. Then $\overline{\Gamma}(X) = \bar{\iota}(X)$. Hence from Proposition 3.4 (and in particular equation (3.14)) we may infer that $E(X, \overline{\Gamma}(X))$ and hence $E(X, \bar{\iota}(X))$.

Second suppose $A[E]$ is injection invariant on abstracts. To show that $A[E]$ is naturally relatively categorical, suppose that

$$\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <, \partial_1, \partial_2) \tag{4.3}$$

is a model of $A^2[E]$. So we must show that the natural bijection $\Gamma : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism. By Cardinal Comparability (2.10),

$$|M \setminus \mathrm{rng}(\partial_1)| \leq |M \setminus \mathrm{rng}(\partial_2)| \quad \text{or} \quad |M \setminus \mathrm{rng}(\partial_2)| \leq |M \setminus \mathrm{rng}(\partial_1)| \tag{4.4}$$

First suppose that $|M \setminus \mathrm{rng}(\partial_1)| \leq |M \setminus \mathrm{rng}(\partial_2)|$. Then let $\Delta : M \setminus \mathrm{rng}(\partial_1) \to M \setminus \mathrm{rng}(\partial_2)$ be a witnessing injection. Define an injection $\iota : M \to M$ by

$$\iota \restriction \mathrm{rng}(\partial_1) = \Gamma, \qquad \iota \restriction (M \setminus \mathrm{rng}(\partial_1)) = \Delta \tag{4.5}$$

Since Γ has range $\mathrm{rng}(\partial_2)$ and Δ has range $M \setminus \mathrm{rng}(\partial_2)$, the map $\iota : M \to M$ is indeed an injection. Now, we verify natural relative categoricity by verifying equation (3.14). So suppose that $X \in (S_1[M] \cap P(\mathrm{rng}(\partial_1))$. Then by injection invariance on abstracts applied to ι, we have that $\mathcal{M} \models E(X, \bar{\iota}(X))$. But since $X \subseteq \mathrm{rng}(\partial_1)$, we have that $\bar{\iota}(X) = \overline{\Gamma}(X)$, so that $\mathcal{M} \models E(X, \overline{\Gamma}(X))$. Hence, we have finished verifying natural relative categoricity via equation (3.14).

Conversely, suppose that $|M \setminus \mathrm{rng}(\partial_2)| \leq |M \setminus \mathrm{rng}(\partial_1)|$ with witnessing injection $\Delta : M \setminus \mathrm{rng}(\partial_2) \to M \setminus \mathrm{rng}(\partial_1)$. Define an injection $\iota : M \to M$ by

$$\iota \restriction \mathrm{rng}(\partial_2) = \Gamma^{-1}, \qquad \iota \restriction (M \setminus \mathrm{rng}(\partial_2)) = \Delta \tag{4.6}$$

Since Γ^{-1} has range $\mathrm{rng}(\partial_1)$ and Δ has range $M \setminus \mathrm{rng}(\partial_1)$, the map $\iota : M \to M$ is indeed an injection. Now, we verify natural relative categoricity by verifying equation (3.15). So suppose that $Y \in (S_1[M] \cap P(\mathrm{rng}(\partial_2))$. Then by injection invariance on abstracts applied to ι, we have that $\mathcal{M} \models E(Y, \bar{\iota}(Y))$. But since $Y \subseteq \mathrm{rng}(\partial_2)$, we have that $\bar{\iota}(Y) = \overline{\Gamma}^{-1}(Y)$, so that $\mathcal{M} \models E(Y, \overline{\Gamma}^{-1}(Y))$. Hence, we have finished verifying natural relative categoricity via equation (3.15).

Having shown the equivalence of natural relative categoricity and injection invariance on abstracts, we now show that these are equivalent to cardinality coarsening on abstracts. First, note that cardinality coarsening on abstracts trivially implies injection invariance on abstracts. For, suppose that we're working in a model of $A[E]$ and there's an injection $\iota : V \to V$ and $X \subseteq \mathrm{rng}(\partial_1)$. Then let $Y = \bar{\iota}(X)$, so that $Y \approx X$ and $X \subseteq \mathrm{rng}(\partial_1)$. Then by cardinality coarsening on abstracts, we have $E(X, Y)$, which is just to say $E(X, \bar{\iota}(X))$, so that we have verified injection invariance on abstracts.

Now assume that $A[E]$ is injection invariant on abstracts. Consider a model $\mathcal{M}^* = (M, S_1[M], S_2[M], \ldots, <, \partial_1)$ of $A[E]$, and suppose that X_0, Y_0 are members of $S_1[M]$ with $Y_0 \approx X_0$ and $X_0 \subseteq \mathrm{rng}(\partial_1)$. We must show that $E(X_0, Y_0)$. There are several cases to consider, which for the sake of readability, we enumerate separately. For the abbreviations of cardinality notions which we employ here, see in particular (2.4)–(2.6) from Section 2.

Case I: $|Y_0| = |X_0| = |V|$. Then $V \approx X_0 \subseteq \mathrm{rng}(\partial_1) \subseteq V$ implies that $|\mathrm{rng}(\partial_1)| = |V|$. Choose a bijection $\pi : \mathrm{rng}(\partial_1) \to V$ and define a map $\partial_2 : S_1[M] \to M$ by

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

$\partial_2(X) = \pi(\partial_1(X))$, so that since $\pi : \mathrm{rng}(\partial_1) \to V$ is a surjection, we have that $\mathrm{rng}(\partial_2) = V$. Then observe that for any X, Y we have the following in $\mathcal{M}^*$ since π is an injection:

$$\partial_2(X) = \partial_2(Y) \iff \partial_1(X) = \partial_1(Y) \iff E(X, Y) \tag{4.7}$$

Hence, the structure $\mathcal{M}^\dagger = (M, S_1[M], S_2[M], \ldots, <, \partial_2)$ is likewise a model of $A[E]$, and so by injection invariance on abstracts we have

$$\mathcal{M}^\dagger \models [\forall \text{ injection } \iota : V \to V \ \forall X \subseteq \mathrm{rng}(\partial_2)\ E(X, \overline{\iota}(X))] \tag{4.8}$$

But since $\mathrm{rng}(\partial_2) = V$, this can be simplified to:

$$\mathcal{M}^\dagger \models [\forall \text{ injection } \iota : V \to V \ \forall X\ E(X, \overline{\iota}(X))] \tag{4.9}$$

Now, since $Y_0 \approx X_0 \approx V$, choose bijections $j_1 : V \to X_0$ and $j_2 : V \to Y_0$. Then by the previous equation, we have $E(V, \overline{j_1}(V))$ and $E(V, \overline{j_2}(V))$, or what is the same $E(V, X_0)$ and $E(V, Y_0)$. Since E is an equivalence relation, we have that $E(X_0, Y_0)$, which was to be demonstrated.

Case II: $|Y_0| = |X_0| < |V|\ \&\ |V| < \omega$. By Cardinal Comparability (2.10),

$$|V \setminus X_0| \leq |V \setminus Y_0| \quad \text{or} \quad |V \setminus Y_0| \leq |V \setminus X_0| \tag{4.10}$$

First suppose that $|V \setminus X_0| \leq |V \setminus Y_0|$. Choose injection $\iota : V \to V$ such that $\iota \restriction X_0 : X_0 \to Y_0$ is a bijection and $\iota \restriction (V \setminus X_0) : V \setminus X_0 \to V \setminus Y_0$ is an injection. Then since $A[E]$ is injection invariant on abstracts, we have that $E(X_0, \overline{\iota}(X_0))$, which is the same as $E(X_0, Y_0)$. Second suppose that $|V \setminus Y_0| \leq |V \setminus X_0|$. Choose injection $j : V \to V$ such that $j \restriction Y_0 : Y_0 \to X_0$ is a bijection and $j \restriction (V \setminus Y_0) : V \setminus Y_0 \to V \setminus X_0$ is an injection. Since $|V| < \omega$, we have that the injection $j : V \to V$ is actually a bijection. Let $\iota = j^{-1}$, so that $\iota : V \to V$ is a bijection and $\overline{\iota}(X_0) = Y_0$. Then by injection invariance on abstracts, we again have $E(X_0, \overline{\iota}(X_0))$ or $E(X_0, Y_0)$.

Case III: $|Y_0| = |X_0| < |V|\ \&\ |V| \geq \omega$. Then by Infinite Sums are Maxs (2.12), our case assumptions imply that $|V| = \max\{|V \setminus X_0|, |X_0|\} = |V \setminus X_0|$ and $|V| = \max\{|V \setminus Y_0|, |Y_0|\} = |V \setminus Y_0|$. Hence, trivially one has that $|V \setminus X_0| \leq |V \setminus Y_0|$. Choose injection $\iota : V \to V$ such that $\iota \restriction X_0 : X_0 \to Y_0$ is a bijection and $\iota \restriction (V \setminus X_0) : V \setminus X_0 \to V \setminus Y_0$ is an injection. Then since $A[E]$ is injection invariant on abstracts, we have that $E(X_0, \overline{\iota}(X_0))$, which is the same as $E(X_0, Y_0)$. □

Finally, let's note an instructive corollary to Theorem 1.1. This corollary tells us that in the context of natural relative categoricity, the only definable isomorphisms between the induced structures are identical to the natural bijection:

COROLLARY 4.1. *Suppose that $A[E]$ is naturally relatively categorical. Suppose that $\mathcal{M}$ is a model of $A^2[E]$ with induced structures $\mathcal{M}_1$ and $\mathcal{M}_2$. Suppose that $\overline{H} : \mathcal{M}_1 \to \mathcal{M}_2$ is a map whose restriction $H \restriction \mathrm{rng}(\partial_1)$ to $\mathrm{rng}(\partial_1)$ is an element of $S_2[M]$. If $\overline{H} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism, then it is equal to the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$.*

Proof. So suppose that $\overline{H} : \mathcal{M}_1 \to \mathcal{M}_2$ is an isomorphism. To show that $\overline{H}$ is equal to the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \to \mathcal{M}_2$, it suffices by Proposition 3.5 to show that

$$\mathcal{M} \models [\forall X\ (X \subseteq \mathrm{rng}(\partial_1) \to E(X, \overline{H}(X))] \tag{4.11}$$

Suppose that $X \subseteq \mathrm{rng}(\partial_1)$, and let $Y = \overline{H}(X)$, so that $Y \approx X$ and $X \subseteq \mathrm{rng}(\partial_1)$. Since $A[E]$ is naturally relatively categorical, by Theorem 1.1 it is cardinality coarsening on abstracts (1.3), from which we can infer $E(X, Y)$, which is the same as $E(X, \overline{H}(X))$, so that indeed the previous equation is satisfied. □

