CLASS THEORY IN HYPE

MARIA BEATRICE BUONAGUIDI AND CARLO NICOLAI

Abstract. The paper studies class theory over the logic **HYPE** recently introduced by Hannes Leitgeb. We formulate suitable abstraction principles and show their consistency by displaying a class of fixed-point (term) models. By adapting a classical result by Brady, we show their inconsistency with standard extensionality principles, as well as the incompatibility of our semantics with weak extensionality principles introduced in the literature. We then formulate our version of weak extensionality (appropriate to the behaviour of the conditional in **HYPE**) and show its consistency with one of the abstraction principles previously introduced. We conclude with observations and examples supporting the claim that, although arithmetical axioms over **HYPE** are as strong as classical arithmetical axioms, the behaviour of classes over **HYPE** is akin to the one displayed by classes in other nonclassical class theories.

1. Introduction

The logic **HYPE** has been recently put forward by Hannes Leitgeb as a framework to study several phenomena that appear to be incompatible with classical logic [13]. **HYPE** has close relationships with well-known logical systems: Odintsov and Wansing [17] recently showed that **HYPE** is equivalent to the logic N_i^* (Heyting-De Morgan logic or modal symmetric propositional calculus), introduced by Moisil [14] and later explored by Monteiro [15]. **HYPE** can be seen, roughly (but see §2 for a precise definition), as the result of extending First-Degree Entailment (**FDE**) with an intuitionistic conditional $\rightarrow$.[1]

One application studied by Leitgeb is to semantic paradoxes. Leitgeb provides fixed-point models based on **HYPE** which satisfy the T-schema $\mathrm{T}\ulcorner A \urcorner \leftrightarrow A$ where A belongs to a restricted class of sentences, possibly containing the conditional but only in the context of **HYPE**-logical truths. This appears to be an improvement over standard fixed-point models proper of Kripke's theory of truth [12], since the equivalence between A and $\mathrm{T}\ulcorner A \urcorner$ can now be expressed in the object language.

The expressive power of fixed-point models based on **HYPE** has impact on the construction of formal (axiomatic) theories of truth in **HYPE**. Nonclassical axiomatizations of fixed-point semantics are known to be deductively weak, if compared to classical alternatives [11]. As shown in [9], **HYPE** enables one to overcome such weakness, and reach the same strength of classical fixed-point theories, notably of the Kripke-Feferman theory [7]. Axiomatizations of Kripke's theory of truth are typically formulated over classical arithmetical theories, such as Peano Arithmetic. Therefore, such theories already assume a nontrivial amount of classical mathematics. *Over this base*, truth-theoretic axioms in **HYPE** are known to yield significant strength.

To measure the foundational significance of **HYPE**, it is then natural to ask whether abstraction principles obtained via fixed-point models for **HYPE** are sufficient to develop a nontrivial amount of mathematical objects and concepts and whether they are compatible with suitable extensionality principles characteristic of sets or classes. It's clear that, given the intuitionistic

2010 *Mathematics Subject Classification.* Primary 03E70, Secondary 03B60.
Key words and phrases. Nonclassical Logics, Set Theory, Abstraction, Extensionality.

[1] A standard reference on **FDE** is [1].

flavour of **HYPE**, Curry's paradox prevents one from considering *naïve* abstraction principles, but one can still consider restricted abstraction principles allowing for a non-trivial amount of self-membership. In the paper we explore several such abstraction and extensionality principles. The main results of the paper are:

(i) The consistency of abstraction and comprehension principles over **HYPE** restricted to a class of formulae Φ whose properties are analogous to the class of sentences appearing in Leitgeb's T-schema (Proposition 3.7), and the inconsistency of such principles with the standard formulation of extensionality (Lemma 3.8).

(ii) The incompatibilty of the so-called *weak extensionality principle* with fixed-point semantics for abstraction in **HYPE**.[2] Specifically, we show that no fixed-point model of Φ-abstraction nor models of further natural restrictions of it are models of weak extensionality (Proposition 3.9).

(iii) The formulation of a suitable extensionality principle for **HYPE** (a weak extensionality principle formulated with contraposable biconditionals) and its consistency with a class of formulae $\Psi \subseteq \Phi$ obtained by suitably regimenting the behaviour of identity (Theorem 3.11). This result requires some non-trivial adaptation of a classic proof by Brady [3, 4, 5].

Although, as shown in [9], arithmetical axioms over **HYPE** display the same proof-theoretic strength as classical arithmetic, the development of classical set/class theory in **HYPE** is severely impeded. As we will exemplify in §4, the abstraction and extensionality principles studied in the paper are already at odds with standard definitions of singletons and power sets. By contrast, class theory in **HYPE** displays some interesting features if compared with other nonclassical theories; for instance, as shown in §4, it delivers stable notions of empty and universal sets, and a well-behaved notion of subset.

2. HYPE

We recall the basics of the logic **HYPE**. We consider the language $\mathcal{L}_{\in}^{\rightarrow}$, with logical symbols $\neg, \vee, \rightarrow, \forall, =$, a propositional constant $\bot$, a set of variables $\mathtt{Var}$ and whose signature contains the binary membership predicate $\in$. $\mathcal{L}_{\in}^{\rightarrow}$ features an abstraction operator $\{\cdot : -\}$, where $\cdot$ stands for a variable and $-$ stands for a formula of $\mathcal{L}_{\in}^{\rightarrow}$ in which the variable indicated by $\cdot$ may be free. Terms and formulae of $\mathcal{L}_{\in}^{\rightarrow}$, with their free variables $\mathtt{FV}(\cdot)$ are inductively defined in a standard fashion. The clause for abstraction terms is:

– If $x, x_1, \dots, x_n \in \mathtt{Var}$ and φ is a formula of $\mathcal{L}_{\in}^{\rightarrow}$, then $\{x \colon \varphi(x, x_1, \dots, x_n)\}$ is a term of $\mathcal{L}_{\in}^{\rightarrow}$ with free variables $\mathtt{FV}(\{x \colon \varphi(x, x_1, \dots, x_n)\}) = \mathtt{FV}(\varphi) \setminus \{x\}$;

The symbols $\wedge, \exists, \leftrightarrow$ are defined as usual, $\top$ is defined as $\neg\bot$. Intuitionistic negation $\sim A$ is defined as $A \rightarrow \bot$ and the material conditional $A \supset B$ as $\neg A \vee B$. Material equivalence $A \equiv B$ is defined as $(A \supset B) \wedge (B \supset A)$. We define $\mathcal{L}_{\in}$ as the $\rightarrow$-free version of $\mathcal{L}_{\in}^{\rightarrow}$.

Proof-thoretically, **HYPE** can be described via the multi-conclusion sequent calculus $\mathbf{G1h_{cd}}$ introduced by Fischer et al. [9], which is recalled below. This calculus combines standard rules of **FDE** with rules for a multi-conclusion calculus for intuitionistic implication. Sequents are expressions of form $\Gamma \Rightarrow \Delta$, where Γ, Δ are multisets of $\mathcal{L}_{\in}^{\rightarrow}$-formulae. For a multiset $\Gamma = \gamma_1, \dots, \gamma_n$, we define $\neg\Gamma = \neg\gamma_1, \dots, \neg\gamma_n$. A formula A is derivable if the sequent $\Rightarrow A$ is provable in $\mathbf{G1h_{cd}}$.

(ID) $A \Rightarrow A$ (L$\bot$) $\bot \Rightarrow$

[2]Weak extensionality is a set-theoretic rendering of the indiscernibility of identicals: if x and y are coextensive, then they are members of the same sets.

$$(\mathtt{LW})\ \frac{\Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta} \qquad (\mathtt{RW})\ \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A}$$

$$(\mathtt{LC})\ \frac{A, A, \Gamma \Rightarrow \Delta}{A, \Gamma \Rightarrow \Delta} \qquad (\mathtt{RC})\ \frac{\Gamma \Rightarrow \Delta, A, A}{\Gamma \Rightarrow \Delta, A}$$

$$(\mathtt{L}\vee)\ \frac{A, \Gamma \Rightarrow \Delta \qquad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta} \qquad (\mathtt{R}\vee)\ \frac{\Gamma \Rightarrow A, B, \Delta}{\Gamma \Rightarrow A \vee B, \Delta}$$

$$(\mathtt{L} \rightarrow)\ \frac{\Gamma \Rightarrow \Delta, A \qquad B, \Gamma \Rightarrow \Delta}{A \rightarrow B, \Gamma \Rightarrow \Delta} \qquad (\mathtt{R} \rightarrow)\ \frac{\Gamma, A \Rightarrow B}{\Gamma \Rightarrow A \rightarrow B, \Delta}$$

$$(\mathtt{ConCp})\ \frac{\Gamma \Rightarrow \neg\Delta}{\Delta \Rightarrow \neg\Gamma} \qquad (\mathtt{ClCp})\ \frac{\neg\Gamma \Rightarrow \Delta}{\neg\Delta \Rightarrow \Gamma}$$

$$(\mathtt{L}\forall)\ \frac{A(t), \Gamma \Rightarrow \Delta}{\forall x A, \Gamma \Rightarrow \Delta} \qquad (\mathtt{R}\forall)\ \frac{\Gamma \Rightarrow \Delta, A(y)}{\Gamma \Rightarrow \Delta, \forall x A}$$

y not free in $\Gamma, \Delta, \forall x A$.

$$\mathtt{Cut}\ \frac{\Gamma \Rightarrow \Delta, A \qquad A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

Identity behaves classically.

$$(\mathtt{Ref}) \qquad \Rightarrow t = t$$

$$(\mathtt{Rep}) \qquad s = t, A(s) \Rightarrow A(t)$$

The extension of $\mathbf{G1h_{cd}}$ with `Ref` and `Rep` is called $\mathbf{G1h^{=}_{cd}}$.

A semantics for **HYPE** can be given in a Routley-style version introduced by Speranski [22]. An involutive Routley frame is a triple $\mathfrak{F} = \langle W, \leq, * \rangle$ such that:

- W is a non-empty set of states;
- $\leq$ is a pre-ordering on W;
- $* : W \longrightarrow W$ is an antimonotone, involutive function on W, that is, for all $w, v \in W$, if $v \leq w$ then $w^* \leq v^*$ and $w^{**} = w$.

Let $\mathfrak{M} = \langle \mathfrak{F}, D, X, I \rangle$ be a constant domain **HYPE**-model for $\mathcal{L}^{\rightarrow}_{\in}$ iff:

- $\mathfrak{F}$ is an involutive Routley frame.
- D is a constant domain.
- $X : W \longrightarrow \mathcal{P}(D^2)$ is a valuation function interpreting $\in$ at each state.
- $I : \mathtt{Term} \longrightarrow D$ is an interpretation function giving the semantic values of terms of $\mathcal{L}^{\rightarrow}_{\in}$ over D. The semantic value of terms $I(t)[\vec{a}]$ is given as follows, where $\vec{a}$ is a tuple of elements of D:

$$I(v_i)[\vec{a}] = a_i,$$
$$I(\{x\colon \varphi(x, z_1, \ldots, z_k)\})[\vec{a}] = I(\{x\colon \varphi(x, a_1, \ldots, a_k)\}).$$

- If $u, v \in W$ and $u \leq v$, then $X(u) \subseteq X(v)$.

