



**HAL**  
open science

# Trakhtenbrot's Theorem in Coq: Finite Model Theory through the Constructive Lens

Dominik Kirst, Dominique Larchey-Wendling

► **To cite this version:**

Dominik Kirst, Dominique Larchey-Wendling. Trakhtenbrot's Theorem in Coq: Finite Model Theory through the Constructive Lens. *Logical Methods in Computer Science*, 2022, 18 (2), pp.17:1-17:29. 10.46298/lmcs-18(2:17)2022 . hal-03359508

**HAL Id: hal-03359508**

**<https://inria.hal.science/hal-03359508>**

Submitted on 17 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## TRAKHTENBROT'S THEOREM IN COQ: FINITE MODEL THEORY THROUGH THE CONSTRUCTIVE LENS \*

DOMINIK KIRST<sup>a</sup> AND DOMINIQUE LARCHEY-WENDLING<sup>b</sup>

<sup>a</sup> Saarland University, Saarland Informatics Campus, Saarbrücken, Germany  
*e-mail address:* [kirst@cs.uni-saarland.de](mailto:kirst@cs.uni-saarland.de)

<sup>b</sup> Université de Lorraine, CNRS, LORIA, Vandœuvre-lès-Nancy, France  
*e-mail address:* [dominique.larchey-wendling@loria.fr](mailto:dominique.larchey-wendling@loria.fr)

**ABSTRACT.** We study finite first-order satisfiability (FSAT) in the constructive setting of dependent type theory. Employing synthetic accounts of enumerability and decidability, we give a full classification of FSAT depending on the first-order signature of non-logical symbols. On the one hand, our development focuses on Trakhtenbrot's theorem, stating that FSAT is undecidable as soon as the signature contains an at least binary relation symbol. Our proof proceeds by a many-one reduction chain starting from the Post correspondence problem. On the other hand, we establish the decidability of FSAT for monadic first-order logic, i.e. where the signature only contains at most unary function and relation symbols, as well as the enumerability of FSAT for arbitrary enumerable signatures. To showcase an application of Trakhtenbrot's theorem, we continue our reduction chain with a many-one reduction from FSAT to separation logic. All our results are mechanised in the framework of a growing Coq library of synthetic undecidability proofs.

### 1. INTRODUCTION

In the wake of the seminal discoveries concerning the undecidability of first-order logic by Turing and Church in the 1930s, a broad line of work has been pursued to characterise the border between decidable and undecidable fragments of the original decision problem. These fragments can be grouped either by syntactic restrictions controlling the allowed function and relation symbols or the quantifier prefix, or by semantic restrictions on the admitted models (see [BGG97] for a comprehensive description).

Concerning signature restrictions, already predating the undecidability results, Löwenheim had shown in 1915 that monadic first-order logic, admitting only signatures with at most unary symbols, is decidable [Lö15]. Therefore, the successive negative results usually presuppose non-trivial signatures containing an at least binary symbol.

Turning to semantic restrictions, Trakhtenbrot proved in 1950 that, if only admitting finite models, the satisfiability problem over non-trivial signatures is still undecidable [Tra50]. Moreover, the situation is somewhat dual to the unrestricted case, since finite satisfiability (FSAT) is still enumerable while, in the unrestricted case, validity is enumerable. As a consequence, finite validity cannot be characterised by a complete finitary deduction

\* extended version of [KLW20].

system and, resting on finite model theory, various natural problems in database theory and separation logic are undecidable. The latter will be subject of a case study in Section 7.

Conventionally, Trakhtenbrot's theorem is proved by (many-one) reduction from the halting problem for Turing machines (see e.g. [BGG97, Lib10]). An encoding of a given Turing machine  $M$  can be given as a formula  $\varphi_M$  such that the models of  $\varphi_M$  correspond to the runs of  $M$ . Specifically, the finite models of  $\varphi_M$  correspond to terminating runs of  $M$  and so a decision procedure for FSAT of  $\varphi_M$  would be enough to decide whether  $M$  terminates or not.

Although this proof strategy is in principle explainable on paper, already the formal definition of Turing machines, not to mention their encoding in first-order logic, is not ideal for mechanisation in a proof assistant. So for our Coq mechanisation of Trakhtenbrot's theorem, we follow a different strategy by starting from the Post correspondence problem (PCP), a simple matching problem on strings. Similar to the conventional proof, we proceed by encoding every instance  $R$  of PCP as a formula  $\varphi_R$  such that  $R$  admits a solution iff  $\varphi_R$  has a finite model. Employing the framework of synthetic undecidability [FKS19, FLWD<sup>+</sup>20], the computability of  $\varphi_R$  from  $R$  is guaranteed since all functions definable in constructive type theory are computable without reference to a concrete model of computation.

Both the conventional proof relying on Turing machines and our elaboration starting from PCP actually produce formulas in a custom signature well-suited for the encoding of the seed decision problems. The sharper version of Trakhtenbrot's theorem, stating that a signature with at least one binary relation (or one binary function and one unary relation) is enough to turn FSAT undecidable, is in fact left as an exercise in e.g. Libkin's book [Lib10]. However, at least in a constructive setting, this generalisation is non-trivial and led us to mechanising a chain of signature transformations eliminating and compressing function and relation symbols step by step.

The constructive and type-theoretic setting introduces subtleties that remain hidden from view in a classical approach to (finite) model theory. Among these subtleties, quotients are critical for signature reductions but not generally constructively available. With suitable notions of finiteness (here defined as listability) and discreteness (decidable equality), however, we are able to build finite and decidable quotients (Theorem 2.9) sufficient for our purposes. Moreover, the usual set-theoretic constructions in model theory can be simulated in type theory just to some extent, with the prominent lack of a (computable) power set. Fortunately, it is possible to use the notion of weak power set (Lemma 2.11) for these constructions. Relatedly, finiteness *does not entail* computability in the constructive setting, so the Tarski semantics has to be refined. In a critically useful result, we establish that finite satisfiability is not impacted by the further requirement of the discreteness of the model (Theorem 4.7).

Complementing the undecidability result, we further formalise that FSAT is enumerable for enumerable signatures and decidable for monadic signatures. Again, both of these standard results come with their subtleties when explored in a constructive approach to finite model theory.

In summary, the main contributions of this paper are the following:

- we provide an axiom-free Coq mechanisation comprising a full classification of finite satisfiability with regards to the signatures allowed;<sup>1</sup>

---

<sup>1</sup>Downloadable from <http://www.ps.uni-saarland.de/extras/fo1-trakh-ext/> and systematically hyperlinked with the definitions and theorems in this PDF.

- we present a streamlined proof strategy for Trakhtenbrot's theorem well-suited for mechanisation and simple to explain informally, basing on PCP;
- we give a constructive account of signature transformations and the treatment of interpreted equality typically neglected in a classical development;
- compared to the conference version of this paper [KLW20], we contribute a refined analysis of the conditions allowing for decidable FSAT in the case of monadic signatures, introducing the notion of discernability of symbols (e.g. Theorem 6.2);
- additionally to [KLW20] also, we mechanise a many-one reduction from FSAT to the satisfiability problem of separation logic following [CYO01] (e.g. Theorem 7.4);
- finally, we point out that some of the involved proofs that were just sketched in [KLW20] have been expanded (e.g. Theorem 4.7), and reworked for better readability (e.g. Theorem 4.13 and Lemma 5.5).

The paper is structured as follows. We first describe the type-theoretical framework for undecidability proofs and the representation of first-order logic in Section 2. We then outline our variant of Trakhtenbrot's theorem for a custom signature in Section 3. This is followed in Section 4 by a development of enough constructive finite model theory to reach the stronger form of Trakhtenbrot's theorem where the signature is only assumed to contain one at least binary symbol. In Section 5 we switch to decidability results for FSAT over monadic signatures, first assuming that symbols enjoy decidable equality, then maximally strengthening the decidability results to the case of (the weaker notion of) decidable Boolean discernability. In Section 6, we conclude with the precise decidability/undecidability classification of FSAT. Section 7 comprises the case study on the undecidability of separation logic and we end with a brief discussion of the Coq development and future work in Section 8.

## 2. FIRST-ORDER SATISFIABILITY IN CONSTRUCTIVE TYPE THEORY

In order to make this paper accessible to readers unfamiliar with constructive type theory, we outline the required features of Coq's underlying type theory, the synthetic treatment of computability available in constructive mathematics, some properties of finite types, as well as our representation of first-order logic.

**2.1. Basics of Constructive Type Theory.** We work in the framework of a constructive type theory such as the one implemented in Coq, providing a predicative hierarchy of type universes  $\mathbb{T}$  above a single impredicative universe  $\mathbb{P}$  of propositions. On type level, we have the unit type  $\mathbb{1}$  with a single element  $*$  :  $\mathbb{1}$ , the void type  $\mathbb{0}$ , function spaces  $X \rightarrow Y$ , products  $X \times Y$ , sums  $X + Y$ , dependent products  $\forall x : X. F x$ , and dependent sums  $\{x : X \mid F x\}$ . On propositional level, these types are denoted using the usual logical notation ( $\top$ ,  $\perp$ ,  $\rightarrow$ ,  $\wedge$ ,  $\vee$ ,  $\forall$ , and  $\exists$ ).

We employ the basic inductive types of Booleans ( $\mathbb{B} ::= \text{tt} \mid \text{ff}$ ), of Peano natural numbers ( $n : \mathbb{N} ::= 0 \mid 1+n$ ), the option type ( $\mathbb{O} X ::= \ulcorner x \urcorner \mid \emptyset$ ), and lists ( $l : \mathbb{L} X ::= [] \mid x :: l$ ). We write  $|l|$  for the *length* of a list,  $l \# m$  for the *concatenation* of  $l$  and  $m$ ,  $x \in l$  for *membership*, and simply  $f [x_1; \dots; x_n] := [f x_1; \dots; f x_n]$  for the *map* function. We denote by  $X^n$  the type of vectors of length  $n : \mathbb{N}$  and by  $\mathbb{F}_n$  the finite types understood as indices  $\{0, \dots, n-1\}$ . The definitions/notations for lists are shared with vectors  $\vec{v} : X^n$ . Moreover, when  $i : \mathbb{F}_n$  and  $x : X$ , we denote by  $\vec{v}_i$  the  $i$ -th component of  $\vec{v}$  and by  $\vec{v}[x/i]$  the vector  $\vec{v}$  with  $i$ -th component updated to value  $x$ .

**2.2. Synthetic (Un-)decidability.** We review the main ingredients of our synthetic approach to decidability and undecidability [FHS18, FKS19, FLW19, FLWD<sup>+</sup>20, LWF19, SF20], based on the computability of all functions definable in constructive type theory<sup>2</sup> or other constructive foundations of mathematics. We first introduce standard notions of computability theory without referring to a formal model of computation, e.g. Turing machines.

**Definition 2.1.** A problem or predicate  $p : X \rightarrow \mathbb{P}$  is

- **decidable** if there is  $f : X \rightarrow \mathbb{B}$  with  $\forall x. px \leftrightarrow f x = \mathbf{tt}$ .
- **enumerable** if there is  $f : \mathbb{N} \rightarrow \mathbb{O} X$  with  $\forall x. px \leftrightarrow \exists n. f n = \ulcorner x \urcorner$ .

These notions generalise to predicates of higher arity. Moreover, a type  $X$  is

- **enumerable** if there is  $f : \mathbb{N} \rightarrow \mathbb{O} X$  with  $\forall x. \exists n. f n = \ulcorner x \urcorner$ .
- **discrete** if equality on  $X$  (i.e.  $\lambda xy : X. x = y$ ) is decidable.
- a data type if it is both enumerable and discrete.

Using the expressiveness of dependent types, we equivalently tend to establish the decidability of a predicate  $p : X \rightarrow \mathbb{P}$  by giving a function  $\forall x : X. px + \neg px$ . Note that it is common to mechanise decidability results in this synthetic sense (e.g. [BP10, MS15, SST15]). Next, decidability and enumerability transport along reductions:

**Definition 2.2.** A problem  $p : X \rightarrow \mathbb{P}$  (many-one) reduces to  $q : Y \rightarrow \mathbb{P}$ , written  $p \preceq q$ , if there is a function  $f : X \rightarrow Y$  such that  $px \leftrightarrow q(f x)$  for all  $x : X$ .<sup>3</sup>

**Fact 2.3.** Assume  $p : X \rightarrow \mathbb{P}$ ,  $q : Y \rightarrow \mathbb{P}$  and  $p \preceq q$ : (1) if  $q$  is decidable, then so is  $p$  and (2) if  $X$  and  $Y$  are data types and  $q$  is enumerable, then so is  $p$ .

*Proof.* These are Fact 2.11 and Lemma 2.12 in [FKS19]. □

Item (1) implies that we can justify the undecidability of a target problem by reduction from a seed problem known to be undecidable, such as the halting problem for Turing machines. This is in fact the closest rendering of undecidability available in a synthetic setting, since the underlying type theory is consistent with the assumption that every problem is decidable.<sup>4</sup> Nevertheless, we believe that in the intended effective interpretation for synthetic computability, a typical seed problem is indeed undecidable and so are the problems reached by verified reductions. More specifically, since the usual seed problems are not co-enumerable, (2) implies that the reached problems are not co-enumerable either.

Given its simple inductive characterisation involving only basic types of lists and Booleans, the (binary) Post correspondence problem (BPCP) is a well-suited seed problem for compact encoding into first-order logic.

