# The Mathematics of Abstraction:
# Preliminary Results

**Roy T Cook**
**University of Minnesota**
**cookx432@umn.edu**

Note: All results are unpublished and due to Roy T Cook, unless otherwise noted.

## 1: Technical Preliminaries

An abstraction principle is any principle of the form:

$A_E$: $(\forall\alpha)(\forall\beta)[@_E(\alpha) = @_E(\beta) \leftrightarrow E(\alpha, \beta)]$

where $\alpha$ and $\beta$ are variables (or sequences of variables) ranging over entities of same 'type' (or sequences whose elements are, pairwise, of the same type), E is an equivalence relation on entities of that type (or, if $\alpha$ and $\beta$ are sequences, then E is an equivalence relation on sequences of the relevant sort), and $@_E$ a term-forming operator – an abstraction operator – mapping entities of the relevant type (or sequences of entities) onto objects. The paradigm instances of abstraction principles are *Hume's Principle*:

HP: $(\forall X)(\forall Y)[\#(X) = \#(Y) \leftrightarrow X \approx Y]$

(i.e. HP = $A_{X \approx Y}$, where "$X \approx Y$" abbreviates the second-order formula expressing that there is a one-one-onto function from the X's to the Y's).

and *Basic Law* V:

BLV: $(\forall X)(\forall Y)[\S(X) = \S(Y) \leftrightarrow (\forall z)(X(z) \leftrightarrow Y(z))]$

(i.e. BLV = $A_{(\forall z)(X(z) \leftrightarrow Y(z))}$).

In addition, Finite Hume's Principle:

FHP: $(\forall X)(\forall Y)[\#(X) = \#(Y) \leftrightarrow (X \approx Y \vee (Inf(X) \wedge Inf(Y))]$

(where "Inf(x)" abbreviates the claim that X has infinitely many instances) shall be useful in what follows. We shall take a cue from these central cases and restrict our attention to second-level abstraction principles of the form:

$A_E$: $(\forall X)(\forall Y)[@_E(X) = @_E(Y) \leftrightarrow E(X, Y)]$

where variables X and Y range over first-level concepts, and where E(Y, Y) is a formula in pure third-order logic (Of course, the cases of most interest are those where E is an equivalence relation on unary first-level concepts). The reader should note that our approach individuates abstraction principles purely syntactically – hence *Hume's Principle*$_2$:

$HP_2$: $(\forall X)(\forall Y)[\#(X) = \#(Y) \leftrightarrow Y \approx X]$

is, strictly speaking, a distinct principle from HP, since HP = $A_{X\approx Y}$ and $HP_2$ = $A_{Y\approx X}$ (further, neither HP $\rightarrow$ $HP_2$ nor $HP_2$ $\rightarrow$ HP is a logical truth, since HP and $HP_2$ involve distinct abstraction operators $@_{X\approx Y}$ and $@_{Y\approx X}$). We shall address subtle issues regarding the equivalence of abstraction principles below.

We shall call this language *L*. At times we shall wish to restrict our attention to formulas that do not contain one or more abstraction operators. Thus:

DEFINITION 1.1: $L\backslash @_E$ is the language obtained by removing from *L* all formulas containing $@_E$. Similarly, if S is a set of abstraction operators, then $L\backslash S$ is the language obtained by removing from *L* all formulas containing any abstraction operator $@_E \in S$.

Hence, if A is the set of all abstraction operators, then $L\backslash A$ is pure third-order logic. The following notations will be useful:

DEFINITION 1.2: Given any formula $\Phi$, $R(\Phi)$ is the *ramsification* of $\Phi$.

DEFINITION 1.3: Given any formula $\Phi$ and unary predicate $\Psi$ (where $\Psi$ might be a second-order variable), $\Phi^{\Psi}$ is the *relativization* of (the quantifiers of) $\Phi$ to $\Psi$.

The following terminology emphasizes that the satisfiability of abstraction principles depends solely on the cardinality of the domain in question:

DEFINITION 1.4: Given any cardinal $\kappa$, an abstraction principle $A_E$ is $\kappa$-*satisfiable* if and only if $A_E$ is satisfiable in a (and hence any) domain of cardinality $\kappa$.

Additionally, the following constructions will be used repeatedly:

DEFINITION 1.5: Given an abstraction principle $A_E$ and any formula $\Phi$, $A_E\nabla\Phi = A_{(R(\Phi)\vee E(X,Y))} = (\forall X)(\forall Y)[@_{(\Phi\vee E(X,Y))}(X) = @_{(\Phi\vee E(X,Y))}(Y) \leftrightarrow (R(\Phi) \vee E(X, Y))]$.