Here and in the following we will let a name in $\mathcal{L}^{\rightarrow}_{\in}$ the element a of the domain D.

Truth in a **HYPE** *model* for $\mathcal{L}^{\rightarrow}_{\in}$ can now be defined. Let $\mathfrak{M}$ be a constant domain **HYPE**-model for $\mathcal{L}^{\rightarrow}_{\in}$. For every $w \in W$, we define $\mathfrak{M}, w \Vdash \varphi$ inductively as follows:

(i) $\mathfrak{M}, w \Vdash (t_1 \in t_2)[\vec{a}]$ iff $(I(t_1)[\vec{a}], I(t_2)[\vec{a}]) \in X(w)$;
(ii) $\mathfrak{M}, w \Vdash (t_1 = t_2)[\vec{a}]$ iff $I(t_1)[\vec{a}] = I(t_2)[\vec{a}]$;

(iii) $\mathfrak{M}, w \Vdash \varphi \wedge \psi$ iff $\mathfrak{M}, w \Vdash \varphi$ and $\mathfrak{M}, w \Vdash \psi$;
(iv) $\mathfrak{M}, w \Vdash \varphi \vee \psi$ iff $\mathfrak{M}, w \Vdash \varphi$ or $\mathfrak{M}, w \Vdash \psi$;
(v) $\mathfrak{M}, w \Vdash \neg\varphi$ iff $\mathfrak{M}, w^* \nVdash \varphi$;
(vi) $\mathfrak{M}, w \Vdash \varphi \to \psi$ iff for all $w' \in W$ such that $w \leq w'$, if $\mathfrak{M}, w' \Vdash \varphi$ then $\mathfrak{M}, w' \Vdash \psi$;
(vii) $\mathfrak{M}, w \Vdash \forall x\varphi$ iff, for all $a \in D$, $\mathfrak{M}, w \Vdash \varphi(a)$;
(viii) $\mathfrak{M}, w \nVdash \bot$.

We can now define *logical consequence* and *validity*. Let Γ, Δ be sets of sentences. Then:

– $\mathfrak{M}, w \Vdash \Gamma \Rightarrow \Delta$ iff, if $\mathfrak{M}, w \Vdash \gamma$ for all $\gamma \in \Gamma$, then $\mathfrak{M}, w \Vdash \delta$ for some $\delta \in \Delta$;
– $\Gamma \Vdash \Delta$ iff, for all $\mathfrak{M}, w$: $\mathfrak{M}, w \Vdash \Gamma \Rightarrow \Delta$.

Speranski [22, Theorem 7.8] establishes a strong completeness result for **HYPE** formulated in a Hilbert-style calculus QN° with respect to the class of involutive Routley frames. As shown by Fischer et al. [9, Lemma 2], $\mathbf{G1h_{cd}} \vdash \Gamma \Rightarrow \Delta$ iff $\mathrm{QN}^\circ \vdash \bigwedge\Gamma \to \bigvee\Delta$, so $\Gamma \Vdash \Delta$ iff there is a finite $\Delta_0 \subseteq \Delta$ such that $\Gamma \vdash_{\mathrm{QN}^\circ} \Delta_0$. Therefore $\mathbf{G1h_{cd}}$ is also complete with respect to the class of involutive Routley frames.

3. Abstraction and Extensionality in HYPE

The extensions of $\mathbf{G1h^=_{cd}}$ with an unrestricted abstraction principle

$$\Rightarrow \forall x(x \in \{u\colon \varphi(u)\} \leftrightarrow \varphi(x)), \qquad \text{with } \varphi \in \mathcal{L}^{\to}_{\in}$$

or with the unrestricted comprehension principle

$$\Rightarrow \forall x\exists y(x \in y \leftrightarrow \varphi(x)), \qquad \text{with } \varphi \in \mathcal{L}^{\to}_{\in}$$

are inconsistent due to Curry's paradox. Therefore, one needs suitable restrictions.

Definition 3.1. *Let X be a collection of formulae of $\mathcal{L}^{\to}_{\in}$. X-*abstraction *is the schema:*

(X-Abs) $$\Rightarrow \forall x(x \in \{u\colon \varphi(u)\} \leftrightarrow \varphi(x)), \qquad \textit{with } \varphi \in X.$$

We first study a theory based on an abstraction axiom schema restricted to a set of formulae Φ, which roughly corresponds to the restriction of the T-schema operated by Leitgeb in [13, Theorem 42].

Definition 3.2. *Define the set Φ of formulae of $\mathcal{L}^{\to}_{\in}$ as follows:*

– *If φ is a formula of $\mathcal{L}_{\in}$, then $\varphi \in \Phi$.*
– *If φ is a* **HYPE***-logical truth, then $\varphi \in \Phi$.*
– *If φ if of the form $\psi \to \chi$, where ψ does not contain $\in$, and $\chi \in \mathcal{L}_{\in}$, then $\varphi \in \Phi$.*
– *The set Φ is closed under $\wedge, \vee, \neg, \exists, \forall$.*

The formulae in the set Φ can be thought of as "stable" formulae, that is, formulae whose truth-value at a state is invariant under potential changes to the state-space structure. For instance, **HYPE**-logical truths clearly satisfy this property since they are true at every state of any frame. Moreover, every state s in a **HYPE** model has a star state s^*, and since the definition of s^* does not depend on the structure of the frame, but only on s, negated formulae are also stable. On the other hand, formulae whose main connective is the conditional which are not logical truths are non-stable because they are evaluated depending on the partial ordering of states in the model under consideration, hence their value changes if we modify the frame structure.

Definition 3.3 ($\mathbf{HBST}_\Phi$). *The theory $\mathbf{HBST}_\Phi$ is obtained by extending $\mathbf{G1h^=_{cd}}$ in $\mathcal{L}^{\to}_{\in}$ with the Φ-abstraction schema*

(Φ-Abs) $$\Rightarrow \forall x(x \in \{u\colon \varphi(u)\} \leftrightarrow \varphi(x)), \qquad \textit{with } \varphi \in \Phi.$$

A comprehension axiom schema can easily be derived from Φ-ABS.

(Φ-COM) $\Rightarrow \forall x \exists y (x \in y \leftrightarrow \varphi(x))$, with $\varphi \in \Phi$.

The compatibility of extensionality axioms with abstraction schemata—a defining feature of set theories as opposed to property theories in **HYPE**—will be thoroughly investigated in section 3.2. To this end, we introduce a weakening of the abstraction schema to a set of formulae $\Psi \subseteq \Phi$, which restricts the formulae containing identity allowed in instances of abstraction.[3]

Definition 3.4. *Define the set Ψ of formulae of $\mathcal{L}_{\in}^{\rightarrow}$ as follows:*

- *If φ is of the form $x \in y$, then $\varphi \in \Psi$.*
- *If φ is a **HYPE**-logical truth or a propositional constant, then $\varphi \in \Psi$.*
- *If φ if of the form $\psi \to \chi$, where $\psi, \chi \in \Psi$, ψ does not contain $\in$, and χ does not contain $\to$, then $\varphi \in \Psi$.*
- *The set Ψ is closed under $\wedge, \vee, \neg, \exists, \forall$.*

Definition 3.5 ($\mathbf{HBST}_\Psi$)**.** *The theory $\mathbf{HBST}_\Psi$ is obtained by extending $\mathbf{G1h}^{=}_{\mathbf{cd}}$ in $\mathcal{L}_{\in}^{\rightarrow}$ with the Ψ-abstraction schema*

(Ψ-ABS) $\Rightarrow \forall x (x \in \{u \colon \varphi(u)\} \leftrightarrow \varphi(x))$, *with $\varphi \in \Psi$.*

Because of the limitations imposed on identity in instances of abstraction, Ψ-ABS (unlike Φ-ABS) is compatible with some extensionality principles. In particular, we consider three extensionality axioms. The first one, which we will label *strong extensionality*—corresponding to the standard formulation of extensionality in classical set theory—is the following:

(EXTS) $\Rightarrow \forall x \forall y (\forall u (u \in x \leftrightarrow u \in y) \to x = y)$.

Next, we introduce a so called axiom of *weak extensionality*, whose consequent is inspired by a notion of identity as indiscernibility:[4]

(EXTW) $\Rightarrow \forall x \forall y (\forall u (u \in x \leftrightarrow u \in y) \to \forall w (x \in w \leftrightarrow y \in w))$.

As we shall see shortly, EXTW is not a suitable extensionality principle for **HYPE**. Instead, we will show that the right version of weak extensionality is a modification of it obtained by guaranteeing that the antecedent and the consequent of the axiom are formulated with contraposable biconditionals. We introduce the abbreviation $\varphi \rightleftarrows \psi$ standing for $(\varphi \leftrightarrow \psi) \wedge (\neg\varphi \leftrightarrow \neg\psi)$, to formulate the following axiom:

(EXTC) $\Rightarrow \forall x \forall y (\forall u (u \in x \rightleftarrows u \in y)) \to \forall w (x \in w \rightleftarrows y \in w))$.[5]

We will show in particular that:

(i) EXTS is inconsistent with $\mathbf{HBST}_\Phi$.
(ii) none of the natural, fixed-point models for $\mathbf{HBST}_\Psi$ we will construct in section 3.1 validates EXTW. As a consequence of this, none of the natural fixed-point models for $\mathbf{HBST}_\Phi$ validates EXTW, and none of the natural fixed-point models for $\mathbf{HBST}_\Psi$ validates EXTS.
(iii) EXTC is *consistent* with $\mathbf{HBST}_\Psi$, and in fact it has a natural model. On the other hand, none of the fixed-point models for $\mathbf{HBST}_\Phi$ constructed in section 3.1 validates EXTC.

[3]Without the restriction, the "natural" model for abstraction cannot be extended to suitable extensionality principles. Details are provided in Section 3.2; see in particular Observation 3.1.

[4]In the context of nonclassical class theory, weak extensionality has received much attention: see [3, 8].

[5]Note that Brady's constructions for the consistency of abstraction and weak extensionality both in partial and paraconsistent settings [4, 5] employ logics with contrapositive conditionals. Field et al. [8] generalise these constructions to settings with primitive conditionals which do not contrapose, like **HYPE**'s $\to$.