**Definition 2.4.** Given a list  $R : \mathbb{L}(\mathbb{L}\mathbb{B} \times \mathbb{L}\mathbb{B})$  of pairs  $s/t$  of Boolean strings<sup>5</sup>, we define derivability of a pair  $s/t$  from  $R$  (denoted by  $R \triangleright s/t$ ) and solvability (denoted by  $\text{BPCP } R$ ):

$$\frac{s/t \in R}{R \triangleright s/t} \qquad \frac{s/t \in R \quad R \triangleright u/v}{R \triangleright (s \# u)/(t \# v)} \qquad \frac{R \triangleright s/s}{\text{BPCP } R}$$

<sup>2</sup>A result shown and applied for many variants of constructive type theory and which Coq designers are committed to maintain as Coq evolves.

<sup>3</sup>Or equivalently, the **dependently typed characterisation**  $\forall x : X. \{y : Y \mid px \leftrightarrow qy\}$ .

<sup>4</sup>As witnessed by classical set-theoretic models satisfying  $\forall p : \mathbb{P}. p + \neg p$  (cf. [Wer97]).

<sup>5</sup>Notice that the list  $R$  is viewed as a (finite) set of pairs  $s/t \in R$  (hence ignoring the order or duplicates), while  $s$  and  $t$ , which are also lists, are viewed as strings (hence repetitions and ordering matter for  $s$  and  $t$ ).

**Fact 2.5.** *Given a list  $R : \mathbb{L}(\mathbb{L}\mathbb{B} \times \mathbb{L}\mathbb{B})$ , the derivability predicate  $\lambda s t. R \triangleright s/t$  is *decidable*. However, the halting problem for Turing machines reduces to BPCP.*

*Proof.* We give of proof of the decidability of  $R \triangleright s/t$  by induction on  $|s| + |t|$ . We also provide a *trivial proof* of the equivalence of two definitions of BPCP. See [FHS18, FLW19] for details on the reduction from the halting problem to BPCP.  $\square$

It might at first appear surprising that derivability  $\lambda s t. R \triangleright s/t$  is decidable while BPCP is reducible from the halting problem (and hence undecidable). This simply illustrates that undecidability is rooted in the unbounded existential quantifier in the equivalence  $\text{BPCP } R \leftrightarrow \exists s. R \triangleright s/s$ .

In summary, the approach to undecidability used in this and other papers [FKS19, FHS18, FLW19, LWF19, SF20, Dud20, KH21] contributing to the Coq Library of Undecidability Proofs [FLWD<sup>+</sup>20] is to verify (synthetic) many-one reductions from a problem known to be undecidable, rooted by the halting problem for Turing machines. In our case, we start from BPCP as well-suited seed for finite first-order satisfiability, backed by the reduction from Turing machine halting to BPCP verified in [FHS18]. Therefore, our mechanised undecidability results as reported in Sections 6 and 7 are statements of the form  $\text{BPCP} \preceq P$  for problems  $P$  expressing finite first-order satisfiability and satisfiability in separation logic, respectively. Given the constructive setting, these reductions expressed in Coq's type theory may then be interpreted as computable functions, transporting the undecidability (specifically, non-co-enumerability) of BPCP to said problems, along the lines of Fact 2.3. This improves on the ubiquitous pen-and-paper practice to sketch an algorithm and leave its computability implicit in that, by formally defining the algorithm as a Coq term, its computability is guaranteed by its very construction.

The even more explicit alternative would be to resort to a concrete model of computation, e.g. by implementing high-level reduction functions operating on abstract data structures as Turing machines operating over textual or binary encodings of those data-structures. Then the notion of undecidability could be formally bootstrapped by showing that there is no Turing machine deciding the (textually encoded) halting problem, and other problems could be shown undecidable by verifying reductions computable by a Turing machine. However, this approach would require low-level coding in such a model, which in principle can be supported by tools to a certain extent [FK19], but still introduces enormous overhead unnecessary to cope with in a setting providing its own implicit notion of computability.

**2.3. Constructive Finiteness.** We present four tools for manipulating finite types: the *finite pigeon hole principle* (PHP) here established without assuming discreteness, the *well-foundedness* of strict orders over finite types, *quotients* over strongly decidable equivalences that map onto  $\mathbb{F}_n = \{0, \dots, n-1\}$ , and the *weak powerset* finitely enumerating every weakly decidable predicate over a finite type. But first, let us fix a definition of finiteness.

**Definition 2.6.** *A type  $X$  is *finite* if there is a list  $l_X$  s.t.  $\forall x : X. x \in l_X$ , and a predicate  $p : X \rightarrow \mathbb{P}$  is *finite* if there is a list  $l_p$  s.t.  $\forall x. px \leftrightarrow x \in l_p$ .*

Note that in constructive settings there are various alternative characterisations of finiteness<sup>6</sup> (bijection with  $\mathbb{F}_n$  for some  $n$ ; negated infinitude for some definition of infiniteness; etc) and we opted for the above since it is easy to work with while transparently capturing the

<sup>6</sup>And these alternative characterisations are not necessarily constructively equivalent.

expected meaning. One can distinguish *strong* finiteness in  $\mathbb{T}$  (i.e.  $\{l_X : \mathbb{L}X \mid \forall x. x \in l_X\}$ ) from *weak* finiteness in  $\mathbb{P}$  (i.e.  $\exists l_X : \mathbb{L}X. \forall x. x \in l_X$ ), the list  $l_X$  being required computable in the strong case. Strong finiteness implies weak finiteness but the converse holds only in restricted contexts, e.g. inside proofs of propositions in  $\mathbb{P}$ .

For the finite PHP, the typical classical proof requires the discreteness of  $X$  to design transpositions/permutations. Here we avoid discreteness completely, the existence of a duplicate being established without actually computing one.

**Theorem 2.7** (Finite PHP). *Let  $R : X \rightarrow Y \rightarrow \mathbb{P}$  be a binary relation and  $l : \mathbb{L}X$  and  $m : \mathbb{L}Y$  be two lists where  $m$  is shorter than  $l$  ( $|m| < |l|$ ). If  $R$  is total from  $l$  to  $m$  ( $\forall x. x \in l \rightarrow \exists y. y \in m \wedge Rxy$ ) then the values at two distinct positions in  $l$  are related to the same  $y$  in  $m$ , i.e. there exist  $x_1, x_2 \in l$  and  $y \in m$  such that  $l$  has shape  $l = \dots \# x_1 :: \dots \# x_2 :: \dots$  and  $Rx_1y$  and  $Rx_2y$ .*

*Proof.* We start with the case where  $R$  is the identity relation  $=_X$  on  $X$ , hence we want to establish that  $l$  contains a duplicate. We first prove the following **generalised statement**: if  $|m| \leq |l|$  and  $l \subseteq m$  (i.e.  $\forall x. x \in l \rightarrow x \in m$ ) then either  $l$  contains a duplicate or  $l$  and  $m$  are permutable. We establish the generalised statement by structural induction on  $m$ .

In particular, when  $|m| < |l|$  then  $l$  and  $m$  cannot be permutable (because permutations preserve length), hence  **$l$  must contain a duplicate**. Generalizing from  $=_X$  to an arbitrary relation  $R : X \rightarrow Y \rightarrow \mathbb{P}$  is then a simple exercise.  $\square$

Using the PHP, given a strict order<sup>7</sup> over a finite type  $X$ , any descending chain has length bounded by the size of  $X$  as measured by the length of the list enumerating  $X$ .

**Fact 2.8.** *Every strict order on a finite type is well-founded.*

*Proof.* For a constructive proof, one can for instance show that descending chains cannot contain a duplicate (otherwise this would give an impossible cycle in a strict order), hence by the PHP, the length of descending chains is bounded by the length of the enumerating list of the finite type.  $\square$

Coq's type theory does not provide quotients in general (see e.g. [Coh13]) but one can build computable quotients in certain conditions, here for a decidable equivalence relation of which representatives of equivalence classes are listable.

**Theorem 2.9** (Finite decidable quotient). *Let  $\sim : X \rightarrow X \rightarrow \mathbb{P}$  be a decidable equivalence with  $\{l_r : \mathbb{L}X \mid \forall x \exists y. y \in l_r \wedge x \sim y\}$ , i.e. finitely many equivalence classes.<sup>8</sup> Then one can compute the quotient  $X/\sim$  onto  $\mathbb{F}_n$  for some  $n$ , i.e.  $n : \mathbb{N}$ ,  $c : X \rightarrow \mathbb{F}_n$  and  $r : \mathbb{F}_n \rightarrow X$  s.t.  $\forall p. c(rp) = p$  and  $\forall xy. x \sim y \leftrightarrow cx = cy$ .*

*Proof.* From the list  $l_r$  of representatives of equivalence classes, remove duplicate representatives using the strong decidability of  $\sim$ . This gives a list  $l'_r$  which now contains exactly one representative for each equivalence class. Convert  $l'_r$  to a vector  $\vec{v}$ . The function  $r$  (representative) is defined by  $r := \lambda p. \vec{v}_p$ . The function  $c$  (for class) is simple search:  $cx$  is the first (and unique)  $p$  such that  $\vec{v}_p \sim x$ .  $\square$

Using Theorem 2.9 with identity over  $X$  as equivalence, we get bijections between finite, discrete types and the type family  $(\mathbb{F}_n)_{n:\mathbb{N}}$ .<sup>9</sup>

<sup>7</sup>i.e. an irreflexive and transitive binary relation.

<sup>8</sup>Hence  $l_r$  denotes a list of *representatives* of equivalence classes.

<sup>9</sup>For a given  $X$ , the value  $n$  (usually called cardinal) is unique by the PHP.

**Corollary 2.10.** *If  $X$  is a finite and discrete type then one can compute  $n : \mathbb{N}$  and a bijection from  $X$  to  $\mathbb{F}_n$ .*

We conclude this section with the question of the finiteness of powersets. There is no notion of finite powerset in type theory: even for the unit/singleton type  $\mathbb{1}$ , there is no list enumerating the predicates in  $\mathbb{1} \rightarrow \mathbb{P}$ , even up to extensional equivalence.<sup>10</sup> However, there is a notion of *weak powerset*. Recall that a predicate  $p : X \rightarrow \mathbb{P}$  is *weakly decidable* if it satisfies  $\forall x. px \vee \neg px$ , its specialized instance of the Law of Excluded Middle (LEM).<sup>11</sup> Assuming  $X$  is finite, one can compute a list containing all the weakly decidable predicates in  $X \rightarrow \mathbb{P}$ .

**Lemma 2.11** (Weak powerset). *For every finite type  $X$ , one can compute a list  $ll : \mathbb{L}(X \rightarrow \mathbb{P})$  which contains every weakly decidable predicate in  $X \rightarrow \mathbb{P}$  up to extensional equivalence, i.e.  $ll$  satisfies  $\forall p : X \rightarrow \mathbb{P}. (\forall x. px \vee \neg px) \rightarrow \exists q. q \in ll \wedge \forall x. px \leftrightarrow qx$ .*

*Proof.* The list  $ll$  is built by induction on the list  $l_X : \mathbb{L}X$  enumerating  $X$ . If  $l_X$  is  $[\ ]$  then  $X$  is a void type and thus  $ll := (\lambda z. \top) :: [\ ]$  fits. If  $l_X$  is  $x :: l$  then we apply the induction hypothesis to  $l$  and get  $ll$  for the finite sub-type composed of the elements of  $l$  and we define  $ll' := (\lambda p z. x \neq z \wedge pz) ll \# (\lambda p z. x = z \vee pz) ll$ . We check that  $ll'$  contains every weakly decidable predicate over  $x :: l$ . Notice that  $|ll'| = 2|ll|$  in the induction step, hence one could easily show that  $|ll| = 2^{|l_X|}$ , recovering the cardinality of the (classical) powerset.  $\square$

Notice that while the weak powerset contains all weakly decidable predicates, it may contain predicates which are not weakly decidable (unless  $X$  is moreover weakly discrete).

**2.4. Representing First-Order Logic.** We briefly outline our representation of the syntax and semantics of first-order logic in constructive type theory (cf. [FKW21, KH21]). Concerning the *syntax*, we describe terms and formulas as dependent inductive types over a **signature**  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  of function symbols  $f : \mathcal{F}_\Sigma$  and relation symbols  $P : \mathcal{P}_\Sigma$  with arities  $|f|$  and  $|P|$ , using binary connectives  $\dot{\square} \in \{\dot{\rightarrow}, \dot{\wedge}, \dot{\vee}\}$  and quantifiers  $\dot{\nabla} \in \{\dot{\forall}, \dot{\exists}\}$ :

$$\begin{aligned} t : \mathbf{Term}_\Sigma &::= x \mid f \vec{t} & (x : \mathbb{N}, f : \mathcal{F}_\Sigma, \vec{t} : \mathbf{Term}_\Sigma^{|f|}) \\ \varphi, \psi : \mathbf{Form}_\Sigma &::= \dot{\perp} \mid P \vec{t} \mid \varphi \dot{\square} \psi \mid \dot{\nabla} \varphi & (P : \mathcal{P}_\Sigma, \vec{t} : \mathbf{Term}_\Sigma^{|P|}) \end{aligned}$$

Negation is defined as the abbreviation  $\dot{\neg} \varphi := \varphi \dot{\rightarrow} \dot{\perp}$ .