DEFINITION 1.6: Given an abstraction principle $A_E$ and any formula $\Phi$, $A_E\Delta\Phi = A_{(R(\Phi)\wedge E(X,Y))} = (\forall X)(\forall Y)[@_{(\Phi\wedge E(X,Y))}(X) = @_{(\Phi\wedge E(X,Y))}(Y) \leftrightarrow (R(\Phi) \wedge E(X, Y))]$.

DEFINITION 1.7: Given an abstraction principle $A_E$, $\eta A_E = A_{(\sim R(AE)\vee(\forall z)(X(z)\leftrightarrow Y(z)))}$.

Note that $\eta A_E \neq \sim A_E$, since the former, but not the latter, is an abstraction principle. We do obtain the following:

THEOREM 1.8: For any abstraction principle $A_E$ and any cardinal number $\kappa$, the following are equivalent:

- $\eta A_E$ is $\kappa$-satisfiable.
- $\sim A_E$ is $\kappa$-satisfiable.
- $\sim R(A_E)$ is true on models of cardinality $\kappa$.
- $R(\eta A_E)$ is true on models of cardinality $\kappa$.
- $\sim R(A_E)$ is logically equivalent to $R(\eta A_E)$.

PROOF: Straightforward, left to the reader.

The following abstraction principle, which we will call the *Trivial Abstraction Principle*, is useful:

$$\text{Triv:} \quad (\forall X)(\forall Y)[\P(X) = \P(Y) \leftrightarrow (\forall z)((X(z) \leftrightarrow X(z)) \wedge (Y(z) \leftrightarrow Y(z)))]$$

Triv is $\kappa$-satisfiable for any $\kappa > 0$. We note the following facts regarding $\nabla$, $\Delta$, and $\eta$:

THEOREM 1.9: For any abstraction principle $A_E$, formula $\Phi$, and cardinal $\kappa$, $A_E\nabla\Phi$ is $\kappa$-satisfiable if and only if either $A_E$ is $\kappa$-satisfiable or $R(\Phi)$ is true on models of cardinality $\kappa$.

PROOF: Straightforward, left to the reader.

COROLLARY 1.10: For any formula $\Phi$ and cardinal $\kappa$, $BLV\nabla\Phi$ is $\kappa$-satisfiable if and only if $R(\Phi)$ is true on models of cardinality $\kappa$.

THEOREM 1.11: For any abstraction principle $A_E$, formula $\Phi$, and cardinal $\kappa$, $A_E\Delta\Phi$ is $\kappa$-satisfiable if and only if $A_E$ is $\kappa$-satisfiable and $R(\Phi)$ is true on models of cardinality $\kappa$.

PROOF: Straightforward, left to the reader.

COROLLARY 1.12: For any formula $\Phi$ and cardinal $\kappa$, $Triv\Delta\Phi$ is $\kappa$-satisfiable if and only if $R(\Phi)$ is true on models of cardinality $\kappa$.

THEOREM 1.13: For any abstraction principle $A_E$ and cardinal $\kappa$, $\eta A_E$ is $\kappa$-satisfiable if and only if $A_E$ is not $\kappa$-satisfiable.

PROOF: Straightforward, left to the reader.

Given these three operations, the set of abstraction principles can be viewed as a Boolean algebra – a fact that shall be of interest in later results. Our final definitions in this section concern various senses in which one abstraction principle can imply another, or two abstraction principles can be equivalent. We shall reserve the symbols "$\rightarrow$" and "$\leftrightarrow$" for the standard (classical) material conditional, recalling that:

THEOREM 1.14: If $A_{E1} \rightarrow A_{E2}$ is a logical truth, then either $E_1$ is typographically identical to $E_2$, or $A_{E1}$ is inconsistent.

PROOF: If $E_1$ is typographically distinct from $E_2$, then $@_{E1}$ and $@_{E2}$ are distinct function terms.

COROLLARY 1.15: If $A_{E1} \leftrightarrow A_{E2}$ is a logical truth, then either $E_1$ is typographically identical to $E_2$, or $A_{E1}$ and $A_{E2}$ are inconsistent.

COROLLARY 1.16: If $A_{E1} \rightarrow A_{E2}$ is a logical truth and $A_{E1}$ is consistent, then $A_{E1} \leftrightarrow A_{E2}$ is a logical truth.