3.1. **A term model for $\mathbf{HBST}_\Phi$.** We construct a constant domain **HYPE** model for $\mathcal{L}_\in^\rightarrow$, adapting to the present setting of **HYPE** a construction by Brady [3, 4], which will turn out to be a model for $\mathbf{HBST}_\Phi$ (and, consequently, for $\mathbf{HBST}_\Psi$). The domain of our model will be the set of constant abstraction terms of $\mathcal{L}_\in^\rightarrow$. The membership predicate is interpreted by adapting the strategy used by Leitgeb in [13, Theorem 42]. Starting with a **HYPE** model assigning arbitrary pairs of constant terms to membership, we carve out a lattice of fixed points via a suitable monotone operator on states. The **HYPE** model with this lattice as its state space will then be our intended model for Φ-abstraction (and Ψ-abstraction). Later, we show that a submodel of our model for Ψ-abstraction also satisfies ExtC, by our adaptation of Brady's results.

The domain of the model is the set of constant abstraction terms $\mathtt{CT} := \{\{x\colon \varphi\} \mid \mathtt{FV}(\varphi) \subseteq \{x\}\}$. Take a countable set of points $\mathbb{W}$, which will be the set of states in our frame. A *valuation* F on $\mathbb{W}$ can be seen as a pair $\langle F^+, F^-\rangle$, with $F^\pm\colon \mathbb{W} \longrightarrow \mathcal{P}(\mathtt{CT}^2)$, where F^+ and F^- denote functions that assign to elements of $\mathbb{W}$ an extension and an antiextension of $\in$, respectively.

We will be interested in the specific valuation $\mathtt{F}$, which distributes sets of pairs over elements of $\mathbb{W}$ without constraints, in such a way that all possible combinations of pairs of elements of $\mathtt{CT}$ are assigned to some element of $\mathbb{W}$, that is:

$$\bigcup_{s\in\mathbb{W}} \mathtt{F}^+(s) = \mathcal{P}(\mathtt{CT}^2), \qquad \bigcup_{s\in\mathbb{W}} \mathtt{F}^-(s) = \mathcal{P}(\mathtt{CT}^2).$$

The valuation $\mathtt{F}$ can be used to induce an involutive Routley frame on $\mathbb{W}$:

- $\mathbb{W}$ is a nonempty set;
- Let $\leq$ be the following ordering on elements of $\mathbb{W}$ based on $\mathtt{F}$: if $v, w \in \mathbb{W}$, $v \leq w$ iff $\mathtt{F}^+(v) \subseteq \mathtt{F}^+(w)$ and $\mathtt{F}^-(v) \subseteq \mathtt{F}^-(w)$ (henceforth abbreviated as $\mathtt{F}(v) \subseteq \mathtt{F}(w)$);
- Given $\mathtt{F}$, we can define the star function $*\colon \mathbb{W} \longrightarrow \mathbb{W}$ as follows:

$$\mathtt{F}^+(w^*) := \{(a,b) | (a,b) \notin \mathtt{F}^-(w)\}, \qquad \mathtt{F}^-(w^*) := \{(a,b) | (a,b) \notin \mathtt{F}^+(w)\}.$$

  Note that $w^{**} = w$ and that for all $x, y \in \mathbb{W}$, if $x \leq y$ then $y^* \leq x^*$.

Let $\mathfrak{F} = \langle \mathbb{W}, \leq, *\rangle$ be the frame induced on $\mathbb{W}$ by $\mathtt{F}$. The model $\mathfrak{M} = \langle \mathfrak{F}, \mathtt{CT}, \mathtt{F}, \mathtt{I}\rangle$ is then specified by $\mathtt{F}$ and an interpretation function $\mathtt{I}$ which gives the semantic value of abstraction terms. We first define the semantic value of terms $\mathtt{I}(t)[\vec{a}]$, where $\vec{a}$ is a tuple of elements of $\mathtt{CT}$:

$$\mathtt{I}(v_i)[\vec{a}] := a_i,$$
$$\mathtt{I}(\{x\colon \varphi(x, z_1, \dots, z_k)\})[\vec{a}] := \{x\colon \varphi(x, a_1, \dots, a_k)\}.$$

As an obvious consequence, for any $\vec{a}$, $\mathtt{I}(t)[\vec{a}] = t$ for $t \in \mathtt{CT}$. The definition of truth in a **HYPE**-model above can be adapted to the present setting to obtain the relation $\mathfrak{M} \Vdash \varphi[\vec{a}]$, whose key clauses are now:

$$\mathfrak{M}, w \Vdash (s \in t)\,[\vec{a}] \text{ iff } (\mathtt{I}(s)[\vec{a}], \mathtt{I}(t)[\vec{a}]) \in \mathtt{F}^+(w),$$
$$\mathfrak{M}, w \Vdash (s \notin t)\,[\vec{a}] \text{ iff } (\mathtt{I}(s)[\vec{a}], \mathtt{I}(t)[\vec{a}]) \in \mathtt{F}^-(w).$$

We can now define a jump operator $\mathcal{J}\colon \mathcal{P}(\mathcal{P}(\mathtt{CT}^2))^2 \to \mathcal{P}(\mathcal{P}(\mathtt{CT}^2))^2$, with $\mathcal{J} = \langle \mathcal{J}^+, \mathcal{J}^-\rangle$ such that:

$$\mathcal{J}^+(\mathtt{F}(w)) := \{(a, \{u\colon \varphi(u)\}) \mid \mathfrak{M}, w \Vdash \varphi\,[a]\},$$
$$\mathcal{J}^-(\mathtt{F}(w)) := \{(a, \{u\colon \varphi(u)\}) \mid \mathfrak{M}, w \Vdash \neg\varphi\,[a]\}.$$

The jump operator selects, within $\mathtt{F}$, the states in which the interpretation of $\in$ corresponds to satisfiability at the relevant states. As a consequence, one has that

$$\mathfrak{M}, w \Vdash \varphi(a) \text{ iff } \mathfrak{M}, w' \Vdash a \in \{u\colon \varphi(u)\} \tag{1}$$

$$\mathfrak{M}, w \Vdash \neg\varphi(a) \text{ iff } \mathfrak{M}, w' \Vdash a \notin \{u \colon \varphi(u)\} \tag{2}$$

where w' is such that $\mathtt{F}(w') = \mathcal{J}(\mathtt{F}(w))$. It's worth noting that if $(a, \{u \colon \neg\varphi\}) \in \mathcal{J}^+(\mathtt{F}(w))$, then $(a, \{u \colon \varphi\}) \in \mathcal{J}^-(\mathtt{F}(w))$.

Moreover, the jump is monotone:

$$\text{if } v \leq w \text{ then } v' \leq w', \text{ where } \mathtt{F}(v') = \mathcal{J}(\mathtt{F}(v)), \mathtt{F}(w') = \mathcal{J}(\mathtt{F}(w)). \tag{3}$$

To show (3), note that, if $v \leq w$, then for all $\varphi \in \mathcal{L}_{\in}^{\rightarrow}$, if $\mathfrak{M}, v \Vdash \varphi$ then $\mathfrak{M}, w \Vdash \varphi$, and if $\mathfrak{M}, v \Vdash \neg\varphi$ then $\mathfrak{M}, w \Vdash \neg\varphi$.

$\langle \mathbb{W}, \leq \rangle$, so defined, is a complete lattice: given any $V \subseteq \mathbb{W}$, we can find its infimum and supremum by taking the intersection and union of the extension and antiextension of membership for the states in V.[6] By the the Knaster-Tarski theorem [23]:

Lemma 3.6. *The operator $\mathcal{J}$ will have a set $\mathtt{FIX}$ of fixed points in $\mathbb{W}$, such that, if $w \in \mathtt{FIX}$, $\mathtt{F}^+(w) = \mathcal{J}^+(\mathtt{F}(w))$ and $\mathtt{F}^-(w) = \mathcal{J}^-(\mathtt{F}(w))$. Moreover, $\langle \mathtt{FIX}, \leq \rangle$ is a complete lattice.*

We let $\sup(\mathtt{FIX}) = \mathtt{MAX}$ and $\inf(\mathtt{FIX}) = \mathtt{MIN}$. We note in particular that, by construction, if $w \in \mathtt{FIX}$, for all $\varphi \in \mathcal{L}_{\in}^{\rightarrow}$,

$$\mathfrak{M}, w \Vdash a \in \{u \colon \varphi(u)\} \text{ iff } \mathfrak{M}, w \Vdash \varphi(a) \tag{4}$$

$$\mathfrak{M}, w \Vdash a \notin \{u \colon \varphi(u)\} \text{ iff } \mathfrak{M}, w \Vdash \neg\varphi(a) \tag{5}$$

(4) and (5) amount to metatheoretic versions of naïve abstraction, and they hold at fixed-point states in the model $\mathfrak{M}$.[7]

In order to obtain the satisfaction of some object-linguistic abstraction biconditional, however, we need to define a new model, $\mathfrak{M}'$, whose state space is restricted to the set of fixed points of $\mathcal{J}$ only. However, since the lattice ordering changes when restricting the state space, paradox will force the loss of full naïveté – for instance, the Curry set will immediately generate paradox with full naïveté when $\leq$ is restricted to fixed-point states. In fact, the model $\mathfrak{M}'$ will satisfy the abstraction axiom at most for formulae in Φ.

More precisely, let $\mathtt{F}' = \mathtt{F} \upharpoonright \mathtt{FIX}$. We then define a new structure

$$\mathfrak{M}' = \langle \langle \mathtt{FIX}, \leq', *' \rangle, \mathtt{CT}, \mathtt{F}', \mathtt{I} \rangle$$

such that $\leq' = \leq \upharpoonright \mathtt{FIX}$ and $*' = * \upharpoonright \mathtt{FIX}$. Note also that $*' = *$, since if $w \in \mathtt{FIX}$ then $w^* \in \mathtt{FIX}$. It can easily be observed that $\mathfrak{M}'$ is a constant domain **HYPE** model for $\mathcal{L}_{\in}^{\rightarrow}$.

Next, we check that the fixed-point property (4)-(5) satisfied by the fixed-point states relative to the ordering $\leq$ is satisfied by the new model $\mathfrak{M}'$ for the restricted class of formulae Φ (see Definition 3.2).

Proposition 3.7. *If $\varphi \in \Phi$, then for all $w \in \mathtt{FIX}$, $\mathfrak{M}', w \Vdash a \in \{u \colon \varphi\}$ iff $\mathfrak{M}', w \Vdash \varphi(a)$, and $\mathfrak{M}', w \Vdash a \notin \{u \colon \varphi\}$ iff $\mathfrak{M}', w \Vdash \neg\varphi(a)$.*

Proof. To prove the claim, it is sufficient to show that, for formulae $\varphi \in \Phi$, for all $s \in \mathtt{FIX}$, $\mathfrak{M}, s \Vdash \varphi(a)$ iff $\mathfrak{M}', s \Vdash \varphi(a)$. If φ is a formula of $\mathcal{L}_{\in}$, by the construction of $\mathtt{F}'$, $\mathfrak{M}, s \Vdash \varphi$ iff

[6] This follows from the fact that $\langle \mathcal{P}(\mathcal{P}(\mathtt{CT}^2))^2, \subseteq \rangle$ is a complete lattice.