In the chosen de Bruijn representation [dB72], a bound variable is encoded as the number of quantifiers shadowing its binder, e.g.  $\forall x. \exists y. P x u \rightarrow P y v$  may be represented by  $\dot{\forall} \dot{\exists} P 1 4 \dot{\rightarrow} P 0 5$ . The variables  $2 = 4 - 2$  and  $3 = 5 - 2$  in this example are the *free* variables, and variables that do not occur freely are called *fresh*, e.g.  $0$  and  $1$  are fresh. For the sake of legibility, we write concrete formulas with named binders and defer de Bruijn representations to the Coq development. For a formula  $\varphi$  over a signature  $\Sigma$ , we define the list  $\mathbf{FV}(\varphi) : \mathbb{L}\mathbb{N}$  of **free variables**, the list  $\mathcal{F}_\varphi : \mathbb{L}\mathcal{F}_\Sigma$  of **function symbols** and the list  $\mathcal{P}_\varphi : \mathbb{L}\mathcal{P}_\Sigma$  of **relation symbols** that actually occur in  $\varphi$ , all by recursion on  $\varphi$ . We say that  $\varphi$  is closed if it is void of free variables, i.e.  $\mathbf{FV}(\varphi) = [\ ]$ .

Turning to *semantics*, we employ the standard (Tarski-style) model-theoretic semantics, evaluating terms in a given domain and embedding the logical connectives into the constructive meta-logic (cf. [VW96]):

<sup>10</sup>As there is no list extensionally enumerating  $\mathbb{P}$  itself.

<sup>11</sup>The general LEM ( $\forall P : \mathbb{P}. P \vee \neg P$ ) cannot be established constructively.



**Definition 2.12.** A **model**  $\mathcal{M}$  over a domain  $D : \mathbb{T}$  is described by a pair of functions  $\forall f. D^{|f|} \rightarrow D$  and  $\forall P. D^{|P|} \rightarrow \mathbb{P}$  denoted by  $f^{\mathcal{M}}$  and  $P^{\mathcal{M}}$ . Given a variable assignment  $\rho : \mathbb{N} \rightarrow D$ , we recursively extend it to a **term evaluation**  $\hat{\rho} : \text{Term} \rightarrow D$  with  $\hat{\rho}x := \rho x$  and  $\hat{\rho}(f \vec{v}) := f^{\mathcal{M}}(\hat{\rho} \vec{v})$ , and to the **satisfaction** relation  $\mathcal{M} \vDash_{\rho} \varphi$  by

$$\begin{aligned} \mathcal{M} \vDash_{\rho} \perp &:= \perp & \mathcal{M} \vDash_{\rho} \varphi \dot{\square} \psi &:= \mathcal{M} \vDash_{\rho} \varphi \square \mathcal{M} \vDash_{\rho} \psi \\ \mathcal{M} \vDash_{\rho} P \vec{t} &:= P^{\mathcal{M}}(\hat{\rho} \vec{t}) & \mathcal{M} \vDash_{\rho} \dot{\nabla} \varphi &:= \nabla a : D. \mathcal{M} \vDash_{a \cdot \rho} \varphi \end{aligned}$$

where each logical connective  $\dot{\square}/\dot{\nabla}$  is mapped to its meta-level counterpart  $\square/\nabla$  and where we denote by  $a \cdot \rho$  the **de Bruijn extension** of  $\rho$  by  $a$ , defined by  $(a \cdot \rho) 0 := a$  and  $(a \cdot \rho) (1+x) := \rho x$ .<sup>12</sup>

A  $\Sigma$ -*model* is thus a dependent triple  $(D, \mathcal{M}, \rho)$  composed of a domain  $D$ , a model  $\mathcal{M}$  for  $\Sigma$  over  $D$  and an assignment  $\rho : \mathbb{N} \rightarrow D$ . It is *finite* if  $D$  is finite, and *decidable* if  $P^{\mathcal{M}} : D^{|P|} \rightarrow \mathbb{P}$  is decidable for all  $P : \mathcal{P}_{\Sigma}$ .

**Fact 2.13.** *Satisfaction  $\lambda \varphi. \mathcal{M} \vDash_{\rho} \varphi$  is decidable for finite, decidable  $\Sigma$ -models.*

*Proof.* By induction on  $\varphi$ ; finite quantification preserves decidability.  $\square$

In this paper, we are mostly concerned with *finite satisfiability* of formulas. However, since some of the compound reductions hold for more general or more specific notions, we introduce the following variants:

**Definition 2.14** (Satisfiability). *For a formula  $\varphi$  over a signature  $\Sigma$ , we write*

- **SAT**( $\Sigma$ )  $\varphi$  if there is a  $\Sigma$ -model  $(D, \mathcal{M}, \rho)$  such that  $\mathcal{M} \vDash_{\rho} \varphi$ ;
- **FSAT**( $\Sigma$ )  $\varphi$  if additionally  $D$  is finite and  $\mathcal{M}$  is decidable;
- **FSATEQ**( $\Sigma; \equiv$ )  $\varphi$  if the signature contains a distinguished binary relation symbol  $\equiv$  interpreted as equality, i.e.  $x \equiv^{\mathcal{M}} y \leftrightarrow x = y$  for all  $x, y : D$ .

Notice that in a classical treatment of finite model theory, models are supposed to be given *in extension*, i.e. understood as tables providing computational access to functions and relations values. To enable this view in our constructive setting, we restrict to decidable relations in the definition of FSAT, and from now on, *finite satisfiability is always meant to encompass a decidable model*. One could further require the domain  $D$  to be discrete to conform more closely with the classical view; discreteness is in fact enforced by FSATEQ. However, we refrain from this requirement and instead show in Section 4.1 that FSAT and FSAT over discrete models are constructively equivalent.

### 3. TRAKHTENBROT'S THEOREM FOR A CUSTOM SIGNATURE

In this section, we show that BPCP reduces to FSATEQ( $\Sigma_{\text{BPCP}}; \equiv$ ) for the special purpose signature  $\Sigma_{\text{BPCP}} := (\{\star^0, e^0, f_{\text{tt}}^1, f_{\text{ff}}^1\}; \{P^2, \prec^2, \equiv^2\})$ .<sup>13</sup> To this end, we fix an instance  $R : \mathbb{L}(\mathbb{L}\mathbb{B} \times \mathbb{L}\mathbb{B})$  of BPCP (to be understood as a finite set of pairs of Boolean strings) and we construct a formula  $\varphi_R$  such that  $\varphi_R$  is finitely satisfiable if and only if  $R$  has a solution.

Informally, we axiomatise a family  $\mathcal{B}_n$  of models over the domain of Boolean strings of length bounded by  $n$  and let  $\varphi_R$  express that  $R$  has a solution in  $\mathcal{B}_n$ . The axioms express enough equations and inversions of the constructions included in the definition of BPCP such that a solution for  $R$  can be recovered.

<sup>12</sup>The notation  $a \cdot \rho$  illustrates that  $a$  is pushed ahead of the sequence  $\rho_0, \rho_1, \dots$

<sup>13</sup>We use superscripts to succinctly describe the arity of each symbol.

Formally, the symbols in  $\Sigma_{\text{BPCP}}$  are used as follows: the functions  $f_b$  and the constant  $e$  represent  $b :: (\cdot)$  and  $[]$  for the encoding of strings  $s$  as terms  $\bar{s}$  (before a term  $\tau$ , possibly  $e$ ):

$$\overline{[]} \# \tau := \tau \quad \overline{b :: s} \# \tau := f_b(\bar{s} \# \tau) \quad \bar{s} := \bar{s} \# e$$

The constant  $\star$  represents an undefined value for strings too long to be encoded in the finite model  $\mathcal{B}_n$ . The relation  $P$  represents derivability from  $R$  (denoted  $R \triangleright \cdot / \cdot$  here) while  $\prec$  and  $\equiv$  represent strict suffixes and equality, respectively.

Expected properties of the intended interpretation can be captured formally as first-order formulas. First, we ensure that  $P$  is proper (only subject to defined values) and that  $\prec$  is a strict order (irreflexive and transitive):

$$\begin{aligned} \varphi_P &:= \dot{\forall}xy. Pxy \dot{\rightarrow} x \not\equiv \star \wedge y \not\equiv \star & (P \text{ proper}) \\ \varphi_{\prec} &:= (\dot{\forall}x. x \not\prec x) \wedge (\dot{\forall}xyz. x \prec y \dot{\rightarrow} y \prec z \dot{\rightarrow} x \prec z) & (\prec \text{ strict order}) \end{aligned}$$

Next, the image of  $f_b$  is forced disjoint from  $e$  and injective, as long as  $\star$  is not reached. We also ensure that the images of  $f_{\text{tt}}$  and  $f_{\text{ff}}$  intersect only at  $\star$ :

$$\varphi_f := \left( \begin{array}{l} f_{\text{tt}} \star \equiv \star \wedge f_{\text{ff}} \star \equiv \star \\ \dot{\forall}x. f_{\text{tt}} x \not\equiv e \\ \dot{\forall}x. f_{\text{ff}} x \not\equiv e \end{array} \right) \wedge \left( \begin{array}{l} \dot{\forall}xy. f_{\text{tt}} x \not\equiv \star \dot{\rightarrow} f_{\text{tt}} x \equiv f_{\text{tt}} y \dot{\rightarrow} x \equiv y \\ \dot{\forall}xy. f_{\text{ff}} x \not\equiv \star \dot{\rightarrow} f_{\text{ff}} x \equiv f_{\text{ff}} y \dot{\rightarrow} x \equiv y \\ \dot{\forall}xy. f_{\text{tt}} x \equiv f_{\text{ff}} y \dot{\rightarrow} f_{\text{tt}} x \equiv \star \wedge f_{\text{ff}} y \equiv \star \end{array} \right)$$

Furthermore, we enforce that  $P$  simulates  $R \triangleright \cdot / \cdot$ , encoding its inversion principle

$$\varphi_{\triangleright} := \dot{\forall}xy. Pxy \dot{\rightarrow} \bigvee_{s/t \in R} \dot{\forall} \left\{ \begin{array}{l} x \equiv \bar{s} \wedge y \equiv \bar{t} \\ \dot{\exists}uv. Puv \wedge x \equiv \bar{s} \# u \wedge y \equiv \bar{t} \# v \wedge u/v \prec x/y \end{array} \right.$$

where  $u/v \prec x/y$  denotes  $(u \prec x \wedge v \equiv y) \dot{\vee} (v \prec y \wedge u \equiv x) \dot{\vee} (u \prec x \wedge v \prec y)$ . Finally,  $\varphi_R$  is the conjunction of all axioms plus the existence of a solution:

$$\varphi_R := \varphi_P \wedge \varphi_{\prec} \wedge \varphi_f \wedge \varphi_{\triangleright} \wedge \dot{\exists}x. Pxx.$$

**Theorem 3.1.**  $\text{BPCP} \preceq \text{FSATEQ}(\Sigma_{\text{BPCP}}; \equiv)$ .

*Proof.* The reduction  $\lambda R. \varphi_R$  is proved correct by Lemmas 3.2 and 3.3.  $\square$

**Lemma 3.2.**  $\text{BPCP } R \rightarrow \text{FSATEQ}(\Sigma_{\text{BPCP}}; \equiv) \varphi_R$ .

*Proof.* Assume  $R \triangleright s/s$  holds for a string  $s$  with  $|s| = n$ . We show that the model  $\mathcal{B}_n$  over Boolean strings bounded by  $n$  satisfies  $\varphi_R$ . To be more precise, we choose  $D_n := \mathbb{O}\{s : \text{LB} \mid |s| \leq n\}$  as domain, i.e. values in  $D_n$  are either an (overflow) value  $\emptyset$  or a (defined) dependent pair  $\ulcorner (s, H_s) \urcorner$  where  $H_s : |s| \leq n$ . We interpret the function and relation symbols of the chosen signature by

$$\begin{aligned} e^{\mathcal{B}_n} &:= [] & f_b^{\mathcal{B}_n} \emptyset &:= \emptyset & P^{\mathcal{B}_n} st &:= R \triangleright s/t \\ \star^{\mathcal{B}_n} &:= \emptyset & f_b^{\mathcal{B}_n} s &:= \text{if } |s| < n \text{ then } b :: s \text{ else } \emptyset & s \prec^{\mathcal{B}_n} t &:= s \neq t \wedge \exists u. u \# s = t \end{aligned}$$

where we left out some explicit constructors and the edge cases of the relations for better readability, see the [Coq code](#) for full detail. As required,  $\mathcal{B}_n$  interprets  $\equiv$  by equality  $=_{D_n}$ .

Considering the desired properties of  $\mathcal{B}_n$ , first note that  $D_n$  can be shown finite by induction on  $n$ . This however crucially relies on the proof irrelevance of the  $\lambda x. x \leq n$  predicate.<sup>14</sup> The atoms  $s \prec^{\mathcal{B}_n} t$  and  $s \equiv^{\mathcal{B}_n} t$  are decidable by straightforward computations

<sup>14</sup>i.e. that for every  $x : \mathbb{N}$  and  $H, H' : x \leq n$  we have  $H = H'$ . In general, it is not always possible to establish finiteness of  $\{x \mid Px\}$  if  $P$  is not proof irrelevant.

on Boolean strings. Decidability of  $P^{\mathcal{B}_n} s t$  (i.e.  $R \triangleright s/t$ ) was established in Fact 2.5. Finally, since  $\varphi_R$  is a closed formula, any variable assignment  $\rho$  can be chosen to establish that  $\mathcal{B}_n$  satisfies  $\varphi_R$ , for instance  $\rho := \lambda x. \emptyset$ . Then showing  $\mathcal{B}_n \vDash_\rho \varphi_R$  consists of verifying simple properties of the chosen functions and relations, with mostly straightforward proofs.  $\square$

**Lemma 3.3.**  $\text{FSATEQ}(\Sigma_{\text{BPCP}}; \equiv) \varphi_R \rightarrow \text{BPCP } R$ .

*Proof.* Suppose that  $\mathcal{M} \vDash_\rho \varphi_R$  holds for some finite  $\Sigma_{\text{BPCP}}$ -model  $(D, \mathcal{M}, \rho)$  interpreting  $\equiv$  as equality and providing operations  $f_b^{\mathcal{M}}, e^{\mathcal{M}}, \star^{\mathcal{M}}, P^{\mathcal{M}}$  and  $\prec^{\mathcal{M}}$ . Again, the concrete assignment  $\rho$  is irrelevant and  $\mathcal{M} \vDash_\rho \varphi_R$  ensures that the functions/relations behave as specified and that  $P^{\mathcal{M}} x x$  holds for some  $x : D$ .