Given the failure of straightforward logical entailment and equivalence to provide useful relations between abstraction principles, we now introduce three additional, and provably distinct, notions of equivalence. The first and weakest of these judges equivalence solely in terms of the cardinalities $\kappa$ such that the principles in question are $\kappa$-satisfiable:

DEFINITION 1.17: Abstraction principle $A_{E1}$ *cardinality-entails* abstraction principle $A_{E2}$ (i.e. $A_{E1} \supset_C A_{E2}$) if and only if, for any cardinal $\kappa$, if $A_{E1}$ is $\kappa$-satisfiable, then $A_{E2}$ is $\kappa$-satisfiable.

DEFINITION 1.18: A set of abstraction principles is *closed under cardinality-entailment* if and only if, for any $A_{E1}$, $A_{E2}$, if $A_{E1} \in S$ and $A_{E1} \supset_C A_{E2}$, then $A_{E2} \in S$.

THEOREM 1.19: For any abstraction principles $A_{E1}$ and $A_{E2}$, the following are equivalent:

- $A_{E1} \supset_C A_{E2}$
- $A_{E1} \rightarrow R(A_{E2})$ is a logical truth.
- $R(A_{E1}) \rightarrow R(A_{E2})$ is a logical truth.

PROOF: Straightforward, left to the reader.

DEFINITION 1.20: $A_{E1} \equiv_C A_{E2}$ if and only if, for any cardinal $\kappa$, $A_{E1}$ is $\kappa$-satisfiable if and only if $A_{E2}$ is $\kappa$-satisfiable.

DEFINITION 1.21: A set of abstraction principles S is *weakly closed under cardinality-equivalence* if and only if, for any $A_{E1}$, $A_{E2}$, if $A_{E1} \in S$ and $A_{E1} \equiv_C A_{E2}$, then $A_{E2} \in S$.

THEOREM 1.22: $A_{E1} \equiv_C A_{E2}$ if and only if $A_{E1} \supset_C A_{E2}$, and $A_{E2} \supset_C A_{E1}$.

PROOF: Straightforward, left to the reader.

COROLLARY 1.23: $A_{E1} \equiv_C A_{E2}$ if and only if $R(A_{E1}) \leftrightarrow R(A_{E2})$ is a logical truth.

We now make the following observations:

THEOREM 1.24: For any formula $\Phi$, the following are all true:

- $BLV\nabla\Phi \equiv_C Triv\Delta\Phi$
- $BLV\Delta\Phi \equiv_C BLV$
- $Triv\nabla\Phi \equiv_C Triv$

PROOF: Straightforward, left to the reader.

We can require more of 'equivalent' abstraction principles than that they be merely cardinality-equivalent, however:

DEFINITION 1.25: Two abstraction principles $A_{E1}$ and $A_{E2}$ are *weakly abstraction-equivalent* (i.e. $A_{E1} \equiv_{WA} A_{E2}$) if and only if $A_{E1} \leftrightarrow A_{E2}[@_{E2}/@_{E1}]$[1] is a logical truth.

DEFINITION 1.26: A set of abstraction principles is *closed under weak-abstraction-equivalence* if and only if, for any $A_{E1}$, $A_{E2}$, if $A_{E1} \in S$ and $A_{E1} \equiv_{WA} A_{E2}$, then $A_{E2} \in S$.

DEFINITION 1.27: Two abstraction principles $A_{E1}$ and $A_{E2}$ are *strongly abstraction-equivalent* (i.e. $A_{E1} \equiv_{SA} A_{E2}$) if and only if $(\forall X)(\forall Y)(E_1(X, Y) \leftrightarrow E_2(X, Y))$ is a logical truth.

DEFINITION 1.28: A set of abstraction principles is *closed under strong-abstraction-equivalence* if and only if, for any $A_{E1}$, $A_{E2}$, if $A_{E1} \in S$ and $A_{E1} \equiv_{SA} A_{E2}$, then $A_{E2} \in S$.

Loosely speaking, two abstraction principles are strongly-abstraction-equivalent if and only if their respective equivalence relations divide concepts into identical equivalence classes on <u>any</u> model, and two abstraction principles are weakly-abstraction-equivalent if and only if they are

[1] Given a formula $\Phi$ and terms $t_1$ and $t_2$, $\Phi[t_1, t_2]$ is the result of replacing all occurrence of $t_1$ in $\Phi$ with $t_2$.

cardinality-equivalent and their respective equivalence relations divide the concepts into identical equivalence classes on any model in which both principles are satisfiable (although the equivalence relations in question need not so 'agree' on domains where the abstraction principles are not satisfiable) The following clarify the relationship between these notions and our previous definition of cardinality equivalence:

THEOREM 1.29: The following entailments hold:

- If $A_{E1} \leftrightarrow A_{E2}$ is a logical truth then $A_{E1} \equiv_{SA} A_{E2}$.
- If $A_{E1} \equiv_{SA} A_{E2}$ then $A_{E1} \equiv_{WA} A_{E2}$.
- If $A_{E1} \equiv_{WA} A_{E2}$ then $A_{E1} \equiv_{C} A_{E2}$.