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Let's turn now to the proof of our other main theorem, namely, Theorem 1.2:

Proof. First suppose that $A[E]$ is surjectively relatively categorical (cf. Definition 3.6). Suppose that $(M, S_1[M], S_2[M], \ldots, <, \partial_1)$ is a model of $A[E]$ where $\partial_1 : S_1[M] \rightarrow M$ is a surjection, and suppose that $\pi : M \rightarrow M$ is a bijection. Let $\partial_2 : S_1[M] \rightarrow M$ be defined by $\partial_2 = \pi \circ \partial_1$. Then since $\pi : M \rightarrow M$ is an injection, we have that the following holds in our model:

$$\partial_2(X) = \partial_2(Y) \Longleftrightarrow \partial_1(X) = \partial_1(Y) \Longleftrightarrow E(X, Y) \tag{4.12}$$

Hence $\mathcal{M} = (M, S_1[M], S_2[M], \ldots, <, \partial_1, \partial_2)$ is a model of $A^2[E]$. Further, since $\pi : M \rightarrow M$ is an surjection, we have that $\partial_2 : S_1[M] \rightarrow M$ is a surjection. Then since by the hypothesis of $A[E]$ being surjectively relatively categorical, we have that the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is an isomorphism of the induced structures (cf. equation (3.3)). Then by surjective relative categoricity (cf. equation (3.14)), we have that $\mathcal{M}$ models $E(X, \overline{\Gamma}(X))$ for any $X \in S_1[M]$. But for any $X \in S_1[M]$, one may use the surjectivity of the operators to check that $\overline{\pi}(X) = \overline{\Gamma}(X)$, which of course implies that $E(X, \overline{\pi}(X))$, so that we are done.

Now suppose that $A[E]$ plus the surjectivity of the abstraction operator proves permutation invariance (1.5). Then we show that $A[E]$ is surjectively relatively categorical. So suppose that $\mathcal{M}$ is a model of $A^2[E]$ wherein ∂_1, ∂_2 are surjective. Then the natural bijection is a bijection $\Gamma : V \rightarrow V$. Then by permutation invariance (1.5), we have that $E(X, \overline{\Gamma}(X))$ for all X. Then by equation (3.14), we have that $\overline{\Gamma} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is an isomorphism.

So we've shown that surjective relative categoricity is equivalent to permutation invariance, assuming that the abstraction operator is surjective. Now we show that these two conditions are equivalent to the condition of bicardinality coarsening, under the hypothesis that the abstraction operator is surjective. First suppose that that $A[E]$ plus the surjectivity of the abstraction operator proves permutation invariance (1.5). Then suppose that X, Y such that $|X| = |Y|$ and $|V \setminus X| = |V \setminus Y|$. Any two witnessing bijections can be conjoined into a bijection $\pi : V \rightarrow V$ such that $\overline{\pi}(X) = Y$ and $\overline{\pi}(V \setminus X) = V \setminus Y$. Then by permutation invariance (1.5), we have that $E(X, \overline{\pi}(X))$ or $E(X, Y)$, which is what we wanted to establish. Finally, suppose that $A[E]$ plus the surjectivity of the abstraction operator proves bicardinality coarsening (1.6). Suppose that $\pi : V \rightarrow V$ is a bijection. Let X be a concept and let $Y = \overline{\pi}(X)$. Then since π is a bijection, we have that $|X| = |Y|$ and $|V \setminus X| = |V \setminus Y|$. So by bicardinality coarsening (1.6), it follows that $E(X, Y)$ or $E(X, \overline{\pi}(X))$. □

In analogue to Corollary 4.1, we have the following result showing that the only isomorphism is the natural bijection in the setting of surjective relative categoricity. We omit the proof since it is entirely analogous to the proof of this earlier corollary.

COROLLARY 4.2. *Suppose that $A[E]$ is surjectively relatively categorical. Suppose that $\mathcal{M}$ is a model of $A^2[E]$, where the abstraction operators are surjective, with induced structures $\mathcal{M}_1$ and $\mathcal{M}_2$. Suppose that $\overline{H} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is a map whose restriction $\overline{H} \restriction \mathrm{rng}(\partial_1)$ to $\mathrm{rng}(\partial_1)$ is an element of $S_2[M]$. If $\overline{H} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ is an isomorphism, then it is equal to the natural bijection $\overline{\Gamma} : \mathcal{M}_1 \rightarrow \mathcal{M}_2$.*

So the equivalent characterizations featuring in Theorems 1.1 and 1.2 concern the abstraction principle $A[E]$. It will be useful to have analogues of these for the underlying equivalence relations E as well. So we define:

$$\textit{Injection Invariant} :\models [\forall \text{ injection } \iota : V \rightarrow V \ \forall X \ E(X, \bar{\iota}(X))] \tag{4.13}$$

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Permutation Invariant :$\models [\forall \text{ bijection } \pi : V \to V \ \forall X \ E(X, \overline{\pi}(X))]$ (4.14)

Cardinality Coarsening :$\models [\forall X, Y \ (|X| = |Y| \to E(X, Y))]$ (4.15)

Bicardinality Coarsening :

$$\models [\forall X, Y \ ((|X| = |Y| \ \& \ |V \setminus X| = |V \setminus Y|) \to E(X, Y))] \tag{4.16}$$

Injection Invariant on Small Concepts :

$$\models [\forall \text{ injection } \iota : V \to V \ \forall X \ (|X| \leq \left|\frac{P(V)}{E}\right| \to E(X, \overline{\iota}(X)))] \tag{4.17}$$

Cardinality Coarsening on Small Concepts :

$$\models [\forall X, Y \ (|Y| = |X| \leq \left|\frac{P(V)}{E}\right| \to E(X, Y))] \tag{4.18}$$

In the statements of these notions, the $\models$ relation is the deduction relation from the deduction system of second-order logic which is the background of all our theories. In the final two conditions (4.17)–(4.18), the clause $|X| \leq \left|\frac{P(V)}{E}\right|$ just means that there's an injection from X to the E-equivalence classes of second-order objects. This can be written in second-order logic as follows:

$$\exists R \ \forall x, y \in X \ (x \neq y \to \neg E(R[x], R[y])) \tag{4.19}$$

wherein R is a binary relation and where $R[x] = \{z : Rxz\}$. These conditions (4.13)–(4.18) are all conditions on an arbitrary formula $E(X, Y)$, which we assume to be an L_0-formula (cf. Definition 2.1) which is provably an equivalence relation in our background second-order logic.

In diagrammatic form, the relationship between these notions are displayed in Figure 1. In the figure, the arrows going in both directions– i.e. "↔"– indicate a provable equivalence, while the double-lined arrows that go in only one direction– i.e. "⇒"– indicate that the implication cannot be reversed. The implications between e.g. injection invariance and cardinality coarsening follows automatically from the analogous part of the proof of Theorem 1.1. The only implications in Figure 1 that are less immediate are the implications from the conditions in the middle column to the conditions in the far-right column. So let's show that if E is cardinality coarsening on small concepts then $A[E]$ is cardinality coarsening on abstracts. To see this, it suffices to note that the Axiom of Choice AC (2.7) implies that

$$|\mathrm{rng}(\partial)| \leq \left|\frac{P(V)}{E}\right| \tag{4.20}$$

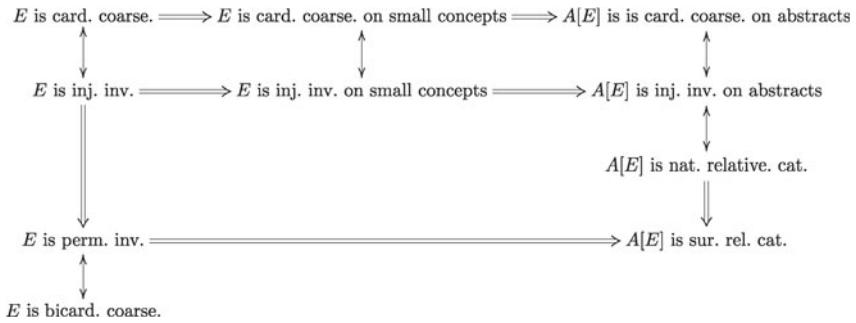


Fig. 1. Provability relation among conditions on E and $A[E]$.

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

For, for each x from $\mathrm{rng}(\partial)$ there is X such that $\partial(X) = x$. Then by AC (2.7), there is R such that for all x from $\mathrm{rng}(\partial)$ one has that $\partial(R[x]) = x$, where $R[x] = \{y : Rxy\}$ which exists by comprehension. Then suppose that $x \neq y$ are both from $\mathrm{rng}(\partial)$ but $E(R[x], R[y])$. Then by $A[E]$ we have $x = \partial(R[x]) = \partial(R[y]) = y$, a contradiction. So indeed we have (4.20).

Let us end this section by discussing the witnessing counterexamples featuring in Figure 1. For an example of an E which is bicardinality coarsening but not cardinality coarsening, see the discussion of the Bicardinality Principle in 5.3 There are different kinds of witnessing counterexamples which show that the conditions on $A[E]$ don't imply the analogous conditions on E. One way to see this is to consider $A[E]$ which are inconsistent. For instance, take $E_0(X, Y) \equiv X = Y$. Then $A[E_0]$ is just the abstraction principle Basic Law V from Frege's *Grundgesetze*, and as we noted earlier this is inconsistent with full comprehension (which we are assuming in this paper). Since the conditions on $A[E]$ are conditions on derivability in $A[E]$, trivially the inconsistent $A[E_0]$ satisfies all of these. One might hope that the right-most arrows could be reversed with the additional assumption of the consistency of $A[E]$. But this is not the case. For consider

$$E(X, Y) \equiv (|V| < \omega \;\&\; X = Y) \vee (|V| \geq \omega \;\&\; X \approx Y) \tag{4.21}$$

Again, for our abbreviations of cardinality-related notions, see (2.4)–(2.6) from Section 2. So on finite domains, E acts like E_0. But for $|V| \geq \omega$, E is just equinumerosity. The principle $A[E]$ is also naturally relatively categorical, since $A[E]$ only has models with infinite first-order domains, and in those models, E acts exactly like equinumerosity, which we showed to be naturally relatively categorical in the earlier paper (Walsh, 2012, proposition 14, p. 1687; see also Section 5). However, E is not cardinality coarsening on small concepts (or bicardinality coarsening), since if $1 < |V| < \omega$ and $a \neq b$ are distinct objects then one has that $\neg E(\{a\}, \{b\})$.