[7] It may be instructive to see how (4) and (5) do not generate problems for typically paradoxical sets. The evaluation of conditionals at $\mathfrak{M}$ is relative to ordering of states involving *both* fixed-point and non-fixed-point states. This ensures that naïve abstraction is satisfied at fixed-point states without entailing the satisfaction of paradoxical sets. For example, the Curry set $\chi := \{u \colon u \in u \to \bot\}$, satisfies naïve abstraction in $\mathfrak{M}$ without behaving paradoxically. Let $w \in \mathbb{W}$ be the top state in $\mathfrak{M}$, i.e. such that that $(\forall v \in \mathbb{W})(v \leq w)$. Then, by the definition of $\mathtt{F}$, $(\chi, \chi) \in \mathtt{F}^+(w)$, and so $\mathfrak{M}, w \nVdash \chi \in \chi \to \bot$. Since **HYPE** modes are hereditary, for all states $v \in \mathbb{W}$, $\mathfrak{M}, v \nVdash \chi \in \chi \to \bot$, and, by the semantic clause for negation, at the bottom state w^* in $\mathfrak{M}$, we have that $\mathfrak{M}, w^* \Vdash \neg(\chi \in \chi \to \bot)$.

$\mathfrak{M}', s \Vdash \varphi$; it's worth noting that the key fact underlying the case of negated formulae is that star states do not change when moving from $\mathfrak{M}$ to $\mathfrak{M}'$.

If φ is a theorem of **HYPE**, φ will be satisfied by all states in every **HYPE**-model. Hence, by construction, $\mathfrak{M}, w \Vdash \varphi$ for all $w \in \mathtt{FIX}$. Since $\langle \mathtt{FIX}, \leq', *' \rangle$ is an involutive Routley frame and **HYPE** has been shown to be complete with respect to the class of involutive Routley frames, $\mathfrak{M}'$ is a model of **HYPE**, so $\mathfrak{M}', w \Vdash \varphi$ for all $w \in \mathtt{FIX}$.

If φ is of the form $\psi \to \chi$, where ψ does not contain $\in$ and $\chi \in \mathcal{L}_{\in}$: $\mathfrak{M}, w \Vdash \psi \to \chi$ iff for all $w' \geq w$, $\mathfrak{M}, w' \nVdash \psi$ or $\mathfrak{M}, w' \Vdash \chi$. Since sentences not containing $\in$ contain $=$, propositional constants, and logical connectives only, and the interpretation of $=$ is constant across all states both in $\mathfrak{M}$ and in $\mathfrak{M}'$, $\mathfrak{M}, w \Vdash \psi$ iff $\mathfrak{M}', v \Vdash \psi$, for all $w \in \mathbb{W}$ and all $v \in \mathtt{FIX}$. Similarly, for $\chi \in \mathcal{L}_{\in}$, for all $w \in \mathtt{FIX}$, $\mathfrak{M}, w \Vdash \chi$ iff $\mathfrak{M}', w \Vdash \chi$ by what established above.

Since the connectives $\wedge, \vee, \forall, \exists$ are evaluated locally, i.e. relative to specific states, closing a formula of Φ under these connectives will satisfy the lemma. Closure under negation follows by the fact that, if $w \in \mathtt{FIX}$, then also $w^* \in \mathtt{FIX}$.

□

It immediately follows by the previous proposition that the following schema

$$\Rightarrow \forall x(x \in \{u \colon \varphi(u)\} \leftrightarrow \varphi(x)) \qquad \text{with } \varphi \in \Phi$$

holds in $\mathfrak{M}'$. Since $\Psi \subseteq \Phi$, Ψ-Abs also holds in $\mathfrak{M}'$. An obvious existential quantification gives us Φ-Com and comprehension for formulae in Ψ.

3.2. **Extensionality.** Since we are interested in sets in **HYPE**, we now study notions of extensionality compatible (or incompatible) with Φ-Abs and Ψ-Abs. We have already introduced the following extensionality principles:

(ExtS) $$\Rightarrow \forall x \forall y(\forall u(u \in x \leftrightarrow u \in y) \to x = y)$$

(ExtW) $$\Rightarrow \forall x \forall y(\forall u(u \in x \leftrightarrow u \in y) \to \forall w(x \in w \leftrightarrow y \in w))$$

(ExtC) $$\Rightarrow \forall x \forall y(\forall u(u \in x \rightleftarrows u \in y) \to \forall w(x \in w \rightleftarrows y \in w))$$

Notice that (ExtC) implies the principle

(ExtC') $$\forall x \forall y(\forall u(u \in x \rightleftarrows u \in y) \to \forall w(x \in w \leftrightarrow y \in w))$$

We will now show that ExtS is inconsistent with Φ-Abs, and that natural models of $\mathbf{HBST}_{\Psi}$ do not validate ExtW. Then, we will show that ExtC is consistent with Ψ-Abs—hence the theory $\mathbf{HBST}_{\Psi}+$ExtC is consistent. The consistency of (ExtC') with $\mathbf{HBST}_{\Psi}$ readily follows.

3.2.1. *Strong and weak extensionality.* Gilmore [10] showed that in a language with a primitive identity predicate like $\mathcal{L}^{\to}_{\in}$, where identity satisfies `Ref` and `Rep` and the identity predicate can appear in abstraction terms, the axiom of strong extensionality ExtS is inconsistent, that is, the empty sequent $\Rightarrow$ is derivable from it. Gilmore's proof can be reproduced in $\mathbf{HBST}_{\Phi}$.

Lemma 3.8. *ExtS is inconsistent with Φ-Abs in* **HYPE**.

Proof. We check that Gilmore's proof [10, Theorem 2] can in fact be carried out in $\mathbf{HBST}_{\Phi}$. In the proof we employ the following abbreviations:[8]

$$\begin{aligned} y \cap z &:= \{x \colon x \in y \wedge x \in z\} \\ \varnothing &:= \{u \colon u \neq u\} \\ \tau &:= \{u \colon u \in u\} \\ A(s) &:= \{x \colon (x = \tau \wedge s \in x)\} \text{ with } x \text{ not free in } s \end{aligned}$$

[8] In the proof and in the remainder of the paper we employ $a \neq b$ as an abbreviation for $\neg(a = b)$.

$$T := \{u\colon (A(u)\cap\varnothing) = A(u)\}$$

We show that the sequents $\Rightarrow \forall u(u \in A(T)\cap\varnothing) \leftrightarrow u \in A(T))$ and $(A(T)\cap\varnothing) = A(T) \Rightarrow$ are derivable in $\mathbf{HBST}_\Phi$. These facts, together with the sequent

$$\forall u(u \in (A(T)\cap\varnothing) \leftrightarrow u \in A(T)) \Rightarrow (A(T)\cap\varnothing) = A(T),$$

easily derivable from ExtS, yield the empty sequent via cut.

We first show that the assumption that $(A(T)\cap\varnothing) = A(T)$ is "explosive", as it entails the empty set of formulae. In the following, where several steps in the derivation are immediate applications of the rules of $\mathbf{G1h}^{=}_{\mathbf{cd}}$, we compress them via a double line. We keep explicit the steps in which abstraction is employed:

$$\dfrac{\begin{array}{c}
(A(T)\cap\varnothing) = A(T) \Rightarrow (A(T)\cap\varnothing) = A(T)\\ \hline
(A(T)\cap\varnothing) = A(T) \Rightarrow T \in T\\ \hline
(A(T)\cap\varnothing) = A(T) \Rightarrow T \in \tau\\ \hline\hline
(A(T)\cap\varnothing) = A(T) \Rightarrow \tau = \tau \wedge T \in \tau\\ \hline
(A(T)\cap\varnothing) = A(T) \Rightarrow \tau \in A(T)\\ \hline\hline
(A(T)\cap\varnothing) = A(T) \Rightarrow \tau \in (A(T)\cap\varnothing)\\ \hline\hline
(A(T)\cap\varnothing) = A(T) \Rightarrow \tau \in \varnothing\\ \hline
(A(T)\cap\varnothing) = A(T) \Rightarrow \tau \neq \tau
\end{array} \qquad \tau \neq \tau \Rightarrow}{(A(T)\cap\varnothing) = A(T) \Rightarrow}$$

Now we show that $(A(T)\cap\varnothing)$ and $A(T)$ are extensionally equivalent.

$$\begin{array}{c}
u \in A(T) \Rightarrow u \in A(T)\\ \hline
u \in A(T) \Rightarrow u = \tau \wedge T \in u\\ \hline\hline
u \in A(T) \Rightarrow T \in \tau\\ \hline
u \in A(T) \Rightarrow T \in T\\ \hline
u \in A(T) \Rightarrow (A(T)\cap\varnothing) = A(T)\\ \hline\hline
u \in A(T) \Rightarrow u \in (A(T)\cap\varnothing)\\ \hline
\Rightarrow u \in A(T) \to u \in (A(T)\cap\varnothing)
\end{array}$$

From the proof above, to conclude $\Rightarrow \forall u(u \in (A(T)\cap\varnothing) \leftrightarrow u \in A(T))$ it suffices to notice that from $u \in A(T) \Rightarrow u \in A(T)$ one can easily derive $u \in (A(T)\cap\varnothing) \Rightarrow u \in A(T)$. ☐

Weak extensionality, to which we now turn, displays a more interesting status in **HYPE**. We first state some negative results concerning (ExtW). Unlike what happens in the case of standard paracomplete and paraconsistent (and inconsistent) settings—see [3, 4, 5]—(ExtW) is not compatible with our **HYPE** models for abstraction.

Proposition 3.9. *No model of* $\mathbf{HBST}_\Psi$ *based on fixed points of* $\mathcal{J}$ *is a model of* (ExtW).

Proof. We can find a counterexample to the axiom. Define the following:

$$\begin{aligned}
\tau &:= \{x\colon x \in x\}\\
b &:= \{x\colon \top\}\\
a &:= \{x\colon \top \wedge \tau \in \tau\}\\
w &:= \{x\colon \forall y(y \in x)\}\\
w' &:= \{x\colon x \notin w\}
\end{aligned}$$

In a model $\mathfrak{M}$ of $\mathbf{HBST}_\Psi$, let s be a state such that $\tau \in \tau$ is a glut at s, that is $\mathfrak{M}, s \Vdash \tau \in \tau$ and $\mathfrak{M}, s \Vdash \tau \notin \tau$. Then, $\mathfrak{M}, s^* \nVdash \tau \in \tau$ and $\mathfrak{M}, s^* \nVdash \tau \notin \tau$.