Instead of trying to show that  $\mathcal{M}$  is isomorphic to some  $\mathcal{B}_n$ , we directly reconstruct a solution for  $R$ , i.e. we find some  $s$  with  $R \triangleright s/s$  from the assumption that  $\mathcal{M} \vDash_\rho \varphi_R$  holds. To this end, we first observe that the relation  $u/v \prec^{\mathcal{M}} x/y$  as defined above is a strict order and thus well-founded as an instance of Fact 2.8.

Now we can show that for all  $x/y$  with  $P^{\mathcal{M}} x y$  there are strings  $s$  and  $t$  with  $x = \bar{s}$ ,  $y = \bar{t}$  and  $R \triangleright s/t$ , by induction on the pair  $x/y$  using the well-foundedness of  $\prec^{\mathcal{M}}$ . So let us assume  $P^{\mathcal{M}} x y$ . Since  $\mathcal{M}$  satisfies  $\varphi_\triangleright$  there are two cases:

- there is  $s/t \in R$  such that  $x = \bar{s}$  and  $y = \bar{t}$ . The claim follows by  $R \triangleright s/t$ ;
- there are  $u, v : D$  with  $P^{\mathcal{M}} u v$  and  $s/t \in R$  such that  $x = \bar{s} \# u$ ,  $y = \bar{t} \# v$ , and  $u/v \prec^{\mathcal{M}} x/y$ . The latter makes the inductive hypothesis applicable for  $P^{\mathcal{M}} u v$ , hence yielding  $R \triangleright s'/t'$  for some strings  $s'$  and  $t'$  corresponding to the encodings  $u$  and  $v$ . This is enough to conclude  $x = \overline{s \# s'}$ ,  $y = \overline{t \# t'}$  and  $R \triangleright (s \# s')/(t \# t')$  as wished.

Applying this fact to the assumed match  $P^{\mathcal{M}} x x$  yields a solution  $R \triangleright s/s$ .  $\square$

#### 4. CONSTRUCTIVE FINITE MODEL THEORY

Combined with Fact 2.5, Theorem 3.1 entails the undecidability (and non-co-enumerability) of FSATEQ over a custom (both finite and discrete) signature  $\Sigma_{\text{BPCP}}$ . By a series of signature reductions, we generalise these results to any signature containing an at least binary relation symbol. In particular, we explain how to reduce FSAT( $\Sigma$ ) to FSAT( $0; \{\in^2\}$ ) for any discrete signature  $\Sigma$ , hence including  $\Sigma_{\text{BPCP}}$ . We also provide a reduction from FSAT( $0; \{\in^2\}$ ) to FSAT( $\{f^n\}; \{P^1\}$ ) for  $n \geq 2$ , which entails the undecidability of FSAT for signatures with one unary relation and an at least binary function. But first, let us show that FSAT is unaltered when further assuming discreteness of the domain.

**4.1. Converting Models to Discrete Ones.** We consider the case of models over a discrete domain  $D$ , i.e. where the equality relation  $=_D := \lambda x y : D. x = y$  is decidable. The question is the following: is FSAT altered when adding this further requirement on models.

Of course, in the case of FSATEQ( $\Sigma; \equiv$ ) the requirement that  $\equiv$  is interpreted as a decidable binary relation which is equivalent to  $=_D$  imposes the discreteness of  $D$ . But in the case of FSAT( $\Sigma$ ) nothing imposes such a restriction on  $D$ . However as we argue below, using Theorem 2.9, we can always quotient  $D$  along a suitable decidable congruence, making the quotient a discrete finite type while preserving first-order satisfaction, from which we deduce that FSAT is unaltered by the discreteness requirement; see Section 4.2.

Let us consider a fixed signature  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$ . In addition, let us fix a finite type  $D$  and a (decidable) model  $\mathcal{M}$  of  $\Sigma$  over  $D$ . Critically, we do not assume the discreteness of  $D$ . We

can conceive an equivalence over  $D$  which is a congruence for all the interpretations of the symbols by  $\mathcal{M}$ , namely *first-order indistinguishability*  $x \stackrel{\circ}{=} y := \forall \varphi \rho. \mathcal{M} \models_{x,\rho} \varphi \leftrightarrow \mathcal{M} \models_{y,\rho} \varphi$ , i.e. first-order semantics in  $\mathcal{M}$  is not impacted when switching  $x$  with  $y$ .

The facts that  $\stackrel{\circ}{=}$  is both an equivalence and a congruence are easy to prove but, with this definition, there is little hope of establishing decidability of  $\stackrel{\circ}{=}$ . The main reason for this is that the signature may contain symbols of infinitely many arities. So we further fix two lists  $l_{\mathcal{F}} : \mathbb{L} \mathcal{F}_{\Sigma}$  and  $l_{\mathcal{P}} : \mathbb{L} \mathcal{P}_{\Sigma}$  of function and relation symbols respectively and restrict the congruence requirement to the symbols in these lists only.

**Definition 4.1** (Bounded first-order indistinguishability). *We say that  $x$  and  $y$  are first-order indistinguishable up to  $l_{\mathcal{F}}/l_{\mathcal{P}}$ , and we write  $x \stackrel{\circ}{=} y$ , if no first-order formula built from the symbols in  $l_{\mathcal{F}}$  and  $l_{\mathcal{P}}$  only can distinguish  $x$  from  $y$ . Formally, this gives:*

$$x \stackrel{\circ}{=} y := \forall \varphi. \mathcal{F}_{\varphi} \subseteq l_{\mathcal{F}} \rightarrow \mathcal{P}_{\varphi} \subseteq l_{\mathcal{P}} \rightarrow \forall \rho : \mathbb{N} \rightarrow D. \mathcal{M} \models_{x,\rho} \varphi \leftrightarrow \mathcal{M} \models_{y,\rho} \varphi.$$

To remain simple, we avoid displaying the dependency on  $l_{\mathcal{F}}$ ,  $l_{\mathcal{P}}$ ,  $D$  and  $\mathcal{M}$  in the notation  $\stackrel{\circ}{=}$  as they remain fixed in this section anyway.

We claim that first-order indistinguishability  $\stackrel{\circ}{=}$  up to  $l_{\mathcal{F}}/l_{\mathcal{P}}$  is a strongly decidable equivalence and a congruence for all the symbols in  $l_{\mathcal{F}}/l_{\mathcal{P}}$ . Remember that congruence (limited to  $l_{\mathcal{F}}/l_{\mathcal{P}}$ ) means commutation with the interpretation of symbols in  $\mathcal{M}$ :

$$\begin{aligned} \forall (f : \mathcal{F}_{\Sigma}) (\vec{v} \vec{w} : D^{|\mathcal{F}|}). f \in l_{\mathcal{F}} \rightarrow (\forall i : \mathbb{F}_{|f|}. \vec{v}_i \stackrel{\circ}{=} \vec{w}_i) \rightarrow f^{\mathcal{M}} \vec{v} \stackrel{\circ}{=} f^{\mathcal{M}} \vec{w} \\ \forall (P : \mathcal{P}_{\Sigma}) (\vec{v} \vec{w} : D^{|\mathcal{P}|}). P \in l_{\mathcal{P}} \rightarrow (\forall i : \mathbb{F}_{|P|}. \vec{v}_i \stackrel{\circ}{=} \vec{w}_i) \rightarrow P^{\mathcal{M}} \vec{v} \leftrightarrow P^{\mathcal{M}} \vec{w}. \end{aligned}$$

We establish the validity of our claim in the following discussion, ending with Theorem 4.5. Equivalence and congruence of  $\stackrel{\circ}{=}$  are easy. However, Definition 4.1 of  $\stackrel{\circ}{=}$  hints at no clue for its decidability. We therefore switch to an alternate definition of  $\stackrel{\circ}{=}$  as a bisimulation.<sup>15</sup> Using Kleene's fixpoint theorem, we would get  $\stackrel{\circ}{=}$  as  $\bigcap_{n < \omega} F^n(\lambda uv. \top)$  for some below defined  $\omega$ -continuous operator  $F$ . Hopefully, we could ensure that only finitely many (as opposed to  $\omega$ ) iterations of the operator  $F$  are needed for the fixpoint to be reached, hence preserving finitary properties such as decidability.

So let us define the operators  $F_{\mathcal{F}}, F_{\mathcal{P}} : (D \rightarrow D \rightarrow \mathbb{P}) \rightarrow (D \rightarrow D \rightarrow \mathbb{P})$  that map a binary relation  $\mathcal{R} : D \rightarrow D \rightarrow \mathbb{P}$  to

$$\begin{aligned} F_{\mathcal{F}}(\mathcal{R}) &:= \lambda(x y : D). \forall f. f \in l_{\mathcal{F}} \rightarrow \forall (\vec{v} : D^{|\mathcal{F}|}) (i : \mathbb{F}_{|f|}). \mathcal{R}(f^{\mathcal{M}} \vec{v}[x/i]) (f^{\mathcal{M}} \vec{v}[y/i]) \\ F_{\mathcal{P}}(\mathcal{R}) &:= \lambda(x y : D). \forall P. P \in l_{\mathcal{P}} \rightarrow \forall (\vec{v} : D^{|\mathcal{P}|}) (i : \mathbb{F}_{|P|}). P^{\mathcal{M}} \vec{v}[x/i] \leftrightarrow P^{\mathcal{M}} \vec{v}[y/i]. \end{aligned}$$

**Fact 4.2.** *The following results hold for the operator  $F_{\mathcal{F}}$  (resp.  $F_{\mathcal{P}}$ ).*

1.  $F_{\mathcal{F}}$  is monotonic, i.e.  $\mathcal{R} \subseteq \mathcal{R}' \rightarrow F_{\mathcal{F}}(\mathcal{R}) \subseteq F_{\mathcal{F}}(\mathcal{R}')$ ;
2.  $F_{\mathcal{F}}$  is  $\omega$ -continuous, i.e.  $\bigcap_n F_{\mathcal{F}}(\mathcal{R}_n) \subseteq F_{\mathcal{F}}(\bigcap_n \mathcal{R}_n)$  with decreasing  $(\mathcal{R}_n)_{n < \omega}$ ;
3.  $F_{\mathcal{F}}$  preserves reflexivity, i.e.  $=_D \subseteq F_{\mathcal{F}}(=_D)$ ;
4.  $F_{\mathcal{F}}$  preserves symmetry, i.e.  $F_{\mathcal{F}}^{-1}(\mathcal{R}) \subseteq F_{\mathcal{F}}(\mathcal{R}^{-1})$ ;
5.  $F_{\mathcal{F}}$  preserves transitivity, i.e.  $F_{\mathcal{F}}(\mathcal{R}) \circ F_{\mathcal{F}}(\mathcal{R}) \subseteq F_{\mathcal{F}}(\mathcal{R} \circ \mathcal{R})$ ;
6.  $F_{\mathcal{F}}$  preserves decidability, i.e. if  $\mathcal{R}$  is decidable then so is  $F_{\mathcal{F}}(\mathcal{R})$ .

Hence the combination  $F(\mathcal{R}) := F_{\mathcal{F}}(\mathcal{R}) \cap F_{\mathcal{P}}(\mathcal{R})$  also preserves these properties.

*Proof.* The proofs of items (1)–(5) are easy, even without assuming boundedness by  $l_{\mathcal{F}}/l_{\mathcal{P}}$ . However, to ensure the preservation of decidability (6), that bound is essential for the quantification over  $f$  in  $l_{\mathcal{F}}$  (resp.  $P$  in  $l_{\mathcal{P}}$ ) in the above definition of  $F_{\mathcal{F}}$  (resp.  $F_{\mathcal{P}}$ ) to stay

<sup>15</sup>That is the greatest fixpoint of an  $\omega$ -continuous operator.

finite. We observe that since  $D$  is finite then so is  $D^{|f|}$  and the remaining quantifications over  $\vec{v} : D^{|f|}$  and  $i : \mathbb{F}_{|f|}$  are finite quantifications again. Hence, all these quantifications behave as finitary conjunctions and thus preserve decidability. Notice that compared to  $F_{\mathcal{F}}$ , the case of  $F_{\mathcal{P}}$  is degenerated because it does not depend in  $\mathcal{R}$ , hence is constant.  $\square$

As a side remark, notice that one can also show that **F preserves first-order definability** where a relation  $\mathcal{R} : D \rightarrow D \rightarrow \mathbb{P}$  is *first-order definable* if there is a formula  $\varphi_{\mathcal{R}}$  built only from  $l_{\mathcal{F}}/l_{\mathcal{P}}$  such that  $\forall \rho. \mathcal{R}(\rho x_0)(\rho x_1) \leftrightarrow \mathcal{M} \vDash_{\rho} \varphi_{\mathcal{R}}$ .