Second, let's show that cardinality coarsening on small concepts does not imply cardinality coarsening. First define the auxiliary formula:

$$\theta(X, Y) \equiv |X| = |Y| \wedge [(X \neq V \;\&\; Y \neq V) \vee (X = V \;\&\; Y = V)] \tag{4.22}$$

Consider then the following equivalence relation E:

$$\begin{aligned} E(X, Y) \;\equiv\;& |V| = \omega_1 \rightarrow [(|X| = |V| \vee |Y| = |V|) \rightarrow \theta(X, Y)] \\ \wedge\;& |V| = \omega_1 \rightarrow [(|X| < |V| \vee |Y| < |V|) \rightarrow |X| = |Y|] \\ \wedge\;& |V| \neq \omega_1 \rightarrow |X| = |Y| \end{aligned} \tag{4.23}$$

Intuitively, this says that if the domain V is the first uncountable cardinal, then V is the only member of its own equivalence class, while all the other concepts of the same cardinality as V form an equivalence class, and then properly smaller concepts are separated into equivalence classes according to cardinality; while if the domain V is any other size, then all the concepts are separated into equivalence classes according to cardinality. Note that $|V| = \omega_1$ can be written in second-order logic in a standard way: for, since we have global choice in the background, it simply consists in the claim that the global well-order of V has a limit point, and that any initial segment of the global well-order is bijective with the initial segment corresponding to the first limit point, but that V itself is not bijective with this initial segment.

Let's verify that this E from (4.23) is cardinality coarsening on small concepts but not cardinality coarsening. For the former, there are two cases to consider. First suppose that the domain V has cardinality ω_1. Then by construction $\left|\frac{P(V)}{E}\right| = \omega$. Suppose that $|Y| = |X| \leq \omega$. Then by definition $E(X, Y)$ since E is sameness of cardinality on

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

smaller concepts. Second suppose that the domain of V is any other cardinality. Then by construction E is just sameness of cardinality and so we are done. Now, let us note that E is not cardinality coarsening. For, consider a model where the domain V has size ω_1. Consider $Y_0 = V$ and $X_0 = Y_0 \setminus \{y_0\}$, where y_0 is any element of $Y_0 = V$. Then by construction we have that $\neg E(X_0, Y_0)$, while they of course have the same cardinality.

§5. Some examples and non-examples. In this section, we give some examples and non-examples of naturally relatively categorical abstraction principles. In the case of abstraction principles which are not naturally relatively categorical, we have always been able to find a single sentence which illustrates this. So we define:

DEFINITION 5.1. *The abstraction principle $A[E]$ is* relatively elementary equivalent *if whenever $\mathcal{M}$ is a model of $A^2[E]$ as in equation (3.2), then for any sentence φ in the signature $L_0[\partial]$ of the induced structures $\mathcal{M}_i$ from equation (3.3) one has that $\mathcal{M}_1 \models \varphi$ if and only if $\mathcal{M}_2 \models \varphi$. If $A[E]$ is not relatively elementarily equivalent, then a sentence φ of $L_0[\partial]$ such that there is a model $\mathcal{M}$ of $A^2[E]$ with $\mathcal{M}_1 \models \varphi$ and $\mathcal{M}_2 \models \neg\varphi$ (or vice-versa) is called a* witness *to the failure of relative elementary equivalence.*

Further, note that natural relative categoricity trivially implies relative elementary equivalence. Prior to stepping into the examples, let us mention that there's an obvious sense in which relative elementarily equivalence is a more apposite formalization of determinacy of truth-value ideas than natural relative categoricity itself. However, it seems difficult to study relative elementarily equivalence directly, and so we study natural relative categoricity instead, since it is the obvious sufficient condition for relative elementarily equivalence in this setting. It is not even obvious to us whether these notions are distinct in the setting of abstraction principles:

QUESTION 5.2. *Suppose that $E(X, Y)$ is an L_0-formula which is provably an equivalence relation in our background second-order logic and such that $A[E]$ is consistent. Suppose further that $A[E]$ is relatively elementarily equivalent. Is it necessarily the case that $A[E]$ is naturally relatively categorical?*

One suspects that the answer to this question is 'no,' but we have been unable to produce a counterexample.

5.1. Hume's Principle. Recall that Hume's Principle is the abstraction principle associated to the equivalence relation of equinumerosity. In (Walsh, 2012, proposition 14, p. 1687), it was shown that Hume's Principle is naturally relatively categorical. In that paper, the result was established by hand. But with our main Theorem 1.1 now in place, we can reduce this to the one-line observation that equinumerosity is trivially cardinality coarsening on abstracts.

5.2. Boolos' New V. In his Boolos (1989), Boolos drew attention to the following equivalence relation:

$$E(X, Y) \equiv (|X| < |V| \vee |Y| < |V|) \rightarrow X = Y \tag{5.1}$$

Intuitively, this equivalence identifies all the concepts bijective with the universe V, but acts like Basic Law V on small concepts. By Theorem 1.1, it's easy to see that this equivalence relation is not naturally relatively categorical. For, it is trivially not cardinality coarsening on abstracts. Indeed, any two distinct small concepts which are equinumerous will not be equivalent under this equivalence relation. Further, it turns out that New V is not relatively elementarily equivalent (5.1), and that the witnessing sentence was previously

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

studied by Jané & Uzquiano (2004). For, they noted that some models of New V generated membership relations which were well-founded while others did not.

To see this, enumerate the finite subsets of ω as $X_1, X_2, X_3, \ldots$, and without loss of generality, suppose that $1 \in X_1$. Then let $\partial_1 : P(\omega) \to \omega$ be any surjection that sends all the infinite subsets of ω to 0, and sends X_1 to 1. Further, let $\partial_2 : P(\omega) \to \omega$ be the surjective map that sends all the infinite subsets of ω to 0, and acts on the finite subsets of ω as follows:

$$\partial_2(X_i) = \min(\omega \setminus (X_i \cup \{0, \partial(X_1), \ldots, \partial(X_{i\text{-}1)}\})) \tag{5.2}$$

Then $\mathcal{M} = (\omega, P(\omega), P(\omega \times \omega), \ldots, \partial_1, \partial_2)$ is model of $A^2[E]$ (1.2). Since both maps are surjective, we have that the induced structures $\mathcal{M}_i$ from equation (3.3) will satisfy $\mathcal{M}_i = (\omega, P(\omega), P(\omega \times \omega), \ldots, \partial_i)$.

If $A[E]$ were naturally relatively categorical, then $A[E]$ would be relatively elementarily equivalent. But let φ be the following sentence in the signature of $L_0[\partial]$:

$$\varphi \equiv \exists\, X\, \exists\, b\, (|X| < |V| \;\&\; b = \partial(X) \;\&\; Xb) \tag{5.3}$$

Then it's easy to see that $\mathcal{M}_1 \models \varphi$ since $b = \partial_1(X_1)$ is a witness. However, by construction, $\mathcal{M}_2 \models \neg\varphi$. So φ is a witness to the failure of relative elementarily equivalence of New V, so that New V is not naturally relatively categorical. Boolos was interested in New V because it allowed one to define an ersatz membership relation:

$$a \eta b \iff \exists\, X\, (|X| < |V| \;\&\; \partial(X) = b \;\&\; Xa) \tag{5.4}$$

Expressed in these terms, φ from (5.3) says that there's a b with $b\eta b$, which intuitively says that the membership relation from (5.4) is not well-founded.

5.3. Bicardinality. Consider the equivalence relation:

$$E(X, Y) \equiv |X| = |Y| \wedge |V \setminus X| = |V \setminus Y| \tag{5.5}$$

It's easy to see that this is provably an equivalence relation in our background second-order logic. Further, this equivalence relation is trivially an example of an equivalence relation which is bicardinality coarsening but not cardinality coarsening, since for instance in infinite structures V and $V \setminus \{a\}$ will have the same cardinality but will not be E-equivalent. Let's call the associated principle $A[E]$ the *Bicardinality Principle*. By Theorem 1.2, we have that the Bicardinality Principle is surjectively relatively categorical. However, it turns out that the Bicardinality Principle is not relatively elementarily equivalent, and hence not naturally relatively categorical.

The non-relative elementary equivalence of the Bicardinality Principle is related to the different ways this principle can interpret the natural numbers. Just as with Hume's Principle, the Bicardinality Principle allows one to build a copy of the natural numbers $\{\overline{0}, \overline{1}, \ldots, \overline{n}, \overline{n+1} \ldots\}$ by setting $\overline{0} = \partial(\emptyset)$ and $\overline{n+1} = \partial(\{\overline{0}, \ldots, \overline{n}\})$. But the Bicardinality principle allows one to build a second copy $\{\widehat{0}, \widehat{1}, \ldots, \widehat{n}, \widehat{n+1} \ldots\}$ of the natural numbers by setting $\widehat{0} = \partial(V)$ and $\widehat{n+1} = \partial(V \setminus \{\widehat{0}, \ldots, \widehat{n}\})$. Then we can define the following sentences for each $n > 0$:

$$\varphi_n \equiv \bigwedge_{m=0}^{n-1} \overline{m} \neq \overline{n}, \qquad \widehat{\varphi}_n \equiv \bigwedge_{m=0}^{n-1} \widehat{m} \neq \widehat{n}, \tag{5.6}$$

Then one has that the Bicardinality Principle proves each φ_n and each $\widehat{\varphi}_n$. The proof is very similar to the proof that Hume's Principle proves the variant ψ_n of φ_n wherein $\overline{n}$ is

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

defined in terms of # instead of ∂ (where again, # is the symbol reserved for the abstraction operator featuring in Hume's Principle). Indeed, the sentences ψ_n are part of the means by which one establishes the principle Nq mentioned in the introduction Section 1 (cf. Walsh, 2014b, equation (34), p. 111).