Hence, $\mathfrak{M}, s \Vdash \forall u(u \in a \leftrightarrow u \in b)$, while $\mathfrak{M}, s^* \nVdash \forall u(u \in a \leftrightarrow u \in b)$, so $\mathfrak{M}, s \Vdash \neg\forall u(u \in a \leftrightarrow u \in b)$.

Since $\mathfrak{M}, s \Vdash \top$, $\mathfrak{M}, s \Vdash \forall y(y \in b)$, and $\mathfrak{M}, s \Vdash \forall y(y \in a)$. Hence, $\mathfrak{M}, s \Vdash a \in w$ and $\mathfrak{M}, s \Vdash b \in w$.

However, we know that $\mathfrak{M}, s^* \nVdash \top \wedge \tau \in \tau$, while, obviously, $\mathfrak{M}, s^* \Vdash \top$. Hence, $\mathfrak{M}, s^* \Vdash \forall y(y \in b)$, while $\mathfrak{M}, s^* \nVdash \forall y(y \in a)$. Hence, $\mathfrak{M}, s^* \Vdash b \in w$, $\mathfrak{M}, s^* \nVdash a \in w$, so $\mathfrak{M}, s \Vdash a \notin w$, $\mathfrak{M}, s \nVdash b \notin w$ by the clause for negation in **HYPE**.

Hence, $\mathfrak{M}, s \Vdash a \in w', \mathfrak{M}, s \Vdash a \notin w'$, while $\mathfrak{M}, s \nVdash b \in w', \mathfrak{M}, s \Vdash b \notin w'$. Hence, $\mathfrak{M}, s \nVdash \forall v(a \in v \leftrightarrow b \in v)$, while $\mathfrak{M}, s \Vdash \forall u(u \in a \leftrightarrow u \in b)$. □

The following counterexample immediately entails also that no model of $\mathbf{HBST}_\Phi$ based on fixed points of $\mathcal{J}$ is a model of (ExtW). Furthermore, it also implies that no model of $\mathbf{HBST}_\Psi$ based on fixed points of $\mathcal{J}$ is a model of (ExtS).

Remark. The previous result only relies on the fact that, in a model of $\mathbf{HBST}_\Psi$, there are states s_0, s_1 with $s_0^* = s_1$, such that a glut in s_0 is a gap in s_1, or viceversa. So, we conjecture that the proof above can be extended to show that *no model* of $\mathbf{HBST}_\Psi$ (and, consequently, no model of $\mathbf{HBST}_\Phi$) can satisfy (ExtW). Given the completeness of the logic, this would then amount to a proof of the inconsistency of (ExtW) with $\mathbf{HBST}_\Psi$.

The previous result essentially rests on the fact that a **HYPE**-conditional can be true at one state even if, say, the antecedent is a glut and the consequent is classically true. This compromises the role of extensional equivalence. It then seems that the appropriate formulation of extensional equivalence in the context of **HYPE** should amount to coincidence of both extension and antiextension, which is what (ExtC) delivers, since it is formulated with contrapositive biconditionals. It turns out that (ExtC) is indeed consistent with $\mathbf{HBST}_\Psi$. However, we also show that no term model of $\mathbf{HBST}_\Phi$ is a model of (ExtC).

3.2.2. *The consistency of ExtC with* $\mathbf{HBST}_\Psi$. To show the consistency of (ExtC), we consider a "minimal" model of $\mathbf{HBST}_\Psi$, obtained by restricting our attention to the least and greatest states in $\mathfrak{M}'$. Indeed, we can show that there are non-minimal models of $\mathbf{HBST}_\Psi$ which do not satisfy (ExtC).

Proposition 3.10. *We can find non-minimal models of* $\mathbf{HBST}_\Psi$ *which do not satisfy* (ExtC).

Proof. Let $\mathfrak{M}$ be a model of $\mathbf{HBST}_\Psi$ with a state s exhibiting two different "truth-teller" sets $\tau := \{x \colon x \in x\}$ and $\tau' := \{x \colon x \in x \wedge x \in x\}$, such that $(\tau', \tau') \in \mathtt{F}^+(s), (\tau', \tau') \notin \mathtt{F}^-(s)$ and $(\tau, \tau) \in \mathtt{F}^-(s), (\tau, \tau) \notin \mathtt{F}^+(s)$. Then, $\mathfrak{M}, s \Vdash \tau' \in \tau$, since $(\tau', \tau') \in \mathtt{F}^+(s)$, and $\mathfrak{M}, s \Vdash \tau \notin \tau'$, since $(\tau, \tau) \in \mathtt{F}^-(s)$.

Then, $\mathfrak{M}, s \Vdash \forall w(w \in \tau \rightleftarrows w \in \tau')$ but $\mathfrak{M}, s \nVdash \forall v(\tau \in v \rightleftarrows \tau' \in v)$, since $\mathfrak{M}, s \Vdash \tau' \in \tau$ but $\mathfrak{M}, s \nVdash \tau \in \tau$. □

The minimal **HYPE**-*model* (or two-state **HYPE** model) $\mathfrak{M}_{\mathtt{min}}$ for Ψ-Abs is the restriction of the model $\mathfrak{M}'$ presented in section 3.1 to the frame consisting of its minimal and maximal states. More precisely, $\mathfrak{M}_{\mathtt{min}} = \langle \mathfrak{F}, \mathtt{CT}, \mathtt{F}, \mathtt{I} \rangle$, $\mathfrak{F} = \langle W, \leq, * \rangle$, $W = \{s_l, s_g\}$, $\mathtt{F}(s_l) = \mathtt{MIN}$, $\mathtt{F}(s_g) = \mathtt{MAX}$, where $\mathtt{MIN}$ and $\mathtt{MAX}$ are, respectively, the minimal and the maximal fixed points of the operator $\mathcal{J}$ defined in section 3.1. We can easily see that $s_l = s_g^*$.

To show the consistency of (ExtC) over minimal models of $\mathbf{HBST}_\Psi$, we largely follow Field et al.'s [8] rewriting of Brady's constructions for partial and paraconsistent logics [3, 4, 5]. We adapt Field et al.'s proof to the **HYPE** setting, using dualities between the minimal and the maximal fixed point to establish the consistency of the axiom over our two-state model $\mathfrak{M}_{\mathtt{min}}$.

It will be useful in what follows to introduce truth values. A sentence φ is said to have truth value 1 at state s in a $\mathbf{HBST}_\Psi$-model $\mathfrak{M}$, in symbols $|\varphi|_{\mathfrak{M},s} = 1$, iff it is determinately true at s, so iff $\mathfrak{M}, s \Vdash \varphi$ and it is not the case that $\mathfrak{M}, s \Vdash \neg\varphi$. Similarly, $|\varphi|_{\mathfrak{M},s} = 0$ iff φ is determinately false at s, that is, iff $\mathfrak{M}, s \Vdash \neg\varphi$ and it is not the case that $\mathfrak{M}, s \Vdash \varphi$. A sentence φ is a gap at s, indicated by $|\varphi|_{\mathfrak{M},s} = n$, iff $\mathfrak{M}, s \nVdash \varphi$ and $\mathfrak{M}, s \nVdash \neg\varphi$, and it is a glut at s, indicated by $|\varphi|_{\mathfrak{M},s} = b$ iff $\mathfrak{M}, s \Vdash \varphi$ and $\mathfrak{M}, s \Vdash \neg\varphi$. Note that $|\varphi|_{\mathfrak{M},s} = n$ iff $|\varphi|_{\mathfrak{M},s^*} = b$, and if $|\varphi|_{\mathfrak{M},s} \in \{0, 1\}$ then $|\varphi|_{\mathfrak{M},s^*} = |\varphi|_{\mathfrak{M},s}$. In the following we will omit specification of the model in case it is clear from the context and simply write $|\varphi|_s$ for the truth value of the sentence φ at state s.

Theorem 3.11. *A two-state model of* $\mathbf{HBST}_\Psi$ *satisfies* ExtC, *that is, for* $s \in \{s_l, s_g\}$, $\mathfrak{M}_{\mathtt{min}}, s \Vdash \forall u(u \in a \rightleftarrows u \in b) \to \forall v(a \in v \rightleftarrows b \in v)$.

We divide the proof of this claim in a few propositions and lemmata. We first recall some results about the minimal and maximal fixed point that will be useful in what follows. A function $f\colon X \longrightarrow Y$, where X, Y are posets, is said to be Scott-continuous if it preserves all directed suprema, that is, if for every directed $V \subseteq X$ with supremum in X, $\bigsqcup f(V) = f(\bigsqcup V)$, where $\bigsqcup$ is the directed join.

Proposition 3.12. *The evaluation of the jump* $\mathcal{J}$ *is Scott-continuous on* $\langle \mathbb{W}, \leq \rangle$.[9]

Proof. This follows immediately from the fact that for any set of states $X \subseteq \mathbb{W}$, $\mathcal{J}(\mathtt{F}(\bigcup_{i\in X} s_i)) = \bigcup_{i \in X} \mathcal{J}(\mathtt{F}(s_i))$, because $s_i \leq s_j$ iff $\mathtt{F}(s_i) \subseteq \mathtt{F}(s_j)$. □

By the fact that the operator $\mathcal{J}$ is Scott-continuous, we can then infer that $\mathtt{MIN}$ can be reached by iterating the restriction of $\mathcal{J}$ over $\inf(\mathbb{W})$,[10] that is, the state $s_0 \in \mathbb{W}$ such that $\mathtt{F}^+(s_0) = \varnothing, \mathtt{F}^-(s_0) = \varnothing$, and taking unions at limit stages. Similarly, $\mathtt{MAX}$ can be reached by iterating the evaluation of $\mathcal{J}$ over $\sup(\mathbb{W})$, and taking intersections at limits.

Definition 3.13 (Level, Classical Level, Substitution Set)**.**

(1) The level $\nu(\varphi)$ *of a sentence* φ *is the ordinal stage in the iteration of the restriction of* $\mathcal{J}$ *over* $\inf(\mathbb{W})$ *at which* φ *obtains its final truth value. When* φ *is a gap or a glut,* $\nu(\varphi) = 0$. *The* classical level $\nu_c(\varphi)$ *is the non-zero ordinal stage in the iteration of the restriction of* $\mathcal{J}$ *over* $\inf(\mathbb{W})$ *at which* φ *acquires a value in* $\{0, 1\}$.

(2) The a/b-substitution set $\Sigma_{a,b}$ *is defined as*

$$\{A(x) \in \Psi \mid \mathtt{FV}(A) \subseteq \{x\}, \mathfrak{M}_{\mathtt{min}}, s_l \nVdash A(a) \rightleftarrows A(b)\}.$$

In other words, $\Sigma_{a,b}$ *is the set of formulae* $A(x)$ *in* Ψ *with the variable* x *free at most, such that* $A(a)$ *and* $A(b)$ *differ in truth value at the least state in* $\mathfrak{M}_{\mathtt{min}}$.