PROOF: Straightforward, left to the reader.

THEOREM 1.30: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1} \equiv_{C} A_{E2}$ but not $A_{E1} \equiv_{WA} A_{E2}$.

PROOF: HP and FHP are cardinality equivalent (since both are κ-satisfiable for all and only infinite κ), but not weakly abstraction-equivalent (since, for any model whose domain is of cardinality κ for $\kappa > \aleph_0$, the respective equivalence relations are differ on uncountable concepts).

THEOREM 1.31: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1} \equiv_{WA} A_{E2}$ but not $A_{E1} \equiv_{SA} A_{E2}$.

PROOF: Let Θ be the second-order formula true only on infinite domains. Consider:

$$A_E: \quad (\forall X)(\forall Y)[@_E(X) = @_E(Y) \leftrightarrow ((X \approx Y \wedge \Theta) \vee (\forall z)(X(z) \leftrightarrow Y(z)))]$$

Then HP $\equiv_{WA} A_E$ but not HP $\equiv_{SA} A_E$

THEOREM 1.32: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1} \equiv_{SA} A_{E2}$ but $A_{E1} \leftrightarrow A_{E2}$ is not a logical truth.

PROOF: This merely reiterates the observation made regarding HP and $HP_2$ earlier.

## 2: The Algebra of Abstraction

The notions developed in the previous section allow us to construct a Boolean algebra whose elements are sets of 'equivalent' abstraction principles. The basic idea is that we construct the meet (∩), join (∩), and complement (¬) operations in terms of our previously defined operations Δ, ∇ and η (additionally, the top (1) and bottom (0) elements will be the sets corresponding to Triv and BLV respectively). The issue is complicated by the following result, which demonstrates that we need to take care in choosing which notion of equivalence to mobilize in the present context:

THEOREM 2.1: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1}\Delta A_{E2}$ is not weakly-abstraction equivalent to $A_{E2}\Delta A_{E1}$.

PROOF: It is not the case that HPΔFHP $\equiv_{WA}$ FHPΔHP.

THEOREM 2.2: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1}\nabla A_{E2}$ is not weakly abstraction equivalent to $A_{E2}\nabla A_{E1}$.

PROOF: Let $\Gamma$ be the second-order formula expressing that the universe is finite. Then it is not the case that $HP\nabla(BLV\nabla\Gamma) \equiv_{WA} (BLV\nabla\Gamma)\nabla HP$.

COROLLARY 2.3: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1}\Delta A_{E2}$ is not strongly-abstraction-equivalent to $A_{E2}\Delta A_{E1}$.

COROLLARY 2.4: There are abstraction principles $A_{E1}$ and $A_{E2}$ such that $A_{E1}\nabla A_{E2}$ is not strongly-abstraction-equivalent to $A_{E2}\nabla A_{E1}$.

The relevant sort of commutativity does hold if we restrict our attention to the weaker notion of cardinality equivalence, however:

LEMMA 2.5: Let A be the class of cardinals $\kappa$ such that $A_{E1}$ is $\kappa$-satisfiable, and B be the class of cardinals such that $A_{E2}$ is $\kappa$-satisfiable, and C be the class of cardinals such that $A_{E1}\nabla A_{E2}$ is $\kappa$-satisfiable. Then $C = A \cup B$.

PROOF: Immediate consequence of THEOREM 1.9.

LEMMA 2.6: Let A be the class of cardinals $\kappa$ such that $A_{E1}$ is $\kappa$-satisfiable, and B be the class of cardinals such that $A_{E2}$ is $\kappa$-satisfiable, and C be the class of cardinals such that $A_{E1}\Delta A_{E2}$ is $\kappa$-satisfiable. Then $C = A \cap B$.

PROOF: Immediate consequence of THEOREM 1.9.

COROLLARY 2.7: For any abstraction principles $A_{E1}$ and $A_{E2}$, $A_{E1}\nabla A_{E2} \equiv_C A_{E2}\nabla A_{E1}$ and $A_{E1}\nabla A_{E2} \equiv_C A_{E2}\nabla A_{E1}$.