However, we have that $\widehat{\varphi}_n$ for each $n > 0$ is a witness to the Bicardinality Principle being not relatively elementarily equivalent. Consider a model of the Bicardinality Principle of the form

$$\mathcal{M}^* = (\omega, P(\omega), P(\omega^2), \ldots, \partial_1) \tag{5.7}$$

wherein $\partial_1 : P(\omega) \to \omega$ is a surjection (cf. Proposition 6.4). Define the injection $\iota : \omega \to \omega$ by $\iota(n) = 2n$, and define $\partial_2 = \iota \circ \partial_1$. Then by an argument with which we are now familiar (cf. equations (4.7) and (4.12)), we have that the following is also a model of $A^2[E]$:

$$\mathcal{M} = (\omega, P(\omega), P(\omega^2), \ldots, \partial_1, \partial_2) \tag{5.8}$$

Since ∂_1 is surjective, we have that the induced structure $\mathcal{M}_1$ from equation (3.3) is identical to the structure $\mathcal{M}^*$ from (5.7), and so $\mathcal{M}_1$ models $\widehat{\varphi}_1$ since it is a model of the Bicardinality Principle. However, if $\mathcal{E}$ denotes the evens, then we have that the induced structure $\mathcal{M}_2$ is equal to the following structure

$$\mathcal{M}_2 = (\mathcal{E}, P(\mathcal{E}), P(\mathcal{E}^2), \ldots, \partial_2 \restriction P(\mathcal{E})) \tag{5.9}$$

Then let $D = \mathcal{E} \setminus \{\partial_2(\mathcal{E})\}$. Then $D, \mathcal{E}$ have cardinality ω, as do the their relative complements $\omega \setminus D, \omega \setminus \mathcal{E}$ in ω. Hence since $\mathcal{M}$ models the Bicardinality Principle, we have that $\partial_2(\mathcal{E}) = \partial_2(D)$. But this implies that on $\mathcal{M}_2$, we have that $\widehat{0} = \widehat{1}$, so that $\mathcal{M}_2$ does not model $\widehat{\varphi}_1$. The same argument works for $\widehat{\varphi}_n$ for all $n \geq 1$.

5.4. The Nuisance Principle. Consider the equivalence relation $E(X, Y) \equiv |X \triangle Y| < \omega$ wherein $X \triangle Y$ denotes the symmetric difference $X \triangle Y = (X \setminus Y) \cup (Y \setminus X)$. In this, recall that the notion of $|Z| < \omega$ is that of Dedekind-finiteness (2.6), so that E is L_0-definable. But per Proposition 2.2, this aligns with the notion of being finite in the sense of being bijective with the initial segments $I_a = \{x : x < a\}$ from our global well-order where a is finite, i.e. is less than the first limit point. So while E is expressible purely in terms of the signature of the background second-order logic L_0 (cf. Definition 2.1), we can use the global well-order to show that it is an equivalence relation. For, using the global well-order we can show that the union of any two finite sets is finite, and so the transitivity of E follows from $X \triangle Z \subseteq (X \triangle Y) \cup (Y \triangle Z)$.

The abstraction principle $A[E]$ has been called the "Nuisance Principle" following Wright's identification of its abstracts as "nuisances" (cf. Wright, 1997), and is related to an earlier principle presented by Boolos (1990). Let's show that $A[E]$ is naturally relatively categorical, and let's begin with the following proposition. This proposition is well-known if one assumes the standard semantics, but we know of no extant proof for the Henkin semantics with choice principles. In the statement of this theorem, the notion of finiteness is Dedekind-finiteness (2.6) or the aforementioned equivalent characterization (cf. again Proposition 2.2).

PROPOSITION 5.3. *$A[E]$ implies that V is finite.*

Proof. So suppose not. Then let's work deductively in the theory $A[E]$ under the assumption that the universe V is infinite. Let ω denote the least limit point in the global well-order. First let's establish that there's a way to enumerate all finite concepts:

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

CLAIM 5.4. *There is a ternary relation R such that for all X, $|X| < \omega$ if and only if there is $a < \omega$ and b such that $R[a, b] = X$, where again $R[a, b] = \{c : R(a, b, c)\}$.*

This claim follows trivially by comprehension from the following claim:

CLAIM 5.5. *There is a ternary relation R such that for all $a < \omega$, and all X with $|X| \leq |I_a|$ there is b with $R[a, b] = X$, where again $R[a, b] = \{c : R(a, b, c)\}$.*

This claim of course follows from the following claim by a single application of AC (2.7):

CLAIM 5.6. *For all $a < \omega$ there is binary relation R_a such that for all X with $|X| \leq |I_a|$ there is b with $R_a[b] = X$.*

We argue for this latter claim by induction on a. For $a = 0$, it holds trivially since we may choose R_a equal to the empty binary concept. Now suppose it holds for a_0 with witness R_{a_0}, and suppose $a = s(a_0)$, where this is the partial successor operation from (2.14). By Infinite Products are Maxs (2.12), there is a bijection $\langle \cdot, \cdot \rangle : V \times V \to V$. Then we define the binary relation R_a as follows:

$$R_a[b] = \{z : \exists\, b_0, c_0\ b = \langle b_0, c_0 \rangle \ \&\ (z = c_0 \vee R_{a_0}[b_0](z))\} \tag{5.10}$$

So this finishes the proof of Claim 5.6 and with it the proof of Claim 5.4.

Using this enumeration of the finite concepts, let's define a partial injective map from finite concepts to objects. For this, let us fix a ternary relation R as in Claim 5.4. Then consider the following partial map from concepts to objects:

$$\nabla(X) = c \iff |X| < \omega \ \&\ c = \min_{<}\{\langle a, b \rangle : a < \omega \ \&\ R[a, b] = X\} \tag{5.11}$$

Then this map is defined on all finite concepts, and on these, it is an injection. For, supposing that $\nabla(X) = \langle a, b \rangle = \nabla(Y)$, we have $X = R[a, b] = Y$. Hence, indeed ∇ is a partial map from concepts to objects which is defined and injective on finite concepts.

So now consider $\mathrm{rng}(\partial)$, where ∂ as usual is the abstraction operator associated to $A[E]$. As in the argument for (4.20) in the previous section, we may use AC (2.7) to show that there is a binary relation S such that for all x from $\mathrm{rng}(\partial)$ one has that $\partial(S[x]) = x$. Then for all concepts X, there is unique x from $\mathrm{rng}(\partial)$ such that $E(X, S[x])$. Hence there are finite concepts F_X, G_X such that $X = (S[x] \cup F_X) \setminus G_X$. Then define a map ∂' from concepts to objects by $\partial'(X) = \langle x, \langle a, b \rangle \rangle$ iff x is from $\mathrm{rng}(\partial)$ and $E(X, S[x])$ and

$$\langle a, b \rangle = \min_{<}\{\langle c, d \rangle : \exists\, F_X, G_X\ X = (S[x] \cup F_X) \setminus G_X \ \&\ \nabla(F_X) = c \ \&\ \nabla(G_X) = d\} \tag{5.12}$$

Then by construction ∂' is an injection from concepts to objects, and so using the full comprehension schema, we may again replicate the Russell paradox to derive a contradiction. □

Using this result, we can now deduce:

PROPOSITION 5.7. *$A[E]$ is naturally relatively categorical.*

Proof. Let's work within $A[E]$. By the above result, V is finite. Hence, since $|X \triangle Y| < \omega$ for all finite X and Y, we have that $E(X, Y)$ for all X, Y. And then trivially $A[E]$ is cardinality coarsening on abstracts. Hence by Theorem 1.1, $A[E]$ is naturally relatively categorical. □

So this proof shows us that naturally relatively categorical abstraction principles are not pairwise consistent. For, as noted earlier, Hume's Principle is naturally relatively categorical

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

and yet it implies that V is infinite, since for instance when one defines the natural numbers using the resources of Hume's Principle, they will be Dedekind-infinite. But of course one would not have expected relatively categorical concepts to be pairwise consistent. For instance, the standard relative categoricity argument for set theory can be deployed to show that the standard axioms plus "there are no inaccessibles", as well as the standard axioms plus "there are inaccessibles," are both relatively categorical, at least assuming (as we are here) that the domain of the model is the entire universe. In Section 6, we'll note by contrast that equivalence relations which are cardinality coarsening on small concepts are pairwise consistent. So this is another way of seeing that the conditions on $A[E]$ in Figure 1 don't have implications for the analogous conditions on E: while the Nuisance Principle is naturally relatively categorical and hence cardinality coarsening on abstracts, it is not the case that the underlying equivalence relation is cardinality coarsening on small concepts.

5.5. The complementation principle. Let's say that a concept X is a *complement*, abbreviated $\mathrm{Cmp}(X)$ if there is $Y \approx X$ such that $X \sqcup Y = V$. Then we define:

$$E(X, Y) \equiv (\mathrm{Cmp}(X) \vee \mathrm{Cmp}(Y)) \rightarrow (X = Y \vee (X \approx Y \;\&\; X \sqcup Y = V)) \tag{5.13}$$

This equivalence relation sorts out concepts as follows: each pair of equinumerous complementary sets constitute a separate equivalence class, everything else occupies a single "junk" equivalence class.

This abstraction principle has finite models only for odd domains, and domains of sizes 2 and 4. For, in the case of odd domains, there are no complements and so there is only one equivalence class. In the case of even domains of size $2n$, the number of E-equivalence classes is exactly $f(n) = \frac{1}{2}\binom{2n}{n} + 1$, and one has that $f(1) = 2$, $f(2) = 4$ and $f(n) \geq 2n + 1$ for $n \geq 3$, as one can establish by induction on n.

Unlike the Nuisance Principle, this $A[E]$ is not naturally relatively categorical and indeed not relatively elementary equivalent. For, consider the domain $M = \{a, b, c, d\}$ of size 4, where there are exactly 4 equivalence classes. Define two abstraction operators by

$$\begin{array}{llll} \partial_1(\{a, b\}) = a, & \partial_1(\{a, c\}) = b, & \partial_1(\{a, d\}) = c, & \partial_1(\emptyset) = d, \\ \partial_2(\{a, b\}) = c, & \partial_2(\{a, c\}) = b, & \partial_2(\{a, d\}) = a, & \partial_2(\emptyset) = d \end{array} \tag{5.14}$$

This determines a model $\mathcal{M}$ of $A^2[E]$ with first-order part M. Then consider the sentence:

$$\forall X\, [|X| = 2 \rightarrow \exists Y (E(X, Y) \wedge Y(\partial_i(X)) \wedge Y(\partial_i(\emptyset))] \tag{5.15}$$

This sentence fails in $\mathcal{M}_1$ with witness $X = \{a, b\}$ but holds in $\mathcal{M}_2$. Hence $A[E]$ is not relatively elementary equivalent and hence not naturally relatively categorical. But like with the Nuisance Principle, $A[E]$ proves V is finite. Hence, it is not this feature alone which permitted the Nuisance Principle to be naturally relatively categorical.

§6. Stability criteria and the bad company problem. One of locus of activity on abstraction principles in recent years has been stability criteria (cf. Cook, 2012). Some of the key stepping stones in this hierarchy are the following:

> The abstraction principle $A[E]$ is *stable* if there is some cardinal λ such that, for all $\kappa \geq \lambda$, $A[E]$ has a standard model of size κ. (6.1)

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

The abstraction principle $A[E]$ is *strongly stable* if there is some cardinal λ such that, for all cardinals κ, $A[E]$ has a standard model of size κ if and only if $\kappa \geq \lambda$. (6.2)

As above, a structure $\mathcal{M} = (M, S_1[M], S_2[M], \ldots)$ is said to be *standard* if $S_n[M] = P(M^n)$ for each $n \geq 1$ (cf. Section 2). Further, here one identifies the *cardinality of* $\mathcal{M}$ as $|M|$, i.e. the cardinality of its first-order part. So this convention on the cardinality of the many-sorted structure $\mathcal{M}$ differs from other settings in which it would be more natural to define its cardinality as the cardinality of the union of its various domains (cf. Walsh, 2014b, sec. 6.1, p. 107, equation (18); Manzano, 1996, p. 231; Ebbinghaus, 1985, pp. 32, 64).