To show Theorem (3.11), we combine two claims. The first, following directly by (Ψ-Abs) from the definition of $\Sigma_{a,b}$, is

$$\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \forall v(a \in v \rightleftarrows b \in v) \text{ iff } \Sigma_{a,b} = \varnothing. \tag{6}$$

The second claim, to which the rest of the section will be devoted, is that

$$\Sigma_{a,b} \neq \varnothing \text{ only if } \mathfrak{M}_{\mathtt{min}}, s_l \nVdash \forall u(u \in a \rightleftarrows u \in b). \tag{7}$$

Define $\Sigma_{a,b,\sigma}$ to be the collection

$$\{A(x) \in \Sigma_{a,b} \mid \nu_c(A(a)) \leq \sigma, \text{ or } \nu_c(A(b)) \leq \sigma\};$$

[9]The jump $\mathcal{J}$ is Scott-continuous on $\langle \mathcal{P}(\mathcal{P}(\mathtt{CT}^2))^2, \subseteq \rangle$, where $\mathtt{F}(v) \subseteq \mathtt{F}(w)$ iff $\mathtt{F}^+(v) \subseteq \mathtt{F}^+(w)$ and $\mathtt{F}^-(v) \subseteq \mathtt{F}^-(w)$, since the jump is defined on the valuation function $\mathtt{F}\colon \mathbb{W} \longrightarrow \mathcal{P}(\mathcal{P}(\mathtt{CT}^2))^2$. However, since $\mathtt{F}$ (and consequently $\mathcal{J}(\mathtt{F})$) is always evaluated at some state $w \in \mathbb{W}$, and the ordering of states in $\mathbb{W}$ is induced by $\mathtt{F}$, by a slight abuse of notation we can identify states with their valuation.

[10]Strictly speaking, we are applying the jump to the restriction of the evaluation function $\mathtt{F}$ over $\inf(\mathbb{W})$.

that is, the set of formulae in $\Sigma_{a,b}$ such that $A(a)$ or $A(b)$ has level less than or equal to σ and value 0 or 1 at the minimal fixed point of $\mathcal{J}$. Then, (7) can be rephrased as: if $\exists\sigma(\Sigma_{a,b,\sigma}) \neq \varnothing$ then $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash \forall u(u \in a \rightleftarrows u \in b)$.[11] We show this claim for s_l and then use the dualities between least and greatest state to obtain the result for s_g as well, as shown in Lemma 3.16. We divide the proof in three claims.

Lemma 3.14. *For all σ: if $\Sigma_{a,b,\sigma} \neq \varnothing$ then $\Sigma_{a.b,\sigma}$ contains a formula of the form $t(x) \in x$, with t a term.*

Proof. Assume there exists an ordinal σ such that $\Sigma_{a,b,\sigma} \neq \varnothing$. Let δ be the smallest such. We show the contrapositive: if no formula of the form $t(x) \in x$ is in $\Sigma_{a,b,\delta}$, then $\Sigma_{a,b,\delta} = \varnothing$. The proof is by induction on the complexity of $\varphi \in \Psi$, since all formulae in $\Sigma_{a,b,\sigma}$ are in Ψ.

If φ is atomic, we consider the case in which $\varphi \in \Sigma_{a,b,\delta}$ is of form $t(x) \in \{y\colon \psi(x,y)\}$—the cases in which φ is a logical truth or a propositional constant are immediate. By assumption, $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash t(a) \in \{y\colon \psi(a,y)\} \rightleftarrows t(b) \in \{y\colon \psi(b,y)\}$ and $\nu_c(t(a) \in \{y\colon \psi(a,y)\}) \leq \delta$ or $\nu_c(t(b) \in \{y\colon \psi(b,y)\}) \leq \delta$. Without loss of generality, we assume the former. By definition of δ and $\mathcal{J}$, there's a $\rho < \delta$ such that $\nu_c(\psi(a,t(a))) = \rho$. Therefore, since $\Sigma_{a,b,\rho} = \varnothing$,

$$\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \psi(a,t(a)) \rightleftarrows \psi(b,t(b))$$

The required contradiction is then readily obtained by Ψ-Abs.

If φ is complex, then we check the case in which it has form $\psi \to \chi$—the cases of negation, other connectives and quantifiers are straightforward by induction hypothesis. If φ is $\psi \to \chi$, where $\psi, \chi \in \Psi$, ψ does not contain $\in$ and χ does not contain $\to$. Let $\psi \to \chi \in \Sigma_{a,b,\delta}$. Then, $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash (\psi \to \chi)(a) \rightleftarrows (\psi \to \chi)(b)$, and we can assume without loss of generality that $\nu_c((\psi \to \chi)(a)) \leq \delta$. By induction hypothesis, $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \psi(a) \rightleftarrows \psi(b)$ and $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \chi(a) \rightleftarrows \chi(b)$. Therefore, $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash (\psi \to \chi)(a) \rightleftarrows (\psi \to \chi)(b)$, contradicting our assumption. □

Lemma 3.15. *Let δ be the least ordinal such that $\Sigma_{a,b,\delta} \neq \varnothing$. If $\Sigma_{a,b,\delta}$ contains a formula of the form $t(x) \in x$, then $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash \forall u(u \in a \rightleftarrows u \in b)$.*

Proof. Assume $\Sigma_{a,b,\delta}$ contains a formula of the form $t(x) \in x$. Then $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash t(a) \in a \rightleftarrows t(b) \in b$, and $\nu_c(t(a) \in a) \leq \delta$ or $\nu_c(t(b) \in b) \leq \delta$. Assume, without loss of generality, that $\nu_c(t(a) \in a) \leq \delta$.

Given that $a := \{y\colon \chi(y)\}$ (since all terms are abstraction terms), there is a $\rho < \delta$ such that, by construction of $\mathcal{J}$, $\nu_c(\chi(t(a))) = \rho$, and, since $\Sigma_{a,b,\rho} = \varnothing$, $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \chi(t(a)) \rightleftarrows \chi(t(b))$. Then, by construction of $\mathcal{J}$, $\nu_c(t(b) \in a) \leq \delta$. Then, since $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash t(a) \in a \rightleftarrows t(b) \in b$ and $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash t(a) \in a \rightleftarrows t(b) \in a$, $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash t(b) \in b \rightleftarrows t(b) \in a$, so $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash \forall u(u \in a \rightleftarrows u \in b)$. □

Lemma 3.16. $\mathfrak{M}_{\mathtt{min}}, s_l \Vdash \psi \rightleftarrows \chi$ *iff* $\mathfrak{M}_{\mathtt{min}}, s_g \Vdash \psi \rightleftarrows \chi$.

Proof.

$$\begin{aligned}
\mathfrak{M}_{\mathtt{min}}, s_g \Vdash \psi \rightleftarrows \chi \quad \text{iff} \quad & \mathfrak{M}_{\mathtt{min}}, s_g \Vdash \psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_g \Vdash \chi \text{ and } \mathfrak{M}_{\mathtt{min}}, s_g \Vdash \neg\psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_g \Vdash \neg\chi \\
& \mathfrak{M}_{\mathtt{min}}, s_l \nVdash \neg\psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_l \nVdash \neg\chi \text{ and } \mathfrak{M}_{\mathtt{min}}, s_l \nVdash \psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_l \nVdash \chi \\
& \mathfrak{M}_{\mathtt{min}}, s_l \Vdash \neg\psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_l \Vdash \neg\chi \text{ and } \mathfrak{M}_{\mathtt{min}}, s_l \Vdash \psi \text{ iff } \mathfrak{M}_{\mathtt{min}}, s_l \Vdash \chi \\
& \mathfrak{M}_{\mathtt{min}}, s_l \Vdash \psi \rightleftarrows \chi
\end{aligned}$$

□

[11]This is the case because, by construction, s_l contains only gaps and no gluts.

Proof of Theorem 3.11. We need to verify that both s_l and s_g satisfy (ExtC). Lemmata 3.14, 3.15 give us (7), which directly covers the claim for s_l. By Lemma 3.16, we can reduce the claim for s_g to the former case. □

It's worth noting that if the set of formulae Φ instead of Ψ is considered, ExtC does not have natural models.

Observation 3.1. If $\mathfrak{M}$ is a term model of $\mathbf{HBST}_\Phi$ such that for any $\vec{a}$, $\mathtt{I}(t)[\vec{a}] = t$ for $t \in \mathtt{CT}$, $\mathfrak{M}$ does not satisfy (ExtC).

Proof. Let $a := \{x \colon \top\}$ and $b := \{x \colon \top \wedge \top\}$. $\mathbf{HBST}_\Phi \vdash \forall u(u \in a \rightleftarrows u \in b)$. However, let $\mathfrak{M}$ be a term model of $\mathbf{HBST}_\Phi$ such that for any $\vec{a}$, $\mathtt{I}(t)[\vec{a}] = t$ for $t \in \mathtt{CT}$. For all states s in $\mathfrak{M}$, $\mathfrak{M}, s \Vdash a \in \{x \colon x = a\}$, however, $\mathfrak{M}, s \nVdash b \in \{x \colon x = a\}$. □

Henceforth, we will refer to the theory $\mathbf{HBST}_\Psi$ + ExtC simply as **HBST**. We will also abbreviate $\forall u(u \in x \rightleftarrows u \in y)$ as $x =_e y$ (x is extensionally identical to y) and $\forall w(x \in w \rightleftarrows y \in w)$ as $x =_w y$ (x is weakly identical to y).

Extensional identity in **HBST** behaves in a more "classical" way than other nonclassical (and paraconsistent) set theories based on some form of abstraction and extensionality, especially inconsistent ones. Indeed, in many paraconsistent and inconsistent set theories based on abstraction and extensionality, such as the one introduced by Zach Weber in [24], identity is not classically reflexive (i.e., there are sets which are both self-identical and not self-identical). On the other hand, **HBST** offers us a picture of extensional identity in which all sets are extensionally identical to themselves only. This is due to the behaviour of the **HYPE** conditional, since $\mathbf{HBST} \vdash \forall u(u \in x \rightleftarrows u \in x)$.

Furthermore, extensional identity and extensional difference are stable across states, since extensional identity is defined in terms of coincidence of both extension and antiextension.