**Theorem 4.3.** *First-order indistinguishability  $\overset{\circ}{\equiv}$  up to  $l_{\mathcal{F}}/l_{\mathcal{P}}$  is extensionally equivalent to  $\equiv_{\mathbb{F}}$  (Kleene's greatest fixpoint of  $F$ ), i.e. for any  $x, y : D$  we have*

$$x \overset{\circ}{\equiv} y \leftrightarrow x \equiv_{\mathbb{F}} y \quad \text{where} \quad x \equiv_{\mathbb{F}} y := \forall n : \mathbb{N}. F^n(\lambda uv. \top) x y.$$

*Proof.* For the  $\rightarrow$  implication, it is enough to show that  $\overset{\circ}{\equiv}$  is a pre-fixpoint of  $F$ , i.e.  $\overset{\circ}{\equiv} \subseteq F(\overset{\circ}{\equiv})$ , and we get this result using **suitable substitutions**. The converse implication  $\leftarrow$  follows from the fact that  $\equiv_{\mathbb{F}}$  is a fixpoint of  $F$ , hence it is a congruence for every symbol in  $l_{\mathcal{F}}/l_{\mathcal{P}}$ , so  $x \equiv_{\mathbb{F}} y$  entails that **formulas built from  $l_{\mathcal{F}}/l_{\mathcal{P}}$  cannot distinguish  $x$  from  $y$** .  $\square$

With  $\equiv_{\mathbb{F}}$ , we now have a more workable characterization of  $\overset{\circ}{\equiv}$  but still no decidability result for it since the quantification over  $n$  in  $\forall n : \mathbb{N}. F^n(\lambda uv. \top) x y$  ranges over the infinite domain  $\mathbb{N}$ . We now establish that the greatest fixpoint is reached after finitely many iterations of  $F$ . Classically one would argue that  $F$  operates over the finite domain of binary relations over  $D$  and since the sequence  $\lambda n. F^n(\lambda uv. \top)$  cannot decrease strictly forever (by the PHP), it must stay constant after at most  $n_0 := 2^{d \times d}$  iterations where  $d := \text{card } D$ . Such reasoning is not constructively acceptable as is, due to the impossibility to build the finite powerset. Fortunately by Fact 4.2 item 6, the iterated values  $F^n(\lambda uv. \top)$  are all decidable, hence belong to the weak powerset. As a consequence, we can apply the finite PHP on the weak powerset.

**Theorem 4.4.** *One can compute  $n : \mathbb{N}$  such that  $\equiv_{\mathbb{F}}$  is equivalent to  $F^n(\lambda uv. \top)$ .*

*Proof.* By a **variant** of Lemma 2.11, we compute the weak powerset of  $D \rightarrow D \rightarrow \mathbb{P}$ ,<sup>16</sup> i.e. a list  $ll$  containing every weakly decidable binary relation over  $D$ , up to extensional equivalence. Since  $\lambda uv. \top : D \rightarrow D \rightarrow \mathbb{P}$  is strongly decidable and  $F$  preserves (both weak and) strong decidability, the sequence  $\lambda n. F^n(\lambda uv. \top)$  is contained in the list  $ll$ , up to extensional equivalence. Hence by Theorem 2.7 (PHP),<sup>17</sup> after  $|ll|$  steps, there must have been a duplicate, i.e. there exists  $a < b \leq |ll|$  such that  $F^a(\lambda uv. \top)$  and  $F^b(\lambda uv. \top)$  are extensionally equivalent. However the values of  $a$  and  $b$  are not computed by the PHP but we can still deduce that  $F^n(\lambda uv. \top)$  must be stalled after  $n = a$ , hence a fortiori after  $n = |ll|$ . It follows that  $F^{|ll|}(\lambda uv. \top)$  is extensionally equivalent to  $\equiv_{\mathbb{F}}$ .  $\square$

We conclude our construction with the main result:  $\overset{\circ}{\equiv}$  is a strongly decidable congruence than can be used to quotient  $\mathcal{M}$  onto a discrete one.

**Theorem 4.5.** *First-order indistinguishability  $\overset{\circ}{\equiv}$  up to  $l_{\mathcal{F}}/l_{\mathcal{P}}$  is a strongly decidable equivalence and a congruence for all the symbols in  $l_{\mathcal{F}}/l_{\mathcal{P}}$ .*

<sup>16</sup>Via  $D \rightarrow D \rightarrow \mathbb{P} \simeq D \times D \rightarrow \mathbb{P}$ , and finiteness of  $D \times D$ .

<sup>17</sup>And here we really need a finite PHP over *non-discrete types*.

*Proof.* Remember that the real difficulty was strong decidability. By Theorem 4.4, the operator  $F$  reaches its fixpoint  $\equiv_F$  after finitely many steps, and by Fact 4.2 item 6,  $F$  preserves decidability, hence by an obvious induction,  $\equiv_F$  is decidable. By Theorem 4.3, the equivalent indistinguishability relation  $\doteq$  is decidable.  $\square$

Back to our side discussion about first-order definability, by Theorems 4.3 and 4.4, there is thus a **first-order formula  $\xi$  which characterises first-order indistinguishability up to  $l_{\mathcal{F}}/l_{\mathcal{P}}$  in  $\mathcal{M}$** , i.e.  $\forall \rho. \rho x_0 \doteq \rho x_1 \leftrightarrow \mathcal{M} \models_{\rho} \xi$ . Since its semantics does not depend on variables other than  $x_0$  and  $x_1$ , one can remap all other variables to e.g.  $x_0$  hence we can even ensure that the first-order formula characterising  $\doteq$  contains only two free variables, namely  $x_0$  and  $x_1$ .

**4.2. Removing Model Discreteness and Interpreted Equality.** We use the strongly decidable congruence  $\doteq$  to quotient models onto discrete ones (in fact  $\mathbb{F}_n$  for some  $n$ ) while preserving first-order satisfaction.

**Definition 4.6.** We write  $\text{FSAT}'(\Sigma) \varphi$  if  $\text{FSAT}(\Sigma) \varphi$  on a discrete model.

**Theorem 4.7.** For every first-order signature  $\Sigma$  and formula  $\varphi$  over  $\Sigma$ , we have  $\text{FSAT}(\Sigma) \varphi$  iff  $\text{FSAT}'(\Sigma) \varphi$ , and as a consequence, both reductions  $\text{FSAT}(\Sigma) \preceq \text{FSAT}'(\Sigma)$  and  $\text{FSAT}'(\Sigma) \preceq \text{FSAT}(\Sigma)$  hold.

*Proof.*  $\text{FSAT}(\Sigma) \varphi$  entails  $\text{FSAT}'(\Sigma) \varphi$  is the non-trivial implication. Hence we consider a finite  $\Sigma$ -model  $(D, \mathcal{M}, \rho)$  of  $\varphi$  and we build a new finite  $\Sigma$ -model of  $\varphi$  which is furthermore discrete. We collect the symbols occurring in  $\varphi$  as the lists  $l_{\mathcal{F}} := \mathcal{F}_{\varphi}$  (for functions) and  $l_{\mathcal{P}} := \mathcal{P}_{\varphi}$  (for relations). By Theorem 4.5, first-order indistinguishability  $\doteq : D \rightarrow D \rightarrow \mathbb{P}$  up to  $\mathcal{F}_{\varphi}/\mathcal{P}_{\varphi}$  is a strongly decidable equivalence over  $D$  and a congruence for the semantics of the symbols occurring in  $\varphi$ . Using Theorem 2.9, we **build the quotient  $D/\doteq$**  on a  $\mathbb{F}_n$  for some  $n : \mathbb{N}$ . We transport the model  $\mathcal{M}$  along this quotient and because  $\doteq$  is a congruence for the symbols in  $\varphi$ , its semantics is preserved along the quotient. Hence,  $\varphi$  has a finite model over the domain  $\mathbb{F}_n$  which is both finite and discrete.  $\square$

**Theorem 4.8.** If  $\equiv$  is a binary relation symbol in the signature  $\Sigma$ , one has a reduction  $\text{FSATEQ}(\Sigma; \equiv) \preceq \text{FSAT}(\Sigma)$ .

*Proof.* Given a list  $l_{\mathcal{F}}$  (resp.  $l_{\mathcal{P}}$ ) of function (resp. relation) symbols such that  $\equiv$  belongs to  $l_{\mathcal{P}}$ , we construct a formula  $\psi(l_{\mathcal{F}}, l_{\mathcal{P}}, \equiv)$  over the function symbols in  $l_{\mathcal{F}}$  and relation symbols in  $l_{\mathcal{P}}$  expressing the requirement that  $\equiv$  is an **equivalence and a congruence** for the symbols in  $l_{\mathcal{F}}/l_{\mathcal{P}}$ . Then we show that  $\lambda \varphi. \varphi \dot{\wedge} \psi(\mathcal{F}_{\varphi}, \equiv :: \mathcal{P}_{\varphi}, \equiv)$  is a correct reduction, where  $\mathcal{F}_{\varphi}$  and  $\mathcal{P}_{\varphi}$  list the symbols occurring in  $\varphi$ .  $\square$

**4.3. From Discrete Signatures to Singleton Signatures.** Let us start by converting a discrete signature to a finite and discrete signature.

**Lemma 4.9.** For any formula  $\varphi$  over a discrete signature  $\Sigma$ , one can compute a signature  $\Sigma_{n,m} = (\mathbb{F}_n; \mathbb{F}_m)$ , arity preserving maps  $\mathbb{F}_n \rightarrow \mathcal{F}_{\Sigma}$  and  $\mathbb{F}_m \rightarrow \mathcal{P}_{\Sigma}$ , and an equi-satisfiable formula  $\psi$  over  $\Sigma_{n,m}$ , i.e.  $\text{FSAT}(\Sigma) \varphi \leftrightarrow \text{FSAT}(\Sigma_{n,m}) \psi$ .

*Proof.* We use the discreteness of  $\Sigma$  and bijectively map the lists of symbols  $\mathcal{F}_{\varphi}$  and  $\mathcal{P}_{\varphi}$  onto  $\mathbb{F}_n$  and  $\mathbb{F}_m$  respectively, using Corollary 2.10. We structurally map  $\varphi$  to  $\psi$  over  $\Sigma_{n,m}$  along this bijection, which preserves finite satisfiability.  $\square$

Notice that  $n$  and  $m$  in the signature  $\Sigma_{n,m}$  depend on  $\varphi$ , hence the above statement cannot be presented as a reduction between (fixed) signatures.

We now erase all function symbols by encoding them with relation symbols. To this end, let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a signature, we set  $\Sigma' := (0; \{\equiv^2\} + \mathcal{F}_\Sigma^{+1} + \mathcal{P}_\Sigma)$  where  $\equiv$  is a new interpreted relation symbol of arity two and in the conversion, function symbols have arity lifted by one, hence the  $\mathcal{F}_\Sigma^{+1}$  notation.

**Lemma 4.10.** *For any finite<sup>18</sup> type of function symbols  $\mathcal{F}_\Sigma$ , one can construct a reduction  $\text{FSAT}'(\mathcal{F}_\Sigma; \mathcal{P}_\Sigma) \preceq \text{FSATEQ}(0; \{\equiv^2\} + \mathcal{F}_\Sigma^{+1} + \mathcal{P}_\Sigma; \equiv^2)$ .*

*Proof.* The idea is to recursively replace a term  $t$  over  $\Sigma$  by a formula which is “equivalent” to  $x \equiv t$  (where  $x$  is a fresh variable not occurring in  $t$ ) and then an atomic formula like e.g.  $P[t_1; t_2]$  by  $\exists x_1 x_2. x_1 \equiv t_1 \wedge x_2 \equiv t_2 \wedge P[x_1; x_2]$ . We complete the encoding with a formula stating that every function symbol  $f : \mathcal{F}_\Sigma$  is encoded into a total functional relation  $P_f : \mathcal{F}_\Sigma^{+1}$  of arity augmented by 1.

Notice that [constructively recovering actual functions from total functional relations](#) is made possible by the finiteness/discreteness of the domain, combined with the decidability of the semantic interpretation of the corresponding relation symbols. In particular, this would fail in the case of SAT with potentially infinite models.  $\square$

Next, assuming that the function symbols have already been erased, we explain how to merge the relation symbols in a signature  $\Sigma = (0; \mathcal{P}_\Sigma)$  into a single relation symbol, provided that there is an upper bound for the arities in  $\mathcal{P}_\Sigma$ .

**Lemma 4.11.** *The reduction  $\text{FSAT}(0; \mathcal{P}_\Sigma) \preceq \text{FSAT}(0; \{Q^{1+n}\})$  holds when  $\mathcal{P}_\Sigma$  is a finite and discrete type of relation symbols and  $|P| \leq n$  holds for all  $P : \mathcal{P}_\Sigma$ .*

*Proof.* This comprises three independent reductions, see Fact 4.12 below.  $\square$

In the following, we denote by  $\mathcal{F}_\Sigma^n$  (resp.  $\mathcal{P}_\Sigma^n$ ) the same type of function (resp. relation) symbols but where the arity is uniformly converted to  $n$ .

**Fact 4.12.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a signature:*

1.  $\text{FSAT}(\mathcal{F}_\Sigma; \mathcal{P}_\Sigma) \preceq \text{FSAT}(\mathcal{F}_\Sigma; \mathcal{P}_\Sigma^n)$  if  $|P| \leq n$  holds for all  $P : \mathcal{P}_\Sigma$ ;
2.  $\text{FSAT}(0; \mathcal{P}_\Sigma^n) \preceq \text{FSAT}(0; \{Q^{1+n}\})$  if  $\mathcal{P}_\Sigma$  is finite;
3.  $\text{FSAT}(\mathcal{F}_\Sigma^0; \mathcal{P}_\Sigma) \preceq \text{FSAT}(0; \mathcal{P}_\Sigma)$  if  $\mathcal{F}_\Sigma$  is discrete.