These criteria have been proposed as solutions to the so-called *Bad Company Problem*, namely the problem of identifying criteria on abstraction principles which would ensure consistency of $A[E]$; perhaps also joint consistency with like abstraction principles and the ability to interpret mathematics of various sorts. Of course, the relative categoricity notions studied here are not candidates for a solution to the Bad Company Problem since they do not ensure consistency. However, the sufficient conditions on the equivalence relations themselves that we have identified for relative categoricity *do* ensure consistency, and indeed stability and joint consistency. In particular, we have the following results:

PROPOSITION 6.1. *Suppose that $E(X, Y)$ is an L_0-formula and suppose that E is cardinality coarsening on small concepts. Then for all infinite cardinals κ there is a standard model of $A[E]$ of cardinality κ.*

PROPOSITION 6.2. *Suppose that $E(X, Y)$ is an L_0-formula and suppose that E is bicardinality coarsening. Then for all infinite cardinals κ there is a standard model of $A[E]$ of cardinality κ.*

These two propositions imply that the equivalence relations satisfying these constraints are stable (6.1). However, this cannot be improved to strong stability (6.2). For, consider the following equivalence relation:

$$E(X, Y) \equiv [|V| < \omega \;\&\; |V| \text{ is even} \rightarrow X \approx Y] \wedge [\neg(|V| < \omega \;\&\; |V| \text{ is even}) \rightarrow X = X] \tag{6.3}$$

This equivalence relation is cardinality coarsening on small concepts. In the case where $|V| < \omega$ is even, we have that E is equinumerosity, so that $\left|\frac{P(V)}{E}\right| \geq |V| + 1$, all concepts are small, and thus trivially $X \approx Y$ implies $E(X, Y)$. But in the other case, E says that all concepts are equivalent, so that again we trivially have that $X \approx Y$ implies $E(X, Y)$. So this E is cardinality coarsening on small concepts, but $A[E]$ is not strongly stable (6.2). For $A[E]$ is satisfiable on all finite odd domains but no finite even domains.

Cook (2012) recommends strong stability as the appropriate solution to the Bad Company Problem, and so the previous examples indicate that the solution offered by cardinality coarsening on small concepts and bicardinality coarsening will be extensionally distinct from Cook's solution. However, it is worth noting that Propositions 6.1 and 6.2 imply that if E is cardinality coarsening on small concepts (respectively, bicardinality coarsening), all structures witnessing that E is not strongly stable will be finite.

But cardinality coarsening on small concepts and bicardinality coarsening have a more mixed scorecard on other desiderata taken to be relevant to judging the success of proposed solutions to the Bad Company problem. On the one hand, Proposition 6.1 and Proposition 6.2 imply that all these abstraction principles are jointly consistent since they

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

all have models of any infinite cardinality. But by the same token, our conditions on E restrict the mathematics that can be recovered in the theory $A[E]$. Inspection of the below proof of these propositions show that the joint theory of all these abstraction principles will be interpretable in second-order Peano arithmetic $\mathtt{PA}^2$ (cf. Simpson, 1999, pp. 4 ff for the definition of $\mathtt{PA}^2$ and cf. Walsh, 2014b, sec. 2 for the definition of interpretation). This is of course because a standard model of cardinality ω is the underlying first-order domain of the standard model of $\mathtt{PA}^2$, and the proofs of the propositions for the case $\kappa = \omega$ are formalizable in $\mathtt{PA}^2$. Hence, there will be no hope of interpreting, e.g., standard $\mathtt{ZFC}$-set theory by restricting attention to such abstraction principles, because $\mathtt{ZFC}$ proves the consistency of $\mathtt{PA}^2$ and so $\mathtt{PA}^2$ cannot interpret $\mathtt{ZFC}$.

But it's difficult to compare this directly to the situation of stability or strong stability. For, suppose that Φ is a sentence of second-order logic that is true on all and only standard models of cardinality $\geq \lambda$. Then $A[E]$ will be strongly stable for $E(X, Y) \equiv (X = Y) \vee \Phi$ (disjunctive abstraction principles like this are due to Heck Jr., 1992). So such $A[E]$ will be able to encode a good deal of interpretability strength, at least when interpretability with parameters is allowed (cf. Walsh, 2014b, sec. 2 for the definition of interpretations with parameters). For instance, Φ could say that there is a model of $\mathtt{ZFC}^2$. It's unknown to us whether the interpretability strength– where the interpretation is *without* parameters– of the theories associated to all stable or strongly stable abstraction principles is strictly above $\mathtt{PA}^2$. This might be related to the fact that stability conditions often depend on undecidable questions in set theory. For instance, the claim that the abstraction principle New V is strongly stable is equivalent to the generalized continuum hypothesis (cf. Shapiro & Weir, 1999, p. 315; Cook, 2007, p. 297). Our conditions, by contrast, concern only what is demonstrable in a sound and complete deductive system.

So let's proceed to the proofs of Propositions 6.1 and 6.2. In these proofs, we employ the notion on equivalence classes from the close of Section 2. We begin with a helpful lemma:

LEMMA 6.3 (Coarsening Dichotomy). *Suppose that E is cardinality coarsening on small concepts. Then for all infinite κ we have $\left|\frac{P(\kappa)}{E}\right| < \kappa$ or $\left|\frac{P(\kappa)}{E}\right| = \left|\frac{P(\kappa)}{\approx}\right| = \kappa$, and hence in either case we have $\left|\frac{P(\kappa)}{E}\right| \leq \kappa$.*

Proof. So suppose that $\left|\frac{P(\kappa)}{E}\right| \geq \kappa$. We must show that $\left|\frac{P(\kappa)}{E}\right| = \left|\frac{P(\kappa)}{\approx}\right| = \kappa$. Since $\left|\frac{P(\kappa)}{E}\right| \geq \kappa$, we of course have that $X \subseteq \kappa$ implies $|X| \leq \left|\frac{P(\kappa)}{E}\right|$. Then we have:

$$[X, Y \subseteq \kappa \ \& \ |Y| = |X|] \Longrightarrow [|Y| = |X| \leq \left|\frac{P(\kappa)}{E}\right|] \tag{6.4}$$

and the consequent of this conditional clearly implies $E(X, Y)$ since E is cardinality coarsening on small concepts. In terms of the equivalence classes, this says that the partition of $P(\kappa)$ by E is coarser than the partition of $P(\kappa)$ by equinumerosity. Then choose representatives $X_0, X_1, \ldots$ for the E-equivalence classes, so that $\frac{P(\kappa)}{E} = \bigsqcup_{i \in I} [X_i]_E$. Then define $I : \frac{P(\kappa)}{E} \to \frac{P(\kappa)}{\approx}$ by $I([X_i]_E) = [X_i]_{\approx}$. Then we claim that I is an injection. For, suppose that $I([X_i]_E) = I([X_j]_E)$ so that $[X_i]_{\approx} = [X_j]_{\approx}$. Then $|X_i| = |X_j|$ and so by the previous equation we have $E(X_i, X_j)$ and hence $[X_i]_E = [X_j]_E$ and so $i = j$ since the X_i, X_j are representatives of the E-equivalence classes. Since $I : \frac{P(\kappa)}{E} \to \frac{P(\kappa)}{\approx}$ is indeed injective, we then have the following, where the last inequality appeals to the infinitude of κ: $\left|\frac{P(\kappa)}{E}\right| \leq \left|\frac{P(\kappa)}{\approx}\right| \leq \kappa$. Since we're assuming $\kappa \leq \left|\frac{P(\kappa)}{E}\right|$, we are done. □

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

So we can now prove Proposition 6.1:

Proof. Let κ be an infinite cardinal, and consider the standard structure, where $<$ is the natural well-ordering on the cardinal: $\mathcal{M}_0 = (\kappa, P(\kappa), P(\kappa \times \kappa), \ldots, <)$. By the Coarsening Dichotomy, we have that $\left|\frac{P(\kappa)}{E}\right| \leq \kappa$. Choose an injection $\widehat{\partial} : \frac{P(\kappa)}{E} \to \kappa$. Then we may define the map $\partial : P(\kappa) \to \kappa$ by $\partial(X) = \widehat{\partial}([X]_E)$. Then by construction we have that $\mathcal{M} = (\kappa, P(\kappa), P(\kappa \times \kappa), \ldots, <, \partial)$ is a model of $A[E]$ of cardinality κ. □

Similarly, we can prove Proposition 6.2:

Proof. Let's use E_0 as an abbreviation for the equivalence relation of bicardinality. So since E is bicardinality coarsening, by definition we have that $E_0(X, Y)$ implies $E(X, Y)$. So as in the proof of the Coarsening Dichotomy, this implies that $\left|\frac{P(\kappa)}{E}\right| \leq \left|\frac{P(\kappa)}{E_0}\right|$. Then one may argue that $\left|\frac{P(\kappa)}{E}\right| \leq \left|\frac{P(\kappa)}{E_0}\right| \leq \kappa \cdot \kappa \leq \kappa$. Since κ is infinite, the only non-trivial inequality is the second. For this, consider the map $[X]_{E_0} \mapsto \langle |X|, |\kappa \setminus X| \rangle$. This is an injection because we have by definition of E_0 that $E_0(X, Y)$ iff $\langle |X|, |\kappa \setminus X| \rangle = \langle |Y|, |\kappa \setminus Y| \rangle$. From $\left|\frac{P(\kappa)}{E}\right| \leq \kappa$, we can use the same construction as employed in the proof of the above proposition to build models of $A[E]$. □

Since E being bicardinality coarsening is a sufficient condition for $A[E]$ being surjectively relatively categorical, it's natural to ask about when there are standard models of $A[E]$ on which the abstraction operator is surjective.