Given COROLLARY 2.7, we can construct a Boolean algebra on sets of cardinality-equivalent abstraction principles as follows:

| | | |
|---|---|---|
| DEFINITION 2.8[2]: | $[A_{E1}]_C$ | $= \{A_{E2} : A_{E2} \equiv_C A_{E1}\}$ |
| | $ELEM_C$ | $= \{[A_E]_C : A_E$ is an abstraction principle$\}$ |
| | $[A_{E1}]_C \cap_C [A_{E2}]_C$ | $= [A_{E1}\Delta A_{E2}]_C$ |
| | $[A_{E1}]_C \cup_C [A_{E2}]_C$ | $= [A_{E1}\nabla A_{E2}]_C$ |
| | $\neg_C[A_E]$ | $= [\eta A_E]_C$ |
| | $1_C$ | $= [Triv]_C$ |
| | $0_C$ | $= [BLV]_C$ |

THEOREM 2.9: $\langle ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C\rangle$ is a Boolean algebra.

PROOF: Straightforward, left to the reader.

The following construction provides a clearer characterization of the structure of the Boolean algebra $\langle ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C\rangle$:

[2] This construction is little more than a variation on the standard Tarksi-Lindenbaum algebra construction.

DEFINITION 2.10: A class of cardinals C is 3OL-definable if and only if there is a purely logical, third-order formula $\Phi$ such that, for any cardinal $\kappa$, $\Phi$ is true on models of size $\kappa$ if and only if $\kappa \in C$.

LEMMA 2.11: For any class of cardinals C, C is 3OL-definable if and only if there is an abstraction principle $A_E$ in L such that, for any cardinal $\kappa$, $A_E$ is $\kappa$-satisfiable if and only if $\kappa \in C$.

PROOF: ($\rightarrow$) If C is 3OL-definable by $\Phi$, then let $A_E = BLV\nabla\Phi$.

($\leftarrow$) Given $A_E$ and C such that $A_E$ is $\kappa$-satisfiable if and only if $\kappa \in C$, then $R(A_E)$ 3-OL-defines C.

DEFINITION 2.12: Let $<X, \cap_X, \cup_X, \neg_X, 1_X, 0_X>$ where $X = \{C : C$ is 3OL-definable$\}$, $A \cap_X B = A \cap B$, $A \cup_X B = A \cup B$ (i.e. the standard class-theoretic intersection and union of A and B), $\neg_X A = \{\kappa : \kappa \notin A\}$, $1_X$ = Card (i.e. the class of all cardinals), and $0_X = \emptyset$.

THEOREM 2.13: $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$ is isomorphic to $<X, \cap_X, \cup_X, \neg_X, 1_X, 0_X>$.

PROOF: The isomorphism f: $X \rightarrow ELEM_C$ is defined as follows: Given $A \in X$ and formula $\Phi$ such that $\Phi$ 3OL defines A, $f(A) = [BLV\nabla\Phi]_C$.

DEFINITION 2.14: A class of cardinals C is weakly-3OL-definable if and only if there is a set purely logical, third-order formulas S such that, for any cardinal $\kappa$, all members of S are true on models of size $\kappa$ if and only if $\kappa \in C$.

THEOREM 2.15: If a class of cardinals C is 3OL-definable, then C is weakly-3OL-definable.

PROOF: Straightforward, left to the reader.

LEMMA 2.16: For any class of cardinals C, C is 3OL-definable if and only if there is a set of abstraction principles S in L where, for any cardinal $\kappa$, $\kappa \in C$ if and only if, for all $A_E \in S$, $A_E$ is $\kappa$-satisfiable.

PROOF: Similar to proof of LEMMA 2.11.

DEFINITION 2.17: An abstraction principle $A_E$ is *stable* iff there is a cardinal $\gamma$ such that, for all cardinals $\kappa \geq \gamma$ $A_E$ is $\kappa$-satisfiable. $A_E$ is *unstable* otherwise.

DEFINITION 2.18: Given set of sentences S, the *stability point* of S, $SP(S) = \inf(\{\kappa$ : for all $\gamma \geq \kappa$, S has a model of size $\gamma\})$ ($SP(S) = 0$ if $\{\kappa$ : for all $\gamma > \kappa$, S has a model of size $\gamma\} = \emptyset$).

LEMMA 2.19: For any abstraction principle $A_{E1}$, there is an abstraction principle $A_{E2}$ such that $SP(\{A_{E2}\}) = SP(\{A_{E1}\}) + 1$.