*Proof.* For the first reduction, every atomic formula of the form  $P \vec{v}$  with  $|\vec{v}| = |P| \leq n$  is converted to  $P \vec{w}$  with  $\vec{w} := \vec{v} \# [x_0; \dots; x_0]$  and  $|\vec{w}| = n$  for an arbitrary term variable  $x_0$ . The rest of the structure of formulas is unchanged.

For the second reduction, we convert every atomic formula  $P \vec{v}$  with  $|\vec{v}| = n$  into  $Q(P::\vec{v})$  where  $P$  now represents a constant symbol ( $Q$  is fixed).

For the last reduction, we replace every constant symbol by a corresponding fresh variable chosen above all the free variables of the transformed formula.  $\square$

<sup>18</sup>In the Coq code, we prove the theorem for finite *or* discrete types of function symbols.

**4.4. Compressing  $n$ -ary Relations to Binary Membership.** Let  $\Sigma_n = (0; \{P^n\})$  be a singleton signature where  $P$  is of arity  $n$ . We now show that  $P$  can be compressed to a binary relation modelling set membership via a construction using hereditarily finite sets [SS16] (useful only when  $n \geq 3$ ).

**Theorem 4.13.**  $\text{FSAT}'(0; \{P^n\}) \preceq \text{FSAT}(0; \{\dot{\in}^2\})$ .

Technically, this reduction is one of the most involved in this work, although in most presentations of Trakhtenbrot's theorem, this is left as an "easy exercise," see e.g. [Lib10]. Maybe it is perceived so because it relies on the encoding of tuples in classical set theory, which is somehow natural for mathematicians, but properly building the finite set model in constructive type theory was not that easy.

Here we only give an overview of the main tools. We encode an arbitrary  $n$ -ary relation  $R : X^n \rightarrow \mathbb{P}$  over a finite and discrete type  $X$  in the theory of *membership* over the signature  $\Sigma_2 = (0; \{\dot{\in}^2\})$ . Membership is much weaker than set theory because the only required set-theoretic axiom is *extensionality*: it suffices to prove the properties of the below described encoding of pairs, ordered pairs and  $n$ -tuples. Two sets are extensionally equal if their members are the same, and we define identity as:

$$x \approx y := \dot{\forall}z. z \dot{\in} x \leftrightarrow z \dot{\in} y$$

Extensionality states that two extensionally equal sets belong to the same sets:

$$\dot{\forall}xy. x \approx y \rightarrow \dot{\forall}z. x \dot{\in} z \rightarrow y \dot{\in} z \quad (4.1)$$

As a consequence of the extensionality axiom (4.1), no first-order formula over  $\Sigma_2$  can distinguish two extensionally equal sets<sup>19</sup> because identity  $\approx$  is a congruence for membership  $\dot{\in}$ . Moreover, establishing the identity  $x \approx y$  of two sets reduces to proving that they have the same elements.

Notice that the language of membership theory (and set theory) does not contain any function symbol, hence, contrary to usual mathematical practices, there is no other way to handle a set than via a characterising formula which makes it a very cumbersome language to work with formally. However, this is how we have to proceed in the Coq development but here, we also use to meta-level terms in the prose for simplicity.

Following Kuratowski, the ordered pair of two sets  $x$  and  $y$  is encoded as  $(x, y) = \{\{x\}, \{x, y\}\}$ . In the first-order theory of membership as implemented in Coq, with terms limited to variables, this means we encode the sentences " $p \dot{=} \{x, y\}$ " and " $p \dot{=} (x, y)$ " by:<sup>20</sup>

$$\begin{aligned} p \dot{=} \{x, y\} &:= \dot{\forall}a. a \dot{\in} p \leftrightarrow a \approx x \dot{\vee} a \approx y \\ p \dot{=} (x, y) &:= \dot{\exists}ab. a \dot{=} \{x, x\} \wedge b \dot{=} \{x, y\} \wedge p \dot{=} \{a, b\} \end{aligned}$$

while the  $n$ -tuple/vector  $(x_1, \dots, x_n)$  is encoded as  $(x_1, (x_2, \dots, x_n))$  recursively. Hence we encode the sentence " $t \dot{=} \vec{v}$ " recursively on  $\vec{v}$  by:

$$t \dot{=} [] := \dot{\forall}a. \dot{\neg} a \dot{\in} t \quad t \dot{=} x :: \vec{v} := \dot{\exists}p. t \dot{=} (x, p) \wedge p \dot{=} \vec{v}$$

Finally we encode "tuple  $\vec{v}$  belongs to  $r$ " as  $\vec{v} \dot{\in} r := \dot{\exists}t. t \dot{=} \vec{v} \wedge t \dot{\in} r$ .

<sup>19</sup>Even if  $x$  and  $y$  have the same elements, this does not imply their identity in every model. However they are identical in the quotient model of Section 4.1, because then  $x$  and  $y$  are first-order indistinguishable.

<sup>20</sup>where the  $\dot{=}$  notation reads as "represents"



We can now describe the reduction function which maps formulas over  $\Sigma_n$  to formulas over  $\Sigma_2$ . We reserve two first-order variables  $d$  (for the domain  $D$ ) and  $r$  (for the relation  $R$ ). We describe the recursive part of the reduction  $\Sigma'_{n \rightsquigarrow 2}$ :

$$\begin{aligned} \Sigma'_{n \rightsquigarrow 2}(P \vec{v}) &:= \vec{v} \in r & \Sigma'_{n \rightsquigarrow 2}(\dot{\forall} z. \varphi) &:= \dot{\forall} z. z \in d \dot{\rightarrow} \Sigma'_{n \rightsquigarrow 2}(\varphi) \\ \Sigma'_{n \rightsquigarrow 2}(\varphi \square \psi) &:= \Sigma'_{n \rightsquigarrow 2}(\varphi) \square \Sigma'_{n \rightsquigarrow 2}(\psi) & \Sigma'_{n \rightsquigarrow 2}(\dot{\exists} z. \varphi) &:= \dot{\exists} z. z \in d \wedge \Sigma'_{n \rightsquigarrow 2}(\varphi) \end{aligned}$$

ignoring the de Bruijn syntax (which would imply adding  $d$  and  $r$  as extra parameters of  $\Sigma'_{n \rightsquigarrow 2}$ ). Notice that we have to prevent  $d$  and  $r$  from occurring freely in  $\varphi$ . In addition, for given  $\varphi$  we set:

$$\begin{aligned} \varphi_1 &:= \dot{\exists} z. z \in d & \text{i.e. } d \text{ is non-empty;} \\ \varphi_2 &:= x_1 \in d \wedge \dots \wedge x_k \in d & \text{where } [x_1; \dots; x_k] = \text{FV}(\varphi). \end{aligned}$$

This gives us the reduction function  $\Sigma_{n \rightsquigarrow 2}(\varphi) := \varphi_1 \wedge \varphi_2 \wedge \Sigma'_{n \rightsquigarrow 2}(\varphi)$ .

The *completeness* of the reduction  $\Sigma_{n \rightsquigarrow 2}$  is the easy part. Given a finite model of  $\Sigma_{n \rightsquigarrow 2}(\varphi)$  over  $\Sigma_2$ , we recover a model of  $\varphi$  over  $\Sigma_n$  by selecting as the new domain the members of  $d$ ,<sup>21</sup> and the interpretation of  $P \vec{v}$  is given by testing whether the  $n$ -tuple  $\vec{v}$  belongs to  $r$ .

The *soundness* of the reduction  $\Sigma_{n \rightsquigarrow 2}$  is the formally involved part, with Theorem 4.14 below containing the key construction.

**Theorem 4.14.** *Given a decidable  $n$ -ary relation  $R : X^n \rightarrow \mathbb{P}$  over a finite, discrete and inhabited type  $X$ , one can compute a finite type  $Y$  equipped with a decidable relation  $\in : Y \rightarrow Y \rightarrow \mathbb{P}$ , two distinguished elements  $d, r : Y$  and a pair of maps  $i : X \rightarrow Y$  and  $s : Y \rightarrow X$  such that:*

1. for any  $x : X$ ,  $i x \in d$  holds;
2. for any  $y \in d$ , there exists  $x$  such that  $y = i x$ ;
3. for any  $\vec{v} : X^n$ ,  $R \vec{v}$  holds iff “tuple  $i(\vec{v})$  belongs to  $r$ ”

*Proof.* We give a brief outline of this proof, referring to the [Coq code](#) for details. The type  $Y$  is built from the type of hereditarily finite sets based on [SS16], and when we use the word “set” below, it means hereditarily finite set. There, sets are finitely branching ordered trees (encoded as binary trees in the standard way), considered up to permutation and contraction equivalence. By [totally ordering those trees](#), a [normal form](#) can be computed to give an effective representative for every equivalence class, implementing the quotient on this infinite type of trees in Coq, which gives us [hereditarily finite sets](#).

Basing on (hereditarily finite) sets, the idea behind the construction of  $Y$  is first to construct  $d$  as a [transitive set](#)<sup>22</sup> of which the elements are in bijection  $i/s$  with the type  $X$ , hence  $d$  is the [cardinal](#) of  $X$  in the set-theoretic meaning. Then the iterated powersets  $\mathcal{P}(d), \mathcal{P}^2(d), \dots, \mathcal{P}^k(d)$  are all transitive sets as well and contain  $d$  both as a member and as a subset. Considering  $\mathcal{P}^{2n}(d)$  which contains all the  $n$ -tuples built from the members of  $d$ , we define  $r$  as the set of  $n$ -tuples collecting the encodings  $i(\vec{v})$  of vectors  $\vec{v} : X^n$  such that  $R \vec{v}$ . We show  $r \in p$  for  $p$  defined as  $p := \mathcal{P}^{2n+1}(d)$ . Using the Boolean counterpart of  $(\cdot) \in p$  for unicity of proofs, we then define  $Y := \{z \mid z \in p\}$ , restrict membership  $\in$  to  $Y$  and this gives the finite type equipped with all the required properties. Notice that the decidability requirement for  $\in$  holds constructively because we work with hereditarily finite sets, and would not hold with arbitrary sets.  $\square$

<sup>21</sup>using the Boolean counterpart of the decidable membership predicate to get a finite dependent sum.

<sup>22</sup>Recall that the members of a transitive set are also subsets of it.

Notice that while the axiom of extensionality for membership (4.1) is absent from the reduction function  $\Sigma_{n \rightsquigarrow 2}$ , it is however satisfied in the model of hereditary finite sets we build to prove Theorem 4.14, and instrumental in the set-theoretic encoding of the  $n$ -ary relation implemented there.

**4.5. Summary: From Discrete Signatures to the Binary Signature.** Combining all of the previous results, we give a reduction from any discrete signature to the binary singleton signature.

**Theorem 4.15.**  $\text{FSAT}(\Sigma) \preceq \text{FSAT}(\mathbb{0}; \{P^2\})$  holds for any discrete signature  $\Sigma$ .

*Proof.* Let us first consider the case of  $\Sigma_{n,m} = (\mathbb{F}_n; \mathbb{F}_m)$ , a signature over the finite and discrete types  $\mathbb{F}_n$  and  $\mathbb{F}_m$ . Then we have a reduction  $\text{FSAT}(\mathbb{F}_n; \mathbb{F}_m) \preceq \text{FSAT}(\mathbb{0}; \{P^2\})$  by combining Theorems 4.7, 4.8, and 4.13 and Lemmas 4.10 and 4.11.

Let us denote by  $f_{n,m}$  the reduction  $\text{FSAT}(\mathbb{F}_n; \mathbb{F}_m) \preceq \text{FSAT}(\mathbb{0}; \{P^2\})$ . Let us now consider a fixed discrete signature  $\Sigma$ . For a formula  $\varphi$  over  $\Sigma$ , using Lemma 4.9, we compute a signature  $\Sigma_{n,m}$  and formula  $\psi_\varphi$  over  $\Sigma_{n,m}$  s.t.  $\text{FSAT}(\Sigma) \varphi \leftrightarrow \text{FSAT}(\mathbb{F}_n; \mathbb{F}_m) \psi$ . The map  $\lambda\varphi.f_{n,m} \psi_\varphi$  is the required reduction.  $\square$

The binary signature can be further reduced to a signature with a non-monadic function.

**Lemma 4.16.**  $\text{FSAT}(\mathbb{0}; \{P^2\}) \preceq \text{FSAT}(\{f^n\}; \{Q^1\})$  when  $n \geq 2$ .

*Proof.* We encode the binary relation  $\lambda x y. P[x; y]$  with  $\lambda x y. Q(f[x; y; \dots])$ , using the first two parameters of  $f$  to encode pairing. But since we need to change the domain of the model, we also use a fresh variable  $d$  to encode the domain as  $\lambda x. Q(f[d; x; \dots])$  and we **restrict all quantifications to the domain** similarly to the encoding  $\Sigma'_{n \rightsquigarrow 2}$  of Section 4.4.  $\square$

We finish the reduction chains with the weakest possible signature constraints. The following reductions have straightforward proofs.

**Fact 4.17.** One has reductions for the three statements below (for  $n \geq 2$ ):

1.  $\text{FSAT}(\mathbb{0}; \{P^2\}) \preceq \text{FSAT}(\mathbb{0}; \{P^n\})$ ;
2.  $\text{FSAT}(\mathbb{0}; \{P^n\}) \preceq \text{FSAT}(\Sigma)$  if  $\Sigma$  contains an  $n$ -ary relation (symbol);
3.  $\text{FSAT}(\{f^n\}; \{Q^1\}) \preceq \text{FSAT}(\Sigma)$  if  $\Sigma$  contains an  $n$ -ary function and a unary relation.