PROPOSITION 6.4. *Suppose that E is bicardinality coarsening. Then for all infinite cardinals κ, there is a standard model of $A[E]$ of cardinality κ wherein the abstraction operator is surjective if and only if $\left|\frac{P(\kappa)}{E}\right| = |\{\lambda : \lambda < \kappa\}| = \kappa$.*

Proof. First suppose that there is a standard model of $A[E]$ of cardinality κ where the abstraction operator is surjective. Then surjectivity implies that $\kappa = \left|\frac{P(\kappa)}{E}\right|$. Then as in the proof of the Proposition 6.1, we may argue that $\kappa = \left|\frac{P(\kappa)}{E}\right| \leq \left|\frac{P(\kappa)}{E_0}\right| \leq \kappa$, so that in fact we have an equality. Since κ is infinite, one further has that $\left|\frac{P(\kappa)}{E_0}\right| = \left|\frac{P(\kappa)}{\approx}\right|$. But trivially one has that $\left|\frac{P(\kappa)}{\approx}\right| = |\{\lambda : \lambda < \kappa\}|$. For the converse, suppose that $\left|\frac{P(\kappa)}{E}\right| = |\{\lambda : \lambda < \kappa\}| = \kappa$, so that trivially $\left|\frac{P(\kappa)}{E}\right| = \kappa$. Then the model construction from the previous proposition yields a model of $A[E]$ in which the abstraction operator is surjective. □

§7. Antonelli and fine on logicality and invariance. The famous Tarski-Sher thesis on logicality (Tarski, 1986; Sher, 1991) suggests that logical notions are those that are invariant under all permutations of the domain. This can be made precise as follows. If $\pi : M \to M$ is a permutation, then π induces permutations $\overline{\pi} : M^n \to M^n$ and $\overline{\pi} : P(M) \to P(M)$. By iterating, a permutation $\pi : M \to M$ induces a permutation of any sort of ω-th order higher-order structure with first-order domain M. Then one says:

> A subset X of any sort of this structure is *permutation-invariant* if $\overline{\pi}(X) = X$ holds for all permutations $\pi : M \to M$. (7.1)

So part of what we want to do in this section is to carefully distinguish this notion coming from Tarski-Sher from our notion of permutation invariance for abstraction principles (1.5).

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

To illustrate the Tarski-Sher notion, consider $X = \{(A, B, C) \in (P(M))^3 : A \cap B = C\}$. Then for any permutation $\pi : M \to M$, we have

$$A \cap B = C \iff \overline{\pi}(A \cap B) = \overline{\pi}(C) \iff \overline{\pi}(A) \cap \overline{\pi}(B) = \overline{\pi}(C) \tag{7.2}$$

So intersection comes out logical in this sense, and similarly for all the other boolean operations.

Antonelli (Antonelli, 2010b) distinguishes several ways that abstraction principles and their associated equivalence relations can be invariant. One way is in terms of the abstraction operator itself. Since the abstraction operator is a function from concepts to objects, its graph is a certain collection of ordered pairs of concepts and objects. Hence an abstraction principle will be logical in the sense of Tarski-Sher just if this collection of ordered pairs is closed under the taking of permutations. This of course happens if and only if $\partial(X) = a$ implies $\partial(\overline{\pi}(X)) = \pi(a)$ and vice-versa for all permutations $\pi : V \to V$. And this is equivalent to the requirement that $\partial(\overline{\pi}(X)) = \pi(\partial(X))$ for all permutations $\pi : V \to V$. Antonelli calls abstraction principles which satisfy this condition *objectually invariant* (Antonelli, 2010b, p. 286). However, objectually invariant abstraction operators are rare. For instance, Hume's Principle is not objectually invariant, nor is the Bicardinality Principle. Hence, if one insisted upon the logicality of abstraction *operators* in the sense of Tarski-Sher, then as Antonelli notes this would rule out most interesting examples (cf. Antonelli, 2010b, p. 286).

The notion which we are calling permutation invariance (1.5) is called *simple invariance* by Antonelli (Antonelli, 2010b, p. 286). This notion was studied earlier in Burgess' discussion of Fine's work on abstraction principles (Burgess, 2005, p. 171). However, the notion which Fine himself was primarily interested in was the permutation invariance of the equivalence relation itself (cf. Fine, 2002, p. 111):

$$\forall \text{ bijection } \pi : V \to V \ (E(X, Y) \to E(\pi(\overline{X}), \pi(\overline{Y})) \tag{7.3}$$

This of course is just the logicality of the equivalence relation E in the sense of Tarski-Sher. However, Fine rather motivates this constraint on equivalence relations by reference to Frege's ideas about the generality of logic (Fine, 2002, p. 109, cf. MacFarlane, 2002, p. 34), and does not at all invoke the Tarski-Sher thesis. But whatever its motivation, all the equivalence relations E studied here will satisfy this constraint since they are assumed to be formulas in the background signature L_0 (cf. Definition 2.1), and so will all be invariant under permutations in the sense of (7.3).

Antonelli's work on abstraction principles and notions of permutation invariance (Antonelli, 2010b) complements his work on abstraction principles and generalized quantifiers (Antonelli, 2010a, 2010c). In this work, the idea was to think about the equivalence relations featuring in abstraction principles as examples of generalized quantifiers. For instance, the truth-condition of the sentence "Just as many students [A] as teachers [B] are hockey fans [C]," is given by the generalized quantifier $Q(A, B, C) \equiv |A \cap C| = |B \cap C|$. So one could then view Hume's Principle as yielding a way to provide generalized quantifiers with first-order truth-conditions. The connection to permutation invariance in the sense of (7.1) and (7.3) is that this is one condition among many that have been developed for assaying which generalized quantifiers actually occur in natural language (cf. Peters & Westerståhl, 2008, pp. 157, 330; Keenan & Moss, 1985, p. 77).

However, neither Antonelli nor Burgess is committed to there being any positive reason for insisting on the constraint of permutation invariance in the sense of (1.5). It clearly does not follow from the Tarski-Sher logicality constraint on the abstraction operator or on

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

the underlying equivalence relation. Further, it's not obviously required by the generality of logic since there are many equivalence relations which are expressible purely in the language L_0 of our background second-order logic which are not permutation invariant in the sense of (1.5). However, our Theorem 1.2 suggests an instrumental reason to be interested in equivalence relations which are permutation invariant in the sense of (1.5). For, it indicates that its a necessary and sufficient condition for a type of relative categoricity. So if one was interested in determinacy of truth-value in the sense of natural relative categoricity, this result would give one a reason to be interested in permutation invariance in the sense of (1.5).

Fine was himself interested in determinacy and established several categoricity results. His first result concerns what we called cardinality coarsening (4.15), a notion which Fine calls *numericality* (Fine, 2002, p. 126). This first categoricity theorem of Fine's can be stated in our terminology as follows:

THEOREM 7.1 (Fine, 2002, p. 126). *Suppose that $\mathcal{M}$ is a model of $A^2[E]$ as in equation (3.2) where E is cardinality coarsening, and suppose that $\mathcal{M}$ models $|M \setminus \mathrm{rng}(\partial_1)| = |M \setminus \mathrm{rng}(\partial_2)|$. For each $i \in \{1, 2\}$, consider the following induced models*

$$\mathcal{N}_i = (M, S_1[M], S_2[M], \ldots, \partial_i) \tag{7.4}$$

Then $\mathcal{N}_1$ is isomorphic to $\mathcal{N}_2$.

This theorem of Fine's occurs far within his book Fine (2002) which treats a great variety of topics and issues. In our view, this result deserves to be better known– for instance, it is not discussed in the *Philosophical Studies* book symposium on Fine's book (cf. Fine, 2005), nor is it treated in Burgess' discussion of Fine (Burgess, 2005, chap. 3). For this reason and for the sake of completeness, let us record its proof:

Proof. By the hypothesis, there is a bijection from $M \setminus \mathrm{rng}(\partial_1)$ to $M \setminus \mathrm{rng}(\partial_2)$. But the natural bijection Γ is a bijection from $\mathrm{rng}(\partial_1)$ to $\mathrm{rng}(\partial_2)$. By joining these two bijections, we may obtain a bijection $\Delta : M \to M$ such that $\Delta \restriction \mathrm{rng}(\partial_1)$ is the natural bijection $\Gamma : \mathrm{rng}(\partial_1) \to \mathrm{rng}(\partial_2)$. Extend to $\overline{\Delta} : \mathcal{N}_1 \to \mathcal{N}_2$ by setting $\overline{\Delta}(X) = \{\Delta(x) : x \in X\}$. Then we claim that $\overline{\Delta} : \mathcal{N}_1 \to \mathcal{N}_2$ is an isomorphism. Let $X \in S_1[M]$. We must show that $\overline{\Delta}(\partial_1(X)) = \partial_2(\overline{\Delta}(X))$. Since Δ extends the natural bijection, this is equivalent to $\partial_2(X) = \partial_2(\overline{\Delta}(X))$. And since $\mathcal{M} \models A^2[E]$, this is equivalent to $E(X, \overline{\Delta}(X))$. But since E is cardinality coarsening, of course $E(X, \overline{\Delta}(X))$ follows from $\overline{\Delta}(X) \approx X$. □

Some special cases of Fine's Theorem can be viewed as a combination of certain parts of our results. For instance, if E is cardinality coarsening, then E is bicardinality coarsening (cf. Figure 1) and hence $A[E]$ is surjectively relatively categorical by one direction of Theorem 1.2. This same result follows from the special case of Fine's theorem wherein one assumes that $M \setminus \mathrm{rng}(\partial_1)$ and $M \setminus \mathrm{rng}(\partial_2)$ are both empty. However, nothing in Fine's Theorem concerns the other direction of our results, namely that certain conditions like cardinality coarsening on small concepts and bicardinality coarsening are necessary for various forms of relative categoricity. Another difference between our work and Fine's work is with the precise induced models which figure in the statement of the results. In particular, the induced structures $\mathcal{N}_i$ in (7.4) are different from the induced structures $\mathcal{M}_i$ from (3.3) with which we have been working in that the structures $\mathcal{M}_i$ restrict their first-order domain down to the range of the abstraction operator ∂_i. Of course, these two types of induced models will align in the case where the abstraction operators are surjective. So while Fine's theorem predicts that Hume's Principle is, in our terminology, surjectively

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

relatively categorical, it does not obviously have any implications for the natural relative categoricity of Hume's Principle.

Fine's second categoricity theorem concerns a general theory of abstraction which he terms $\mathtt{GA}^+$. Roughly, this theory is the amalgamation of abstraction principles $A[E]$ which satisfy a combination of Tarski-Sher permutation invariance conditions (7.1) and varieties of stability-like conditions as discussed in Section 6. For the sake of simplicity, let's consider the case of a single such abstraction principle $A[E]$. Fine's second theorem then indicates that any model $\mathcal{M}$ of $A^2[E]$ which satisfies the constraint that $M \setminus \mathrm{rng}(\partial_1)$ and $M \setminus \mathrm{rng}(\partial_2)$ have the same cardinality will be such that the induced structures $\mathcal{N}_i$ in (7.4) are isomorphic. The official statement of Fine's second theorem is more complicated since one must precisely define the analogue $A^2[E_1, E_2, \ldots]$ of $A^2[E]$ in the case where abstraction principles associated to $E_1, E_2, \ldots$ are present. See Fine (2002, p. 189) for the precise statement of this result, and see Fine (2002, p. 170) for the precise statement of the theory $\mathtt{GA}^+$.