PROOF: Straightforward, left to the reader.

THEOREM 2.20: There exists a class of cardinals C such that C is weakly-3OL definable, but not 3OL-definable.

PROOF: Let S be the set of stable abstraction principles, and let C be the class of cardinals such that $\kappa \in C$ if and only if, for all $A_{E1} \in S$, $A_{E1}$ is $\kappa$-satisfiable. Note that $SP(S) > 0$ and, for all $A_{E1} \in S$, $A_{E1}$ is SP(S)-satisfiable. Assume (for *reductio*) that $\Phi$ 3OL-defines C. Then, for any $\kappa$, BLV∇Φ is κ-satisfiable if and only if $\kappa \in C$, and hence SP({BLV∇Φ}) = SP(S). Let $A_{E2}$ be an abstraction principle such that SP($A_{E2}$) = SP(BLV∇Φ) + 1 (by LEMMA 2.19). $A_{E2} \in S$, and $A_{E2}$ is not SP(S)-satisfiable. Contradiction.

COROLLARY 2.21: $<X, \cap_X, \cup_X, \neg_X, 1_X, 0_X>$ (and hence $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$) is not closed under least upper bounds.

PROOF: Combine THEOREMS 2.13 and 2.20.

## 3. Some Particularly Interesting Classes of Abstraction Principles.

Particular classes of abstraction principles have been singled out for special attention in the literature – in particular, in the literature on the Bad Company Objection. Interestingly, even though these classes were originally identified based on philosophical motivations, a number of these correspond to familiar algebraic constructions in $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$. First, we shall list definitions (a number of these were originally formulated in Wright [1997] & Weir [2003]):

DEFINITION 3.1: An abstraction principle $A_E$ is *satisfiable* if and only if there is a cardinal κ such that $A_E$ is κ-satisfiable. $A_E$ is *unsatisfiable* otherwise. SAT = the set of satisfiable abstraction principles.

DEFINITION 3.2: An abstraction principle $A_E$ is *Field conservative* iff, for any theory T and formula Φ in $L \backslash @_E$, if $T^{\sim(\exists Y)(x = @(Y))} \cup \{A_E\} \Rightarrow \Phi^{\sim(\exists Y)(x = @(Y))}$ then $T \Rightarrow \Phi$. $A_E$ is *Field-non-conservative* otherwise. F-CON = the set of Field-conservative principles.

DEFINITION 3.3: Given a theory T, ⇃(T) = inf({γ : T has a model of cardinality γ}) (⇃(T) = 0 if T is unsatisfiable).

DEFINITION 3.4: An abstraction principle $A_E$ is *pseudo-conservative* iff, for any theory T in $L \backslash @_E$, there is a cardinal κ ≥ ⇃(T) such that $A_E$ is κ-satisfiable. $A_E$ is *pseudo-non-conservative* otherwise. P-CON = the set of pseudo-conservative abstraction principles.

DEFINITION 3.5: An abstraction principle $A_E$ is *unbounded* iff, for any cardinal γ, there is a cardinal κ ≥ γ such that $A_E$ is κ-satisfiable. $A_E$ is *bounded* otherwise. UNB = the set of unbounded principles.

DEFINITION 3.6: An abstraction principle $A_E$ is *stable* iff there is a cardinal γ such that, for all cardinals κ ≥ γ $A_E$ is κ-satisfiable. $A_E$ is *unstable* otherwise. STB = the set of stable principles.

DEFINITION 3.7: An abstraction principle $A_{E1}$ is *copacetic* iff, for any unbounded abstraction principle $A_{E2}$, there is a model that satisfies both $A_{E1}$ and $A_{E2}$. $A_E$ is *uncopacetic* otherwise. COP = the class of copacetic abstraction principles.

DEFINITION 3.8: $A_{E1}$ is *irenic* iff, for any Field-conservative abstraction principle $A_{E2}$, there is a model that satisfies both $A_{E1}$ and $A_{E2}$. $A_{E1}$ is *unirenic* otherwise. IRN = the set of irenic abstraction principles.

DEFINITION 3.9: $A_E$ is *strongly stable* iff there is some κ such that $A_E$ is satisfiable on all and only models whose domains are of cardinality ≥ κ. $A_E$ is *strongly unstable* otherwise. S-STB = the set of strongly stable principles.