We can also note that, if $\mathfrak{M}_{\mathtt{min}}, s \Vdash \forall u(u \in a \rightleftarrows u \in b)$, for $s \in \{s_l, s_g\}$, then $\mathfrak{M}_{\mathtt{min}}, s \Vdash \varphi(a)$ iff $\mathfrak{M}_{\mathtt{min}}, s \Vdash \varphi(b)$ for all formulae $\varphi \in \mathcal{L}^{\rightarrow}_{\in} \setminus \{=\}$, that is, all formulae not containing identity. This can be shown by straightforward induction on the complexity of $\varphi \in \mathcal{L}^{\rightarrow}_{\in} \setminus \{=\}$.

4. Classes in HBST and $\mathbf{HBST}_\Phi$

This section lays out a brief survey of some basic facts about **HBST**. We start by providing a definition of the empty set and the universal set, and then list some results about their behaviour. The possibility to define a universal set, which should behave as the dual of the empty set, constitutes one of the intuitive advantages of considering set theories based on abstraction and extensionality. Indeed, most attempts to formulate one such theory, both based on classical logic (see [10, 6]) and on various nonclassical logics (see [3, 4, 2, 18, 19, 24]) develop some form of empty set and universal set. **HBST** supports well-behaved empty and universal sets, thanks to the behaviour of weak identity and extensional identity licensed by ExtC.

Definition 4.1. *The empty set* $\varnothing := \{x \colon \forall y(x \in y)\}$ *and the universal set* $V := \{x \colon \exists y(x \in y)\}$ *exist by* Ψ*-Abs.*

In **HBST**, the empty set has no members, on pain of explosion. This is proved by exploiting the fact that $\mathbf{HBST} \vdash \forall x(x =_e x)$, so that $\mathbf{HBST} \vdash x \neq_e x \rightarrow \bot$, where $x \neq_e x$ stands for $\neg\forall u(u \in x \rightleftarrows u \in x)$. Indeed, by definition of $\varnothing$, if $x \in \varnothing$ then x is a member of all sets, hence $x \in \{u \colon u \neq_e u\}$. Then, by Ψ-Abs, $x \neq_e x$, obtaining the required contradiction. Similarly, by using the fact that $\mathbf{HBST} \vdash \forall x(x =_e x)$ it can be shown that every set is a member of V.

Proposition 4.2. $\mathbf{HBST} \vdash \neg\exists x(x \in \varnothing)$*,* $\mathbf{HBST} \vdash \forall x(x \in V)$*.*

Since the empty set is provably empty in **HBST**, it can be defined by several other abstraction terms, corresponding to all **HYPE**-logical falsehoods, such as $\{x\colon x \neq_e x\}, \{x\colon x \neq_w x\}, \{x\colon \bot\}$. All these abstraction terms are extensionally identical, and hence weakly identical. Similarly, there are many extensionally identical abstraction terms for the universal set, corresponding to all **HYPE**-logical truths.[12]

The **HYPE** conditional, being reasonably well-behaved, allows us to recover the usual laws of subsets in **HYPE**. This is in contrast to inconsistent naïve set theories with a material conditional, such as Restall's [19] and Omori's [18], which cannot prove the transitivity of subsets.[13] However, in light of our definition of weak identity with the contrapositive biconditional $\rightleftarrows$, subsets need to be defined with contrapositive conditionals to recover their usual properties.

Definition 4.3. *The set x is a subset of the set y, in symbols $x \subseteq y$, iff $\forall u(u \in x \to u \in y \wedge u \notin y \to u \notin x)$. The set x is a proper subset of the set y, in symbols $x \subset y$, iff $x \subseteq y \wedge x \neq_e y$.*

Proposition 4.4. *Subsets have the following properties in* **HBST***:*

(1) **HBST** $\vdash \forall x(\varnothing \subseteq x)$.
(2) *Subsets satisfy the properties which are usually associated with partial orders, i.e.*
  (a) $\Rightarrow \forall x(x \subseteq x)$;
  (b) $\Rightarrow \forall x\forall y(x \subseteq y \wedge y \subseteq x \to x =_w y)$;
  (c) $x \subseteq y, y \subseteq z \Rightarrow x \subseteq z$.

Proposition 4.5. *Proper subsets satisfy the following properties:*

(1) **HBST** $\vdash \forall x(x \not\subset x)$;
(2) *For an arbitrary* **HBST***-model $\mathfrak{M}$, for all objects $a, b \in D$, if at all states s, $\mathfrak{M}, s \Vdash a \subset b$ then for no s' in $\mathfrak{M}$, $\mathfrak{M}, s' \Vdash b \subset a$;*
(3) *For an arbitrary* **HBST***-model $\mathfrak{M}$, for all objects $a, b, c \in D$, if for all states s, $\mathfrak{M}, s \Vdash a \subset b, b \subset c$ then it must be the case that for all states s in $\mathfrak{M}, s \Vdash a \subset c$.*

The complement of a, defined by $a^c := \{u\colon u \notin a\}$, and the restricted complement of a with respect to X, defined as $X \setminus a := \{u\colon u \in X \wedge u \notin a\}$, exist by Ψ-Abs.

It follows from Ψ-Abs that $\vdash \forall x((x^c)^c =_e x)$. Complements also have the following properties, from the De Morgan laws:

$$\vdash \forall x\forall y(x^c \cap y^c =_e (x \cup y)^c)$$
$$\vdash \forall x\forall y(x^c \cup y^c =_e (x \cap y)^c).$$

By definition, **HBST** $\vdash \varnothing =_e V^c$ and, consequently, **HBST** $\vdash V =_e \varnothing^c$. Also, by definition of complement, **HBST** $\vdash \forall x\forall y(x \subseteq y \to y^c \subseteq x^c)$.

Proposition 4.6. *The union $\bigcup w := \{x\colon \exists z(z \in w \wedge x \in z)\}$ of a set w exists by Ψ-Abs. Hence* **HBST** $\vdash \forall x\exists y\forall u(u \in y \leftrightarrow \exists z(z \in x \wedge u \in z))$*; that is,* **HBST** *proves the axiom of Union.*

The intersection $\bigcap w$ of a set w is not available in its standard formulation because of the need to use an instance of $\to$ in the corresponding abstraction term. Correspondingly, the axiom scheme

[12] The behaviour of empty and universal sets in **HBST** differs from the nonclassical set theories developed by Restall [19], Weber [24] and Ripley [20]. In the development of his naïve set theory NST, Ripley defines the empty set as $\varnothing := \{x\colon \bot\}$, the Russell set as $r := \{x\colon x \notin x\}$, and the so-called "Weber set" as $w := \{x\colon r \in r\}$. Then, by his version of extensionality, the empty set is shown to be identical to the Weber set, but NST $\vdash \exists x x \in w$, while NST $\vdash \forall x x \notin \varnothing$. Similar observations are available in Restall's and Weber's frameworks, as they allow for sets x such that $\vdash x = x$ and $\vdash x \neq x$. This reasoning is not available in **HBST**, however, since **HYPE**, although paraconsistent, has no provable gluts.

[13] This is because of the use of the material conditional in an inconsistent set theory. Inconsistent set theories with non-material conditionals, such as Weber's [24], prove the laws of subsets, but display an inconsistent behaviour of identity.

of Separation holds, in **HBST**, only for formulae in Ψ. Finite unions and finite intersections are defined in the standard way; they are associative, commutative and satisfy distribution laws.

Recovering other axioms of **ZF** as theorems of **HBST** proves more difficult. The standard formulation of the axiom of Pairing and of the axiom of Infinity, indeed, require the use of singletons. However, because of the restrictions imposed on the **HYPE** conditional in instances of abstraction, the singleton of a set x, defined as $\{x\} := \{u \colon u =_e x\}$,[14] does not satisfy Ψ-ABS.

The only way to define a form of equivalence which can appear in abstraction terms in **HBST** is to use the material conditional, letting $x =_e^{\supset} y$ (read "x is materially equivalent to y") abbreviate the expression $\forall u(u \in x \equiv u \in y)$, and letting $x =_w^{\supset} y$ ("x is materially weakly equivalent to y") abbreviate $\forall u(x \in u \equiv y \in u)$. Then, the material singleton of a set x is defined as $\{x\}^{\supset} := \{u \colon u =_e^{\supset} x\}$.

However, as is the case for **FDE**, the material conditional in **HYPE** is exceptionally badly behaved, since it invalidates $A \supset A$, modus ponens and the deduction theorem.

This entails that material equivalence and material weak equivalence are not equivalence relations.[15] Hence, if a set x is gappy or glutty at some state in some **HBST** model, then **HBST** $\nvdash x \in \{x\}^{\supset}$.

The intended behaviour of singletons, however, can be recovered by disregarding EXTC and working in $\mathbf{HBST}_\Phi$. Since the identity predicate can now freely appear in instances of abstraction, the abstraction term $\{x\} := \{u \colon u = x\}$ satisfies Φ-ABS, and $\mathbf{HBST}_\Phi \vdash \forall x(x \in \{x\})$. Then, by defining $\{x, y\} := \{u \colon u \in \{x\} \vee u \in \{y\}\}$, the usual Pairing axiom

$$\forall x \forall y \exists v \forall u(u \in v \leftrightarrow u = x \vee u = y)$$

holds in $\mathbf{HBST}_\Phi$ by Φ-ABS. Also, V, defined as above, is easily seen to satisfy the usual formulation of the axiom of Infinity:

$$\exists x(\varnothing \in x \wedge \forall y(y \in x \to y \cup \{y\} \in x)).$$

It must be noted that, since $\mathbf{HBST}_\Phi$ lacks extensionality, and is a theory of properties, rather than sets, abstraction terms are not unique. Hence, for instance, the abstraction term V does not indicate the universal set, but a universal property. Similarly, singletons and pairs are not unique. The downside of $\mathbf{HBST}_\Phi$, then, is that the results provable in **HBST** using EXTC, among which the (weak) uniqueness of the empty and universal set, are not available in $\mathbf{HBST}_\Phi$.

In fact, since identity can appear in the abstraction schema of $\mathbf{HBST}_\Phi$, we can show that that the axioms of Adjunctive Set Theory (see e.g. [16]), which are mutually interpretable with Robinson's arithmetic Q, are derivable in $\mathbf{HBST}_\Phi$:

Proposition 4.7. *The following are provable in* $\mathbf{HBST}_\Phi$*:*

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

Another failure of abstraction, however, cripples both **HBST** and $\mathbf{HBST}_\Phi$; namely, the impossibility to formulate an adequate notion of powerset. Indeed, the abstraction term $\mathcal{P}(x) := \{u \colon u \subseteq x\}$ does not satisfy neither Φ-ABS nor Ψ-ABS.

Similarly to the case of singletons, the only notion of powerset which can appear in instances of abstraction is obtained by using the material conditional. Let $x \sqsubseteq y$ iff $\forall u(u \in x \supset u \in y)$; then, the material powerset of a set x is defined as $\mathcal{P}_{\supset}(x) := \{u \colon u \sqsubseteq x\}$. However, the problems linked to the behaviour of $\supset$ in **HYPE** which we encountered earlier reappear. Indeed, material subsethood is not reflexive and not transitive; hence, for gappy or glutty x, **HBST** $\nvdash x \in \mathcal{P}_{\supset}(x)$.[16]

---

[14] Or, alternatively, as $\{x\} := \{u \colon u =_w x\}$.