## 5. DECIDABILITY RESULTS

Complementing the previously studied negative results, we now examine the conditions allowing for decidable satisfiability problems. Since any binary symbol renders finite satisfiability undecidable (see upcoming Theorem 6.3), we only need to consider monadic signatures where all arities are below one. First we consider the case of discrete mononadic signatures, and then give the most general characterisation of decidability of finite satisfiability, by showing (in the monadic case) that it is equivalent to the decidability of the Boolean discernability of function and relation symbols.

### 5.1. Discrete Monadic Signatures.

**Definition 5.1** (Monadic). *A first-order signature is monadic if the arities of all symbols are either 0 or 1, functions and relations alike. A first-order signature is degenerate/propositional if the arities of all relation symbols is 0.*

Notice that when relation symbols have arity 0, then no term can occur in first-order formulas, hence the degenerate or propositional qualifier.

**Lemma 5.2** (FSAT over a fixed domain). *Given a discrete signature  $\Sigma$  and a discrete and finite type  $D$ , one can decide whether or not a formula over  $\Sigma$  has a (finite) model over domain  $D$ .*

*Proof.* By Fact 2.13, satisfaction in a given finite model is decidable. It is also [invariant under extensional equivalence](#), so we only need to show that there are [finitely many \(decidable\) models](#) over  $D$  up to extensional equivalence.<sup>23</sup>  $\square$

**Lemma 5.3.** *A formula over a signature  $\Sigma$  has a finite and discrete model if and only if it has a (finite) model over  $\mathbb{F}_n$  for some  $n : \mathbb{N}$ .*

*Proof.* If  $\varphi$  has a model over a discrete and finite domain  $D$ , by Corollary 2.10, one can bijectively map  $D$  to  $\mathbb{F}_n$  and transport the model along this bijection.  $\square$

**Lemma 5.4.** *FSAT( $0; \mathcal{P}_\Sigma$ ) is decidable if  $\mathcal{P}_\Sigma$  is discrete with uniform arity 1.*

*Proof.* By Lemma 4.9, we can assume  $\mathcal{P}_\Sigma = \mathbb{F}_n$  for some  $n : \mathbb{N}$  w.l.o.g. We show that if  $\varphi$  has a finite model then it must have [a model over the domain  \$\{\vec{v} : \mathbb{B}^n \mid b \vec{v} = \mathbf{tt}\}\$](#)  for some Boolean subset  $b : \mathbb{B}^n \rightarrow \mathbb{B}$ . Up to extensional equivalence, there are only finitely many such subsets  $b$  and we conclude with Lemma 5.2.  $\square$

The following Lemma 5.5 explains how to remove all function symbols but proceeds entirely inside monadic signatures. Contrary to Lemma 4.10 that encodes  $n$ -ary functions with  $(1+n)$ -ary (functional) relations, here we cannot increase the arity by one because we intend the target signature to be monadic also.

**Lemma 5.5.** *Let  $n : \mathbb{N}$ , and  $\mathcal{P}_\Sigma$  be a finite type of relation symbols. For signatures of uniform arity 1, we have a reduction  $\text{FSAT}(\mathbb{F}_n; \mathcal{P}_\Sigma) \preceq \text{FSAT}(0; (\mathbb{L}\mathbb{F}_n \times \mathcal{P}_\Sigma) + \mathcal{P}_\Sigma)$ .*

*Proof.* We implement a proof somewhat inspired by that of Proposition 6.2.7 (Grädel) in [BGG97, pp. 251] but the invariant suggested in the iterative process described there did not work out formally and we had to proceed in a single conversion step instead, switching from single symbols to lists of function symbols.

Given a formula potentially containing function symbols in the finite type  $\mathbb{F}_n$ , the problem is to remove those functions symbols while preserving FSAT. In the monadic signature  $(\mathbb{F}_n; \mathcal{P}_\Sigma)$  (with uniform arity 1), an atomic formula is necessarily of the form  $P(f_1(\dots f_q(x) \dots))$  where  $P : \mathcal{P}_\Sigma$ ,  $[f_1; \dots; f_q] : \mathbb{L}\mathbb{F}_n$  is a list of function symbols, and  $x : \mathbb{N}$  is a variable. The removal of function symbols proceeds in two steps: in a first phase, by adding more predicate symbols, we reduce to the case where logical atoms are of the form  $P(f(x))$  or  $P(x)$ , i.e. at most one nested function application; and in a second phase, we remove the function symbols by skolemisation.

<sup>23</sup>Without discreteness of  $\Sigma$ , it is impossible to build the list of models over  $D = \mathbb{B}$ .

In the first step, we assume an upper bound  $m$  over the number  $q$  of successive applications to be found in an atomic formula, e.g.  $P(f_1(\dots f_q(x)\dots))$ . This number  $m$  is *computed globally* from the start formula to be converted, e.g. by choosing the **maximum depth of nesting** of function symbols occurring in it. For disambiguation, we denote the relation symbols of the target signature either by  $P_r$  with  $r : \mathcal{P}_\Sigma$  (those that originate in  $\mathcal{P}_\Sigma$ ) or newly the introduced ones by  $Q_{w,r}$  with  $(w,r) : \mathbb{L}\mathbb{F}_n \times \mathcal{P}_\Sigma$ .

To the atom  $P_r(f_1(\dots f_q(x)\dots))$  we **associate** the new atom  $Q_{[f_q;\dots;f_1],r}(x)$ . To ensure the correctness of this encoding, we add the equivalences

$$Q_{[],r}(x) \leftrightarrow P_r(x) \quad \text{and} \quad Q_{f::w,r}(x) \leftrightarrow Q_{w,r}(f(x))$$

all universally quantified over  $x$ . When satisfied, these equivalences uniquely characterise the interpretation of the predicates  $Q_{w,r}$  and ensure that the replacing atom  $Q_{[f_q;\dots;f_1],r}(x)$  has the same interpretation as the original  $P_r(f_1(\dots f_q(x)\dots))$ .

Notice however that there are infinitely many words  $w$  in  $\mathbb{L}\mathbb{F}_n$  (as soon as  $n > 0$ ), and as a consequence, infinitely many instances of  $Q_{f::w,r}(x) \leftrightarrow Q_{w,r}(f(x))$ . As these would not all be embeddable a single first-order formula, we use the bound  $m$  to limit the equivalences to the instances where  $|w| < m$ , of which there are indeed only finitely many.

In the second step (skolemisation), we simply convert the equivalences

$$Q_{f::w,r}(x) \leftrightarrow Q_{w,r}(f(x)) \quad \text{into} \quad Q_{f::w,r}(x) \leftrightarrow Q_{w,r}(x_f)$$

where the  $x_f$  are  $n$  new skolem variables (as many as in the type  $\mathbb{F}_n$ ). To preserve FSAT, we need to existentially quantify over the skolem variables  $x_0, \dots, x_{n-1}$ , these existential quantifications  $\exists x_0 \dots \exists x_{n-1}$  occurring above the conjunction of all the equations  $Q_{f::w,r}(x) \leftrightarrow Q_{w,r}(x_f)$ . On top of this, we keep universal quantification over  $x$ .  $\square$

If functions or relations have arity 0, one can always lift them to arity 1 by filling the hole with an arbitrary term, like in Fact 4.12, item (1).

**Fact 5.6.** *The reduction  $\text{FSAT}(\mathcal{F}_\Sigma; \mathcal{P}_\Sigma) \preceq \text{FSAT}(\mathcal{F}_\Sigma^1; \mathcal{P}_\Sigma^1)$  holds when all arities in  $\Sigma$  are at most 1, where  $\mathcal{F}_\Sigma^1$  and  $\mathcal{P}_\Sigma^1$  denote arities uniformly updated to 1.*

**Fact 5.7.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma^0)$  be a signature where the arities of all relation symbols is 0. Then  $\text{FSAT}(\Sigma)$  and  $\text{FSAT}(0; \mathcal{P}_\Sigma^0)$  are inter-reducible, i.e. we have both  $\text{FSAT}(\Sigma) \preceq \text{FSAT}(0; \mathcal{P}_\Sigma^0)$  and  $\text{FSAT}(0; \mathcal{P}_\Sigma^0) \preceq \text{FSAT}(\Sigma)$ .*

*Proof.* If all relation symbols are constants (arity 0), then we contemplate a degenerate/propositional fragment of first-order logic. Indeed, regardless of  $\mathcal{F}_\Sigma$ , no term can occur in formulas, hence neither can function symbols. Since relations are the same with the same arity 0 in both signatures, there are obvious FSAT preserving structural transformations (back and forth) of formulas of these degenerate first-order signatures.  $\square$

**Theorem 5.8.**  *$\text{FSAT}(\Sigma)$  is decidable if  $\Sigma$  is discrete with arities less or equal than 1, or if all relation symbols have arity 0.*

*Proof.* If all arities are at most 1, then by Fact 5.6, we can assume  $\Sigma$  of uniform arity 1. Therefore, for a formula  $\varphi$  over  $\Sigma$  with uniform arity 1, we need to decide FSAT for  $\varphi$ . By Theorem 4.9, we can compute a signature  $\Sigma_{n,m} = (\mathbb{F}_n; \mathbb{F}_m)$  and a formula  $\psi$  over  $\Sigma_{n,m}$  equi-satisfiable with  $\varphi$ . Using the reduction of Lemma 5.5, we compute a formula  $\gamma$ , equi-satisfiable with  $\psi$ , over a discrete signature of uniform arity 1, void of functions. We decide the satisfiability of  $\gamma$  by Lemma 5.4.

On the other hand, if all relation symbols have arity 0, we use the reduction  $\text{FSAT}(\Sigma) \preceq \text{FSAT}(0, \mathcal{P}_\Sigma^0)$  of Fact 5.7 and we are back to the previous case.  $\square$

**5.2. Discernability and Finite Satisfiability.** We say that two members  $x$  and  $y$  of a type  $X$  (of e.g. symbols) are *Boolean discernable* if there is a Boolean valued map  $\delta : X \rightarrow \mathbb{B}$  giving them two different values, i.e.  $\delta x \neq \delta y$ . In the sequel, we simply say discernable or undiscernable for conciseness. We however recall the reader not to confuse (Boolean) undiscernability with Leibniz's understanding of undiscernability<sup>24</sup> which agrees with "equality" in the case of Coq. Formally this gives the definition:

**Definition 5.9** (Discernability). *Given a type  $X$  and two terms  $x, y : X$ , we say that  $x$  and  $y$  are **discernable** and write  $x \not\equiv_d y$  if  $\exists \delta : X \rightarrow \mathbb{B}. \delta x \neq \delta y$ . On the other hand, we say that  $x$  and  $y$  are **undiscernable** and we write  $x \equiv_d y$  if  $\forall \delta : X \rightarrow \mathbb{B}. \delta x = \delta y$ .*

**Equivalently**,  $x$  and  $y$  are discernable if they can be mapped to two different values of a discrete type. Hence  $\delta$  discerns  $x$  from  $y$  and the decision algorithm for identity between  $\delta x$  and  $\delta y$  eventually produces a proof of discernability. Because  $\mathbb{B}$  is discrete, undiscernability is the negation of discernability, i.e.  $\forall xy : X. x \equiv_d y \leftrightarrow \neg(x \not\equiv_d y)$ .

Undiscernability is an equivalence relation generally weaker than equality but the two notions match on discrete types. As we explain below, it is the proper notion to capture when two symbols of a first-order signature cannot be interpreted differently by any finite first-order model. The reason is that those models are inherently discrete: by Theorem 4.7, the interpretation of function symbols can be restricted to discrete models; and relation symbols are always interpreted by decidable relations.

**Fact 5.10.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a signature such that  $\text{FSAT}(\Sigma)$  is decidable. Then for any two relation symbols  $P$  and  $Q$  in  $\mathcal{P}_\Sigma$ , it is decidable whether  $P$  and  $Q$  are discernable.*

*Proof.* The formula  $P(\dots) \wedge \neg Q(\dots)$  is finitely satisfiable if and only if  $P \not\equiv_d Q$ . The dots  $(\dots)$  are arbitrary terms filling the mandatory arguments of  $P$  and  $Q$ .  $\square$

**Fact 5.11.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a signature with  $P : \mathcal{P}_\Sigma$ , a relation symbol of arity 1. If  $\text{FSAT}(\Sigma)$  is decidable, then for any two function symbols  $f$  and  $g$  in  $\mathcal{F}_\Sigma$ , it is decidable whether  $f$  and  $g$  are discernable.*

*Proof.* The formula  $P(f(\dots)) \wedge \neg P(g(\dots))$  is finitely satisfiable if and only if  $f \not\equiv_d g$ .  $\square$

We restrict the discussion that follows to monadic signatures only because otherwise FSAT is undecidable (see upcoming Theorem 6.3). Hence function and relation symbols have arity either 0 or 1. If a monadic signature contains a unary relation symbol, then decidability of FSAT entails decidability of discernability of both function symbols and relation symbols. On the other hand, if a monadic signature contains no unary relation (degenerate case), then all relations are constant (zero-ary) and in this case we can only obtain an algorithm for discerning relation symbols.

**Lemma 5.12.** *Let  $\Sigma^1 = (\mathcal{F}_\Sigma^1; \mathcal{P}_\Sigma^1)$  be a signature with uniform arity 1 such that both  $\mathcal{F}_\Sigma$  and  $\mathcal{P}_\Sigma$  have decidable discernability. For any formula  $\varphi$  over  $\Sigma^1$  one can compute a discrete signature  $\Sigma_d^1$  of uniform arity 1 and a formula  $\psi$  over  $\Sigma_d^1$  such that  $\varphi$  and  $\psi$  are equi-satisfiable, i.e.  $\text{FSAT}(\Sigma^1) \varphi \leftrightarrow \text{FSAT}(\Sigma_d^1) \psi$ .*

<sup>24</sup> $x$  is Leibniz-undiscernable from  $y$  if for any predicate  $P : X \rightarrow \mathbb{P}$ ,  $P(x)$  implies  $P(y)$ .