DEFINITION 3.10: An abstraction principle $A_E$ is *strongly conservative* if and only if, for any theory T and formula $\Phi$ in $L \backslash @_E$, if Th $\cup$ $\{A_E\}$ $\Rightarrow$ $\Phi$ then Th $\Rightarrow$ $\Phi$. $A_E$ is *strongly non-conservative* otherwise. S-CON = the set of strongly conservative abstraction principles.

DEFINITION 3.11: An abstraction principle $A_{E1}$ is *agreeable* iff, for any consistent abstraction principle $A_{E2}$, there is a model that satisfies both $A_{E1}$ and $A_{E2}$. AGR = the class of agreeable abstraction principles.

THOEREM 3.12: S-CON = AGR $\subset$ S-STB $\subset$ IRN $\subseteq$ STB = COP $\subset$ UNB $\subseteq$ F-CON = P-CON $\subset$ SAT.

Proof: Left to the reader (or see Weir [2004], Cook [unpublished], Linnebo [forthcoming]).

Unfortunately, whether or not IRN = STB (and whether UNB = F-CON) remains an open question at the time of writing. (although it is known that IRN = STB if and only if UNB = F-CON – see Linnebo [forthcoming]).

THEOREM 3.13: SAT, F-CON, UNB, STB, IRN, and S-CON are each closed under cardinality-entailment (and hence closed under cardinality-equivalence, weak abstraction equivalence, and strong abstraction equivalence).

PROOF: Straightforward, left to the reader.

We can now characterize a number of these classes in terms of familiar algebraic constructions.

DEFINITION 3.14: Let $UNSAT_C = \{[A_E]_C : A_E \notin SAT\}$.

THEOREM 3.15: $UNSAT_C$ is an ideal of $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$.

PROOF: Let $A_{E1}$, $A_{E2} \notin$ SAT, and $A_{E3}$ be any other abstraction principle. Then $A_{E1} \nabla A_{E2} \notin$ STB, and $A_{E1} \Delta A_{E3} \notin$ STB.

In this case, $UNSAT_C = \{[BLV]_C\}$, thus $UNSAT_C$ is the trivial, single-element ideal of $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$ consisting solely of the bottom element. We shall see, however, that other interesting classes of abstraction principles correspond to non-trivial ideals.

DEFINITION 3.16: Let $UNFCON_C = \{[A_E]_C : A_E \notin \text{F-CON}\}$.

THEOREM 3.17: $UNFCON_C$ is an ideal of $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$.

PROOF: Let $A_{E1}$, $A_{E2} \notin$ F-CON, and $A_{E3}$ be any other abstraction principle. Then $A_{E1} \nabla A_{E2} \notin$ F-CON, and $A_{E1} \Delta A_{E3} \notin$ F-CON.

Unlike $UNSAT_C$, UNF-CON is a non-trivial ideal of $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$.

DEFINITION 3.18: Let $UNUNB_C = \{[A_E]_C : A_E \notin UNB\}$.

THEOREM 3.19: $UNUNB_C$ is an ideal of $<ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C>$.

PROOF: Let $A_{E1}$, $A_{E2} \notin$ UNB, and $A_{E3}$ be any other abstraction principle. Then $A_{E1} \nabla A_{E2} \notin$ UNB, and $A_{E1} \Delta A_{E3} \notin$ UNB.

Of course, we do not know whether this ideal is identical, or distinct, to the one in THEOREM 3.17.

DEFINITION 3.20: Let $STB_C = \{[A_E]_C : A_E \in STB\}$.

THEOREM 3.21: $STB_C$ is a filter of <$ELEM_C$, $\cap_C$, $\cup_C$, $\neg_C$, $1_C$, $0_C$>.

PROOF: Let $A_{E1}$, $A_{E2} \in$ STB, and $A_{E3}$ be any other abstraction principle. Then $A_{E1} \Delta A_{E2} \in$ STB, and $A_{E1} \nabla A_{E3} \in$ STB.

$STB_C$ is a non-trivial filter. The next notion gives us a trivial filter to complement the trivial ideal $UNSAT_C$.

DEFINITION 3.22: Let $SCON_C = \{[A_E]_C : A_E \in$ S-CON$\}$.

THEOREM: $SCON_C$ is a filter of <$ELEM_C$, $\cap_C$, $\cup_C$, $\neg_C$, $1_C$, $0_C$>.

PROOF: $SCON_C$ is the trivial one-element filter $\{[Triv]_C\}$.