[15] To see this, it is sufficient to consider the Russell set r. Since $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash r \in r$ and $\mathfrak{M}_{\mathtt{min}}, s_l \nVdash r \notin r$, **HBST** $\nvdash \forall u(x \in u \equiv x \in u)$, and **HBST** $\nvdash \forall u(u \in x \equiv u \in x)$.

[16] To show the failure of reflexivity, consider the Russell set: $\mathfrak{M}, s_l \nVdash \forall u(u \notin r \vee u \in r)$, so $\mathfrak{M}, s_l \nVdash r \sqsubseteq r$.

Worse even, the material powerset of the empty set is glutty at the greatest state of a **HBST** model, even though the empty set has a consistent membership structure. To see this, consider the set $w := \{u\colon r \in r\}$, where r is the Russell set. For all **HBST**-models $\mathfrak{M}$ with a least state s_l and a greatest state s_g such that $s_g = s_l^*$, for all $a \in D$, $\mathfrak{M}, s_l \nVdash a \in w$, $\mathfrak{M}, s_l \nVdash a \notin w$, hence $\mathfrak{M}, s_g \Vdash a \in w$ and $\mathfrak{M}, s_g \Vdash a \notin w$. By Ψ-Abs, $\mathfrak{M}, s \Vdash u \in \mathcal{P}_{\supset}(\varnothing)$ iff $\mathfrak{M}, s \Vdash \forall z(z \notin u \vee z \in \varnothing)$. At the greatest state s_g, $\mathfrak{M}, s_g \Vdash w \in \mathcal{P}_{\supset}(\varnothing)$, and $\mathfrak{M}, s_g \Vdash w \notin \mathcal{P}_{\supset}(\varnothing)$.

The previous observations show that using the material conditional in instances of abstraction where use of **HYPE**'s $\rightarrow$ is forbidden by Ψ-Abs does not solve the difficulties encountered in the treatment of singletons and powerset in **HBST**, since the material conditional displays unintended behaviour when dealing with gappy and glutty sets. Such unintended behaviour is exacerbated by the fact that states in a given **HBST**-model may have both gaps and gluts.

The minimal model $\mathfrak{M}_{\mathtt{min}}$, however, has only gaps at the least state s_l and only gluts at the greatest state s_g. Furthermore, s_l and s_g are dual, since $s_l = s_g^*$. Then, the behaviour of $\supset$ is fixed at s_l and s_g. A natural question to ask is whether, locally, that is, relatively to s_l or s_g only, we can define a well-behaved notion of powerset.

As shown by Leitgeb [13, Observation 39], indeed, if $\mathfrak{M}$ is a **HYPE**-model with the property of having a least state s_l and a greatest state s_g such that $s_g = s_l^*$, then $s_l \nVdash A \wedge \neg A$ and $s_g \Vdash A \vee \neg A$. Furthermore, the logical consequence relation of **K3** for a language based only on the symbols $\{\wedge, \vee, \neg, \exists, \forall\}$ coincides with the consequence relation for $\rightarrow$-free formulae in **HYPE** defined as truth preservation at s_l. The logical consequence relation of **LP** for a language based on $\{\wedge, \vee, \neg, \exists, \forall\}$ coincides with the consequence relation for $\rightarrow$-free formulae in **HYPE** defined as truth preservation at s_g.

Also, as a consequence of the clauses for conditionals in **HYPE**, in a two-state **HBST**-model where s_l is the least state, s_g is the greatest state and $s_l = s_g^*$, if $s_l \Vdash \varphi \supset \psi$ then $s_l \Vdash \varphi \rightarrow \psi$, and if $s_g \Vdash \varphi \rightarrow \psi$ then $s_g \Vdash \varphi \supset \psi$.

It is easy to see then that the **HBST** results we obtain locally in the least state and the greatest state of $\mathfrak{M}_{\mathtt{min}}$ correspond, in their $\rightarrow$-free fragment, to set theory in **K3** and **LP**.[17]

More precisely, given the relation between **HYPE**'s $\rightarrow$ and $\supset$ in $\mathfrak{M}_{\mathtt{min}}$, **HBST** has the same consequences as set theory in **LP** at the greatest state s_g, while, at s_l, **HBST** expands set theory in **K3** with the global results obtained with the use of **HYPE**'s $\rightarrow$.

Considering the minimal model $\mathfrak{M}_{\mathtt{min}}$, furthermore, makes it possible to exploit the dualities between s_g and s_l to obtain consequences of ExtC. For instance, note that in $\mathfrak{M}_{\mathtt{min}}$, for all $a, b \in D$, and for $s \in \{s_g, s_l\}$, $\mathfrak{M}_{\mathtt{min}}, s \Vdash a =_e b$ or $\mathfrak{M}_{\mathtt{min}}, s \Vdash a \neq_e b$, while this is not generally so in **HBST**.[18]

However, powerset is still not available, because the material subset behaves badly both at s_l and s_g: indeed, it is not reflexive at s_l, due to the presence of gappy sets, and it is non-transitive at s_g, because s_g inherits all the pathologies of **LP** set theory already encountered in Restall [19] and Omori [18]. The failure of powerset deriving from the lack of a good conditional which can feature in formulae satisfying Ψ-Abs or Φ-Abs leads set theory in **HYPE** to an impasse, similar to that encountered by other nonclassical theories with abstraction and extensionality.

5. Open Questions and Future Work

The paper takes some initial steps in understanding abstraction and extensionality principles in the logic **HYPE**. We list some questions that are left open in our investigation.

---

[17]Classic references for set theory in **K3** and **LP** are, respectively, Skolem [21] and Restall [19].

[18]An example of this is that at the minimal model we recover a weakened version of the law of ordered pairs, with pairs defined using material singletons. Indeed, we can show that the law of ordered pairs holds at $\mathfrak{M}_{\mathtt{min}}$ for sets with no gaps, i.e. sets x such that $s_l \Vdash x =_e^{\supset} x$.

First, it would be interesting to obtain clear proof-theoretic reductions between **HBST**, $\mathbf{HBST}_\Phi$ and some classical (weak) set theories. For instance, Proposition 4.7 provides a lower bound for the strength of $\mathbf{HBST}_\Phi$. Can Adjunctive Set Theory be interpreted in $\mathbf{HBST}_\Phi$? What is the logical strength (measured in terms of classical weak set theories) of **HBST**?

Second, it would be interesting to check the status of **HBST** and $\mathbf{HBST}_\Phi$ as *class theories over an ontology of sets*, such as classical ZFC. By the results of [9], a compositional theory of satisfaction over ZFC will be as strong as—in terms of $\mathcal{L}_\in^\rightarrow$-theorems—Kripke-Feferman truth over ZFC. How does the addition of **HYPE**-abstraction for classes over ZFC compare with such compositional theories of satisfaction?

Furthermore, it would be interesting to consider alternative methods to obtain a (fixed-point) interpretation of class membership over a **HYPE** model. For instance, one promising option is to replace **HYPE**-satisfaction (cf. the definition of our valuation function $\mathtt{F}$ in §3) with a supervaluation schema collecting formulae that are not just **HYPE**-true but **HYPE**-"supertrue". The construction could deliver stronger abstraction principles consistent with **HYPE**.

References

[1] Alan Ross Anderson and Nuel D Belnap Jr. First degree entailments. *Mathematische Annalen*, 149(4):302–319, 1963.

[2] Ross Brady. Universal logic. 2006.

[3] Ross T. Brady. The consistency of the axioms of abstraction and extensionality in a three-valued logic. *Notre Dame Journal of Formal Logic*, 12(4):447–453, 1971.

[4] Ross T. Brady. The simple consistency of a set theory based on the logic CSQ. *Notre Dame Journal of Formal Logic*, 24(4):431–449, 1983.

[5] Ross T. Brady. The non-triviality of dialectical set theory. In Graham Priest, Richard Routley, and Jean Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pages 437–470. Philosophia Verlag, 1989.

[6] Solomon Feferman. Toward useful type-free theories. i. *Journal of Symbolic Logic*, 49(1):75–111, 1984.

[7] Solomon Feferman. Reflecting on incompleteness1. *The Journal of Symbolic Logic*, 56(1):1–49, 1991.

[8] Hartry Field, Harvey Lederman, and Tore Fjetland Øgaard. Prospects for a naive theory of classes. 2017.

[9] Martin Fischer, Carlo Nicolai, and Pablo Dopico. Nonclassical truth with classical strength. a proof-theoretic analysis of compositional truth over hype. *Review of Symbolic Logic*, pages 1–24, 2021.

[10] P. C. Gilmore. The consistency of partial set theory without extensionality. In *Axiomatic Set Theory*, volume 13 of *Proceedings of Symposia in Pure Mathematics*, pages 147–153. American Mathematical Society, 1974.

[11] Volker Halbach and Leon Horsten. Axiomatizing kripke's theory of truth. *The Journal of Symbolic Logic*, 71(2):677–712, 2006.

[12] Saul Kripke. Outline of a theory of truth. *The journal of philosophy*, 72(19):690–716, 1975.

[13] Hannes Leitgeb. HYPE: A system of hyperintensional logic. *Journal of Philosophical Logic*, 48(2):305–405, 2019.

[14] G.C. Moisil. *Logique modale*. 1942.

[15] António Monteiro. Sur les algebres de heyting symétriques. *Portugaliae Mathematica*, 39:1–237, 1980.

[16] Edward Nelson. *Predicative Arithmetic*. Princeton University Press, Princeton, N.J., 1986.

[17] Sergei Odintsov and Heinrich Wansing. Routley star and hyperintensionality. *Journal of Philosophical Logic*, 50:33–56, 2021.

[18] Hitoshi Omori. Remarks on naive set theory based on lp. *The Review of Symbolic Logic*, 8(2):279–295, 2015.

[19] Greg Restall. A note on naive set theory in LP. *Notre Dame Journal of Formal Logic*, 33(3):422–432, 1992.

[20] David Ripley. Naive set theory and nontransitive logic. *Review of Symbolic Logic*, 8(3):553–571, 2015.

[21] Thoralf Skolem. A set theory based on a certain 3-valued logic. *Mathematica Scandinavica*, 8(1):127–136, 1960.

[22] Stanislav O Speranski. Negation as a modality in a quantified setting. *Journal of Logic and Computation*, 31(5):1330–1355, 04 2021.

[23] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. 1955.

[24] Zach Weber. *Paradoxes and Inconsistent Mathematics*. Cambridge University Press, 2021.