In the subsequent section of his book (Fine, 2002, pp. 189 ff), Fine proceeds to examine the interpretability strength of $\mathtt{GA}^+$ and related theories. It is in the context of surveying these results (cf. Burgess, 2005, p. 171) that Burgess introduces the notion what we have called permutation invariance (1.5), which does not occur in Fine's own work. In particular, Burgess introduces a theory (cf. Burgess, 2005, p. 173) in which one assumes that the equivalence relations E are permutation invariant in our sense (1.5) and further that there are at most two equivalence classes:

$$\exists\, X, Y\ \forall\, Z\ (E(Z, X) \vee E(Z, Y)) \tag{7.5}$$

This last condition is less restrictive than one might initially expect, since Burgess is working in a setting where one can consider equivalence relations not only on second-order objects, but on third order-objects, fourth-order objects, etc. However, Burgess uses this theory merely to motivate elements of Fine's theory, so that neither Fine nor Burgess suggest restricting attention to theories which satisfy these specific conditions.

As to its broader philosophical significance, Fine notes that one path to determinacy of truth-value would proceed through categoricity results like Theorem 7.1. For instance, discussing the specialization of this result to the "number of" abstraction operator, Fine writes: "Thus once we know the cardinality of the non-numbers, we are in a position to specify the truth of every arithmetical statement in purely logical terms" (Fine, 2002, p. 86, cf. p. 93). But ultimately Fine rejected this path to determinacy of truth-value, due to concerns about one's access to claims about the cardinality of parts of the domain. He writes: "The difficulty with this approach is to see how someone could grasp what these truth-conditions are without already having access to an infinite domain of abstract objects" (Fine, 2002, p. 94). This was one reason among many that led Fine to develop an alternative approach on which one postulates "a procedure for the construction of the domain" (Fine, 2006, p. 90). This "procedural postulationism" is designed to secure not only determinacy of truth-value but also determinacy of reference (Fine, 2006, p. 89; Fine, 2002, p. 100). While this is no place to discuss Fine's later procedural postulationism, it's worth underscoring that the specific worry which Fine cites with respect to categoricity results does not seem so damaging to one who would insist on the requirement of natural relative categoricity. For, unlike the hypotheses of Fine's theorem (Theorem 7.1), there are no assumptions in natural relative categoricity about the cardinality of the non-abstracts. However, Fine's concern might worry someone who insisted on the requirement of surjective relative categoricity as it is unclear how someone could have advance assurance that every object is an abstract.

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

§8. Hodes and supervaluationism. The work of Hodes (Hodes, 1984; Hodes, 1990; Hodes, 1991) constitutes a sustained attempt to forge techniques and ideas from Frege's *Grundlagen* into a viable version of fictionalism. Hodes writes: "[. . .] mathematical discourse, when carried on within the mathematical object-picture, [is] a special sort of fictional discourse: numbers are fictions 'created' with a special purpose, to encode numerical object-quantifiers and thereby enable us to 'pull down' a fragment of third-order logic, dressing it in first-order clothing" (Hodes, 1984, p. 144). Hodes differs from other fictionalists in his invocation of supervaluationist semantics, by which Hodes can say with fictionalists that mathematical language doesn't refer, but also affirm that certain sentences containing mathematical vocabulary are true in a very demanding sense.

One can implement Hodes-style supervaluationism with respect to an arbitrary equivalence relation. Hodes does this only with respect to equivalence relations associated to cardinality and sets, in Hodes (1990) and Hodes (1991) respectively. But it seems to us that the definitions offered in Hodes (1990, pp. 364–365) and Hodes (1991, p. 158) naturally generalize as follows:

DEFINITION 8.1. *Suppose that $E(X, Y)$ is an L_0-formula and κ is a cardinal, and suppose further that φ is an $L_0[\partial]$-sentence. Then φ is said to be* κ-supertrue relative to $A[E]$ *(resp.* κ-superfalse relative to $A[E]$*) if for all standard models $\mathcal{M}$ of $A[E]$ of cardinality κ, it is also the case that $\mathcal{M} \models \varphi$ (resp. $\mathcal{M} \models \neg\varphi$). Further, φ is said to be* κ-bivalent relative to $A[E]$ *if φ is κ-supertrue or κ-superfalse.*

In this definition, we are, as in the previous sections, identifying the cardinality of the model $\mathcal{M}$ of $A[E]$ with the cardinality of its first-order part.

As this definition makes clear, Hodes' approach was focused on the standard semantics for second-order logic. Moreover, there's a reason for this: if one instead used the Henkin semantics, then the analogue of supertruth would reduce to provability, at least in the case where $A[E]$ has only infinite models. This is the content of the following elementary proposition, whose proof we omit since it is a simple application of the Löwenheim-Skolem theorems.

PROPOSITION 8.2. *Suppose that $E(X, Y)$ is an L_0-formula and $A[E]$ is consistent but has only infinite models. Suppose further that φ is an $L_0[\partial]$-sentence and κ is an infinite cardinal. If all models $\mathcal{M}$ of $A[E]$ of cardinality κ satisfy φ, then $A[E]$ proves φ.*

So this proposition indicates that something is gained by Hodes' invocation of standard models, namely, the idea of supertruth does not reduce to that of deduction.

Hodes' fictionalism suggests the idea of maximizing bivalence, and so Hodes himself established various results indicating that "[. . .] our actual mathematical reasoning makes no use of sentences parsed by non-bivalent sentences [. . .]" (Hodes, 1990, p. 370, cf. Observations 4–5, pp. 367–368). With an eye towards maximizing bivalence, we introduce the following definition of bivalence-compatibility:

DEFINITION 8.3. *The abstraction principle $A[E]$ is* bivalence-compatible *if there is an infinite cardinal κ such that φ is κ-bivalent relative to $A[E]$ for each $L_0[\partial]$-sentence φ.*

There might be other ways of maximizing bivalence, but this at least seems like a natural enough route: the idea is that there's some size such that all standard models of that size must agree on the truth-value of sentences expressible using the abstraction operator.

It's then natural to ask which abstraction principles are bivalence compatible. By using Fine's Theorem 7.1 and the Coarsening Dichotomy (Lemma 6.3), one can easily show:

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

PROPOSITION 8.4. *If E is cardinality coarsening then $A[E]$ is bivalence compatible, and indeed $A[E]$ is ω_1-bivalent.*

Proof. Letting $\kappa = \omega_1$, let's note that the Coarsening Dichotomy Lemma implies that $\left|\frac{P(\kappa)}{E}\right| < \kappa$. Since $\left|\frac{P(\kappa)}{E}\right| < \kappa$, it follows that for any model $\mathcal{M}$ of $A^2[E]$ (1.2) with cardinality κ, we have that $|\mathrm{rng}(\partial_i)| < \kappa$. By Infinite Sums are Maxs (2.12), we have that $|M \setminus \mathrm{rng}(\partial_1)| = \kappa = |M \setminus \mathrm{rng}(\partial_2)|$. Then by Fine's Theorem 7.1, we have $\mathcal{N}_1$ and $\mathcal{N}_2$ as defined in equation (7.4) are isomorphic. So for any sentence φ of $L_0[\partial]$, we have that $\mathcal{N}_1 \models \varphi$ iff $\mathcal{N}_2 \models \varphi$. □

The situation with respect to bicardinality coarsening and cardinality coarsening on small concepts is slightly more subtle. As for bicardinality coarsening, since it was introduced in connection with surjective relative categoricity, it's natural to modify the definition of κ-bivalence (Definition 8.1) so that attention is restricted to models of $A[E]$ where the abstraction operator is surjective. Now it's not necessarily the case that every bicardinality coarsening equivalence relation will have such a model for a given κ. For instance, the Bicardinality Principle from Section 5 won't have such a model for $\kappa = \omega_1$, as one can easily check by reference to Proposition 6.4. However, for those cardinals κ for which there are such models, Theorem 1.2 trivially implies that bicardinality coarsening suffices for κ-bivalence in the modified sense, since again isomorphism suffices for elementary equivalence.

As for cardinality coarsening on small concepts, it is simply unknown to us whether this implies bivalence compatibility. So we record the following question:

QUESTION 8.5. *Suppose that $E(X, Y)$ is an L_0-formula which is provably an equivalence relation in our background second-order logic and which is cardinality coarsening on small concepts. Is it necessarily the case that $A[E]$ is bivalence compatible?*

If this question is answered in the negative, then it would indicate that the way in which natural relative categoricity captures the idea of determinacy of truth-value is distinct from the way in which Hodes' supervaluationism captures this idea. Of course, these two notions are extensionally distinct since an equivalence relation can be bivalence compatible simply by mimicking Hume's Principle on domains of cardinality ω_1 and appealing to Proposition 8.4, while violating cardinality coarsening on small concepts on domains of other sizes. So the interesting direction is that which is at issue in the above question, since a negative answer would indicate that these two determinacy of truth-value ideas studied in our work and Hodes work are orthogonal to one another.

§9. Conclusions. Our goal has been to articulate various notions of relative categoricity for abstraction principles and to study which abstraction principles are accordingly relative categorical. The import of our Theorems 1.1 and 1.2 is that such relatively categorical abstraction principles qualitatively look like Hume's Principle. Our results contravene the general experience we have with relative categoricity. For, this notion is highly non-domain-specific in that one has relatively categorical axiomatizations of number, set, the reals, etc. But when we restrict attention down to abstraction principles, our Theorems 1.1 and 1.2 show that relative categoricity is tied to cardinality coarsening notions.

Finally, it's worth emphasizing the limited scope of our study. First, we have focused exclusively on abstraction principles formed from equivalence relations on unary concepts, and many natural abstraction principles like that associated to the Burali-Forti paradox

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

concern abstraction principles on binary concepts. Antonelli (Antonelli, 2010a, pp. 10–11) notes that the equivalence relation associated to the Burali-Forti paradox is permutation invariant in the sense of (1.5) once one extends this notion naturally from equivalence relations on unary concepts to binary concepts, and similarly it will be injection invariant. So this indicates that the study of relative categoricity concepts will be quite different when one goes from equivalence relations on unary concepts to those on binary concepts. For instance, this example indicates that in this more general setting we can't have that injection invariance suffices for having standard models of infinite cardinality (cf. Proposition 6.1). So it's hard to predict *apriori* how much of the present study holds when we pass from equivalence relations on unary concepts to those on binary concepts or n-ary concepts.