## 4: Abstraction and Cosatisfaction Sets

Amongst the sets of abstraction principles listed in THEOREM 3.12, there were a number of pairs A, B such that:

A = $\{A_{E1}$ : For any $A_{E2} \in$ B, there is a $\kappa$ such that $A_{E1}$, $A_{E2}$ are both $\kappa$-satisfiable$\}$

We now rigorously define this notion, which we shall call co-satisfaction:

DEFINITION 4.1: Given a set of abstraction principles S:

COS(S) = $\{A_{E1}$ : for any $A_{E2} \in$ S, there is a $\kappa$ such that $A_{E1}$ & $A_{E2}$ are both $\kappa$-satisfiable$\}$

As we noted, a number of the sets studied above are defined along these lines, hence:

| | | | |
|---|---|---|---|
| AGR | ( = S-CON) | = | COS(SAT) |
| IRN | | = | COS(F-CON) |
| COP | ( = STB) | = | COS(UNB) |

One natural, but ultimately false, conjecture would be that, for any sets of abstraction principles A and B, if A = COS(B), then B = COS(A). Easy counterexamples can be found by choosing a B not closed under cardinality-equivalence (e.g. let HP $\in$ B but FHP $\notin$ B). Even requiring that B be closed under cardinality-equivalence, however, the conjecture still fails, as he following theorem demonstrates:

THEOREM 4.2: There exist satisfiable sets of abstraction principles A and B such that A and B are closed under cardinality-equivalence, A = COS(B), but B $\neq$ COS(A).

PROOF: Immediate consequence of the fact that, for any satisfiable A, Triv $\in$ COS(A) Hence, if Triv $\notin$ B, then B $\neq$ COS(A).

If we strengthen the antecedent condition to closure under cardinality-entailment, however, then we obtain the desired result:

THEOREM 4.3: Given any sets of abstraction principles A and B such that B is closed under cardinality-entailment, if A = COS(B), then B = COS(A).

PROOF: ($\rightarrow$) Assume $A_{E1} \in$ B. Then every $A_{E2} \in$ A is cosatisfiable with $A_{E1}$. So $A_{E1} \in$ COS(A).

(←) Assume $A_{E1} \notin B$. Then no member of B entails $A_{E1}$. So $\eta A_{E1}$ is cosatisfiable with every $A_{E2} \in B$. So $\eta A_{E1} \in COS(B)$. Hence $\eta A_{E1} \in A$. Thus, $A_{E1} \notin COS(A)$.

This provides the following chart:

| C | COS(C) |
|---|---|
| SAT | S-CON = AGR |
| XX | S-STB |
| F-CON | IRN |
| UNB | W-STB = COP |
| W-STB = COP | UNB |
| IRN | F-CON |
| S-STB | ??? |
| S-CON = AGR | SAT |

Note that if A = COS(B), then A is trivially closed under cardinality-entailment. Hence, S-STB, since not closed under cardinality entailment, is not the co-satisfiability class of any class of principles (hence the "XX"). Of course, there is some class of principles that is the co-satisfiability class of S-STB, but it remains unidentified at present (hence the "???").

The chart above suggests the following question: Are there sets of principles that are their own co-satisfiability class? More formally:

DEFINITION 4.4: A set of abstraction principles S is *fixed* if and only if S = COS(S).

OPEN QUESTION 4.5: Is there a fixed set of abstraction principles?

OPEN QUESTION 4.6: If there is a fixed set of abstraction principles, is it unique?

Although these questions are open, the following result places some constraints on the nature of any fixed classes of abstraction principles, if they in fact exist:

THEOREM 4.7: If A, B, C, and D are sets of abstraction principles closed under cardinality-entailment and where A = COS(B), C = COS(D) and $A \subseteq C$, then $D \subseteq B$.

PROOF: Straightforward, left to the reader.

We conclude this brief survey of results with the following conjecture:

CONJECTURE 4.8: If A = COS(A) then $\{[A_E]_C : A_E \in A\}$ is an ultrafilter in $\langle ELEM_C, \cap_C, \cup_C, \neg_C, 1_C, 0_C \rangle$.

**References:**

Heck, R. (ed.) [1997], *Language, Thought and Logic: Essays in Honour of Michael Dummett*, Oxford, Clarendon Press.

Linnebo, O. [forthcoming], "Come Criteria for Acceptable Abstraction", *Notre Dame Journal of Formal Logic* **: ** - **.

Weir, A. [2003], "Neo-Fregeanism: An Embarassment of Riches", *Notre Dame Journal of Formal Logic* 44: 13 - 48.

Wright, C [1997], "On the Philosophical Significance of Frege's Theorem", in Heck [1997]: 201 – 244.