A similar limitation that should be emphasized is that we have focused on some specific notions of relative categoricity, namely natural relative categoricity and surjective relative categoricity. It would of course be ideal to have analogues of Theorems 1.1 and 1.2 for the notion of relative elementary equivalence and bivalence compatibility (Definition 5.1 and Definition 8.3), but we have been unable to obtain any such characterizations. Further, as indicated in Question 5.2 and Question 8.5, even the relationships between these notions and our notions is not yet resolved. As a final note, it should be underscored that natural relative categoricity concerns a specific bijection (namely, the natural bijection) being an isomorphism. In Corollary 4.1, we noted that this is the only isomorphism when natural relative categoricity does obtain. But it would be important to study the more general notion of relative categoricity which did not restrict attention to the natural bijection. We have been unable to establish analogues of Theorems 1.1 and 1.2 for this more general notion. Our motivation for considering the natural bijection $\Gamma(\partial_1(X)) = \partial_2(X)$ is that this seemed like a natural enough way for agents to exchange information about their abstraction operators, broadly similar in character to Parsons' agents translating their interlocutor's arithmetical vocabulary by their own arithmetical vocabulary.

§10. Acknowledgments. Parts of this material were presented on March 27, 2012 at the workshop on the Mathematics of Abstraction at Birkbeck, University of London; at May 23, 2014 at the workshop Abstraction: Philosophy and Mathematics at the University of Oslo; and on January 7, 2015 at the logic seminar in the department of Logic and Philosophy of Science at the University of California, Irvine. Thanks to the organizers and participants of these events and seminars, as well as to the many others who provided valuable feedback on this work. In particular, thanks to: Aldo Antonelli, Kyle Banick, Roy Cook, Samuel Eklund, Salvatore Florio, J. Ethan Galebach, Jeremy Heis, Graham Leach-Krouse, Greg Lauro, Sarah Lawsky, Øystein Linnebo, Richard Mendelsohn, Christopher Mitsch, Alexander C. R. Oldemeier, Markus Pantsar, Jonathan Payne, Terence Parsons, Agustín Rayo, Sam Roberts, Marian Rogers, J. Schatz, Gabriel Uzquiano, and Kai Wehmeier. Thanks also to the anonymous referees for many helpful comments. Walsh would also like to acknowledge the support of Øystein Linnebo's European Research Council-funded project "Plurals, Predicates, and Paradox" and a Kurt Gödel Research Prize Fellowship.

BIBLIOGRAPHY

Antonelli, G. A. (2010a). The nature and purpose of numbers. *The Journal of Philosophy*, **107**(4), 191–212.

Antonelli, G. A. (2010b). Notions of invariance for abstraction principles. *Philosophia Mathematica*, **18**(3), 276–292.

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Antonelli, G. A. (2010c). Numerical abstraction via the Frege quantifier. *Notre Dame Journal of Formal Logic*, **51**(2), 161–179.

Boolos, G. (1989). Iteration again. *Philosophical Topics*, **17**, 5–21. Reprinted in Boolos (1998).

Boolos, G. (1990). The standard equality of numbers. In Boolos, G., editor. *Meaning and Method: Essays in Honor of Hilary Putnam*, pp. 261–277. Cambridge: Cambridge University Press. Reprinted in Boolos (1998), Demopoulos (1995).

Boolos, G. (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Burgess, J. P. (2005). *Fixing Frege*. Princeton Monographs in Philosophy. Princeton: Princeton University Press.

Button, T., & Walsh, S. (2015). Ideas and results in model theory: Reference, realism, structure and categoricity. arXiv:1501.00472.

Cook, R. T., editor (2007). *The Arché Papers on the Mathematics of Abstraction*. Volume 71 of The Western Ontario Series in Philosophy of Science. Berlin: Springer.

Cook, R. T. (2012). Conservativeness, stability, and abstraction. *British Journal for the Philosophy of Science*, **63**, 673–696.

Demopoulos, W., editor (1995). *Frege's Philosophy of Mathematics*. Cambridge: Harvard University Press.

Ebbinghaus, H.-D. (1985). Extended logics: The general framework. In Barwise, J. and Feferman, S., editors. *Model-Theoretic Logics*, Perspectives in Mathematical Logic, pp. 25–76. New York: Springer.

Enderton, H. B. (2001). *A Mathematical Introduction to Logic* (second edition). Burlington: Harcourt.

Fine, K. (2002). *The Limits of Abstraction*. Oxford: The Clarendon Press.

Fine, K. (2005). Précis. *Philosophical Studies*, **122**(3), 305–313.

Fine, K. (2006). Our knowledge of mathematical objects. In Gendler, T. Z. and Hawthorne, J., editors. *Oxford Studies in Epistemology*, Vol. 1, pp. 89–109. Oxford: Clarendon Press.

Frege, G. (1884). *Die Grundlagen der Arithmetik*. Breslau: Koebner.

Frege, G. (1980). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number* (second edition). Evanston: Northwestern University Press.

Hale, B. (1987). *Abstract Objects*. Oxford: Basil Blackwell.

Hale, B., & Wright, C. (2000). Implicit definition and the a priori. In Boghossian, P. and Peacocke, C., editors. *New Essays on the A Priori*, pp. 286–319. Oxford: Clarendon. Reprinted in Hale & Wright (2001).

Hale, B., & Wright, C. (2001). *The Reason's Proper Study*. Oxford: Oxford University Press.

Heck Jr., R. G. (1992). On the consistency of second-order contextual definitions. *Noûs*, **26**(4), 491–494.

Hodes, H. (1984). Logicism and the ontological commitments of arithmetic. *The Journal of Philosophy*, **81**(3), 123–149.

Hodes, H. (1990). Where do the natural numbers come from? *Synthese*, **84**(3), 347–407.

Hodes, H. (1991). Where do sets come from? *The Journal of Symbolic Logic*, **56**(1), 150–175.

Hrbacek, K., & Jech, T. (1999). *Introduction to Set Theory* (third edition). Volume 220 of Monographs and Textbooks in Pure and Applied Mathematics. New York: Dekker.

Jané, I., & Uzquiano, G. (2004). Well and non-well-founded Fregean extensions. *Journal of Philosophical Logic*, **33**, 437–465.

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Keenan, E. L., & Moss, L. S. (1985). Generalized quantifiers and the expressive power of natural language. In van Benthem, J. and ter Meulen, A., editors. *Generalized Quantifiers in Natural Language*, pp. 73–124. Dordrecht: Floris.

Kunen, K. (1980). *Set Theory*. Volume 102 of Studies in Logic and the Foundations of Mathematics. Amsterdam: North-Holland.

Kunen, K. (2011). *Set Theory*. London: College Publications.

Lavine, S. (1999). Skolem was wrong. Unpublished.

Linnebo, Ø. (2011). Higher-order logic. In Horsten, L. and Pettigrew, R., editors. *The Continuum Companion to Philosophical Logic*, pp. 105–127. London and New York: Continuum.

MacFarlane, J. (2002). Frege, Kant, and the logic in logicism. *The Philosophical Review*, **111**(1), 25–65.

Manzano, M. (1996). *Extensions of First Order Logic*. Volume 19 of Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.

Marker, D. (2002). *Model Theory: An Introduction*. Volume 217 of Graduate Texts in Mathematics. New York: Springer.

McGee, V. (1997). How we learn mathematical language. *Philosophical Review*, **106**(1), 35–68.

Parsons, C. (1990). The uniqueness of the natural numbers. *Iyyun*, **39**(1), 13–44.

Parsons, C. (2008). *Mathematical Thought and its Objects*. Cambridge: Harvard University Press.

Peters, S., & Westerståhl, D. (2008). *Quantifiers in Language and Logic*. Oxford: Oxford University Press.

Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-Order Logic*. Volume 17 of Oxford Logic Guides. New York: The Clarendon Press.

Shapiro, S. (2000). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.

Shapiro, S., editor (2005). *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford: Oxford University Press.

Shapiro, S., & Weir, A. (1999). New V, ZF and abstraction. *Philosophia Mathematica*, **7**(3), 293–321.

Sher, G. (1991). *The Bounds of Logic: A Generalized Viewpoint*. A Bradford Book. Cambridge: MIT Press.

Simpson, S. G. (1999). *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Berlin: Springer.

Tarski, A. (1986). What are logical notions? *History and Philosophy of Logic*, **7**(2), 143–154.

Väänänen, J., & Wang, T. (2014). Internal categoricity in arithmetic and set theory. *Notre Dame Journal of Formal Logic*. To appear.

Walsh, S. (2012). Comparing Hume's Principle, basic law V and Peano arithmetic. *Annals of Pure and Applied Logic*, **163**, 1679–1709.

Walsh, S. (2014a). Fragments of Frege's *Grundgesetze* and Gödel's constructible universe. arXiv:1407.3861.

Walsh, S. (2014b). Logicism, interpretability, and knowledge of arithmetic. *The Review of Symbolic Logic*, **7**(1), 84–119.

Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Volume 2 of Scots Philosophical Monographs. Aberdeen: Aberdeen University Press.

Wright, C. (1997). On the philosophical significance of Frege's theorem. In Heck Jr., R. G., editor. *Language, Thought, and Logic: Essays in Honour of Michael Dummett*, pp. 201–244. Oxford: Oxford University Press. Reprinted in Hale & Wright (2001).

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press

Wright, C. (1998). On the harmless impredictavity of $N^{=}$ (Hume's Principle). In Schirn, M., editor. *Philosophy of Mathematics Today*, pp. 393–368. Oxford: Clarendon Press. Reprinted in Hale & Wright (2001).

Wright, C. (1999). Is Hume's Principle analytic? *Notre Dame Journal of Formal Logic*, **40**(1), 6–30. Reprinted in Hale & Wright (2001).

DEPARTMENT OF LOGIC AND PHILOSOPHY OF SCIENCE
5100 SOCIAL SCIENCE PLAZA
UNIVERSITY OF CALIFORNIA, IRVINE
IRVINE, CA 92697-5100, U.S.A.
E-mail: swalsh108@gmail.com or walsh108@uci.edu

DEPARTMENT OF PHILOSOPHY
NORTHWESTERN UNIVERSITY
1860 CAMPUS DRIVE, EVANSTON
ILLINOIS 60208-2214, U.S.A.
E-mail: s-ebelsduggan@northwestern.edu

https://doi.org/10.1017/S1755020315000052 Published online by Cambridge University Press