*Proof.* Let  $S$  be a type of symbols that has decidable discernability (either  $\mathcal{F}_\Sigma$  or  $\mathcal{P}_\Sigma$ ). Consequently, [undiscernability is also decidable](#). Since undiscernable symbols cannot be distinguished by first-order models, we would like to quotient along undiscernability but it is not always possible to do so constructively.<sup>25</sup> However, using a variant of Theorem 2.9, given a list  $l_S : \mathbb{L}S$  of symbols, we [compute](#) a discrete type<sup>26</sup>  $D$  and a map  $\delta : S \rightarrow D$  such that

$$\forall r.s. r \in l_S \rightarrow s \in l_S \rightarrow (r \equiv_d s \leftrightarrow \delta r = \delta s).$$

In a way,  $\delta$  is able to discern symbols simultaneously, but only those in  $l_S$ . We can now use  $\delta$  to project the symbols of  $S$  on a discrete type while preserving FSAT for formulas that contain symbols in  $l_S$  only.

Given a start formula  $\varphi$  in the signature  $\Sigma^1$  (of uniform arity 1) with decidable discernability of symbols, we compute the function and relation symbols occurring in  $\varphi$  and use those two lists to compute two discrete types  $\mathcal{F}_d, \mathcal{P}_d$  and two maps  $\delta_{\mathcal{F}} : \mathcal{F}_\Sigma \rightarrow \mathcal{F}_d, \delta_{\mathcal{P}} : \mathcal{P}_\Sigma \rightarrow \mathcal{P}_d$  identifying the undiscernable symbols occurring in  $\varphi$ . Using  $\delta_{\mathcal{F}}$  and  $\delta_{\mathcal{P}}$ , we map the formula  $\varphi$  on a formula  $\psi$  in the discrete signature  $\Sigma_d^1 = (\mathcal{P}_d^1; \mathcal{F}_d^1)$  (of uniform arity 1 also) while preserving FSAT, i.e.  $\text{FSAT}(\Sigma^1) \varphi \leftrightarrow \text{FSAT}(\Sigma_d^1) \psi$ . Notice that the target signature depends on (the symbols occurring in)  $\varphi$ .  $\square$

**Theorem 5.13.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a monadic signature (all arities are less than 1) such that both  $\mathcal{F}_\Sigma$  and  $\mathcal{P}_\Sigma$  have decidable discernability. Then  $\text{FSAT}(\Sigma)$  is decidable.*

*Proof.* By Fact 5.6, to show the  $\text{FSAT}(\Sigma)$  is decidable we may assume that  $\Sigma$  is of uniform arity 1, i.e. of the form  $\Sigma^1 = (\mathcal{F}_\Sigma^1; \mathcal{P}_\Sigma^1)$ . Given  $\varphi$  over  $\Sigma^1$ , we compute an equi-satisfiable  $\psi$  over the discrete signature  $\Sigma_d^1$  with Lemma 5.12, and then decide  $\text{FSAT}(\Sigma_d^1) \psi$  using Theorem 5.8.  $\square$

**Theorem 5.14.** *Let  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  be a propositional monadic signature (all arities of relation symbols are 0) such that  $\mathcal{P}_\Sigma$  has decidable discernability. Then  $\text{FSAT}(\Sigma)$  is decidable.*

*Proof.* We combine the reduction  $\text{FSAT}(\Sigma) \preceq \text{FSAT}(0; \mathcal{P}_\Sigma^0)$  of Fact 5.7 and Theorem 5.13.  $\square$

## 6. FINAL SIGNATURE CLASSIFICATION

We conclude with the exact classification of FSAT regarding enumerability, decidability, and undecidability depending on the properties of the signature.

**Theorem 6.1.** *Given  $\Sigma = (\mathcal{F}_\Sigma; \mathcal{P}_\Sigma)$  where both  $\mathcal{F}_\Sigma$  and  $\mathcal{P}_\Sigma$  are data types, the finite satisfiability problem for formulas over  $\Sigma$  is enumerable.*

*Proof.* Using Theorem 4.7 and Lemmas 5.2 and 5.3, one constructs a predicate  $Q : \mathbb{N} \rightarrow \text{Form}_\Sigma \rightarrow \mathbb{B}$  s.t.  $\text{FSAT}(\Sigma) \varphi \leftrightarrow \exists n. Q n \varphi = \text{tt}$ . Then, it is easy to build a [computable enumeration](#)  $e : \mathbb{N} \rightarrow \mathbb{O} \text{Form}_\Sigma$  of  $\text{FSAT}(\Sigma) : \text{Form}_\Sigma \rightarrow \mathbb{P}$ .  $\square$

**Theorem 6.2** (Full Monadic FOL).  *$\text{FSAT}(\Sigma)$  is decidable if either (a) or (b) holds:*

- (a) :  $\Sigma$  is monadic and has decidable discernability for both function and relation symbols;
- (b) : relation symbols in  $\Sigma$  have all arity 0 together with decidable discernability.

<sup>25</sup>e.g. if there is no available enumeration of  $\equiv_d$ -equivalence classes.

<sup>26</sup>Notice that  $D$  is also finite by construction but this property is not needed below.

This simply recollects Theorems 5.13 and 5.14. However, we point out that in the case of monadic signatures, because of Facts 5.10 and 5.11, it is not possible to weaken further the above assumptions (a) or (b) to entail the decidability of  $\text{FSAT}(\Sigma)$ . And for non-monadic signatures, we have the general statement of Trakhtenbrot's theorem:

**Theorem 6.3** (Full Trakhtenbrot). *BPCP reduces to  $\text{FSAT}(\Sigma)$  if either (a) or (b) holds:*

- (a) :  $\Sigma$  contains either an at least binary relation symbol;
- (b) :  $\Sigma$  contains a unary relation symbol together with an at least binary function symbol.

*Proof.* By Theorems 3.1, 4.8 and 4.15, Lemma 4.16, and Fact 4.17.  $\square$

According to the explanation provided in Section 2.2 we may (informally) observe that this verified reduction yields undecidability:

**Observation 6.4.** *For a signature  $\Sigma$  satisfying the conditions (a) or (b) of Theorem 6.3,  $\text{FSAT}(\Sigma)$  is undecidable and, more specifically, not co-enumerable.*

Notice that even if the conditions on arities of Theorems 6.2 and 6.3 fully classify signatures with decidable discernability, *a fortiori* discrete signatures, it is not possible to decide whether a signature is monadic or not, unless the signature is furthermore finite. For a given formula  $\varphi$  though, it is always possible to render it in the finite signature of used symbols.

## 7. APPLICATION: UNDECIDABILITY OF SEPARATION LOGIC

Trakhtenbrot's theorem can be used to establish several negative results concerning first-order logic and other decision problems. Regarding first-order logic, a direct consequence is that the completeness theorem fails in the sense that no (enumerable) deduction system can capture finite validity, as this would turn  $\text{FSAT}(\Sigma)$  co-enumerable in contradiction to Observation 6.4. Examples of decision problems shown undecidable by trivial reduction from  $\text{FSAT}(\Sigma)$  are the finite satisfiability problem in the first-order theory of graphs and problems such as query containment and query equivalence in data base theory. This section is devoted to a still simple but insightful application of Trakhtenbrot's theorem, namely the undecidability of separation logic.

Separation logic [Rey02, IO01] as an assertion language for finite data structures bears an obvious connection to finitely interpreted first-order logic. In particular the formulation in [CYO01] adding pointers to *binary* heap cells ( $t \mapsto t_1, t_2$ ) to the spatial operations  $\varphi * \psi$  and  $\varphi \multimap \psi$  for separating conjunction and implication, as well as **emp** for the emptiness assertion in extension of the pure first-order language, admits a compact reduction from the  $\text{FSAT}$  problem over the *binary* signature. In this section, we outline our adaptation of the undecidability proof given in [CYO01] to our constructive and type-theoretic setting, with a focus on the technical details concerning discreteness and decidability induced by this approach.

We represent the syntax of separation logic as an inductive type **SL** of formulas  $\varphi, \psi$  by

$$\varphi, \psi : \mathbf{SL} ::= (t \mapsto t_1, t_2) \mid \mathbf{emp} \mid \varphi * \psi \mid \varphi \multimap \psi \mid t_1 \equiv t_2 \mid \dot{\perp} \mid \varphi \dot{\square} \psi \mid \dot{\nabla} \varphi \quad (t : \mathbb{O}\mathbb{N})$$

with  $\dot{\square} \in \{\dot{\rightarrow}, \dot{\wedge}, \dot{\vee}\}$  and  $\dot{\nabla} \in \{\dot{\forall}, \dot{\exists}\}$  as in **Form** and isolate a minimal fragment **MSL** by

$$\varphi, \psi : \mathbf{MSL} ::= (t \mapsto t_1, t_2) \mid \dot{\perp} \mid \varphi \dot{\square} \psi \mid \dot{\nabla} \varphi \quad (t : \mathbb{O}\mathbb{N}).$$

Informally, the assertions expressed by SL and MSL are interpreted over a memory model consisting of a finite heap addressing binary cells and a stack mapping variables to addresses. The first-order fragment is interpreted as expected, where quantification ranges over addresses. The pointer  $(t \mapsto t_1, t_2)$  is interpreted strictly as the assertion that the heap consists of a single cell containing the pair denoted by  $t_1$  and  $t_2$  referenced at  $t$ , while  $(t \hookrightarrow t_1, t_2)$  just asserts that the heap contains such a pair.

**Definition 7.1.** *Given a stack  $s : \mathbb{N} \rightarrow \mathbb{V}$  mapping variables to possibly invalid addresses  $\mathbb{V} := \mathbb{O}\mathbb{N}$  and a heap  $h : \mathbb{L}(\mathbb{N} \times (\mathbb{V} \times \mathbb{V}))$  representing a finite map of valid addresses to pairs of addresses, we define the satisfaction relation  $h \models_s \varphi$  for  $\varphi : \mathbf{SL}$  recursively by*

$$\begin{aligned} h \models_s (t \mapsto t_1, t_2) &:= \exists a. \hat{s}t = \ulcorner a \urcorner \wedge h = [(a, (\hat{s}t_1, \hat{s}t_2))] & h \models_s t_1 \equiv t_2 &:= \hat{s}t_1 = \hat{s}t_2 \\ h \models_s \text{emp} &:= h = [] & h \models_s \perp &:= \perp \\ h \models_s \varphi * \psi &:= \exists h_1 h_2. h \approx h_1 \# h_2 \wedge h_1 \models_s \varphi \wedge h_2 \models_s \psi & h \models_s \varphi \dot{\square} \psi &:= h \models_s \varphi \square h \models_s \psi \\ h \models_s \varphi \dot{*} \psi &:= \forall h'. h \# h' \rightarrow h' \models_s \varphi \rightarrow h \# h' \models_s \psi & h \models_s \dot{\nabla} \varphi &:= \nabla v : \mathbb{V}. h \models_{v.s} \varphi \end{aligned}$$

where  $\hat{s}\ulcorner x \urcorner := s x$  and  $\hat{s}\emptyset := \emptyset$ , where  $h \approx h'$  denotes equivalence  $\forall a p. (a, p) \in h \leftrightarrow (a, p) \in h'$  and  $h \# h'$  denotes disjointness  $\neg \exists a p p'. (a, p) \in h \wedge (a, p') \in h'$ , and where for the first-order fragment each logical connective  $\dot{\square}/\dot{\nabla}$  is mapped to its meta-level counterpart  $\square/\nabla$ .

For  $\varphi : \mathbf{MSL}$  the relation  $h \models_s \varphi$  is obtained by the same rules with additionally

$$h \models_s (t \hookrightarrow t_1, t_2) := \exists a. \hat{s}t = \ulcorner a \urcorner \wedge (a, (\hat{s}t_1, \hat{s}t_2)) \in h$$

and we define the satisfiability problem **SLSAT** (**MSLSAT**) on  $\varphi : \mathbf{SL}$  ( $\varphi : \mathbf{MSL}$ ) as the existence of a stack  $s$  and a functional heap  $h$  (i.e.  $\forall a p p'. (a, p) \in h \rightarrow (a, p') \in h \rightarrow p = p'$ ) such that  $h \models_s \varphi$ .

The outline of the following reduction is to first establish  $\mathbf{FSAT} \preceq \mathbf{MSLSAT}$  to emphasise that already the fragment MSL is undecidable and then continue with  $\mathbf{MSLSAT} \preceq \mathbf{SLSAT}$  by a mere syntax embedding. The idea for the main reduction is to encode the binary relation  $P[x; y] : \mathbf{Form}$  on the heap by  $(a \hookrightarrow x, y) : \mathbf{MSL}$  at some address  $a$  while tracking the domain elements  $x$  via empty cells  $(x \hookrightarrow \emptyset, \emptyset)$ . Since the intermediate model transformations require computational access to the domain, we actually need to base the main reduction on  $\mathbf{FSAT}'$  guaranteeing discreteness.

Formally, we translate first-order formulas  $\varphi : \mathbf{Form}$  over the binary signature  $(0; \{P^2\})$  to formulas  $\overline{\varphi} : \mathbf{MSL}$  in the sufficient fragment of separation logic by

$$\begin{aligned} \overline{P[x; y]} &:= (\dot{\exists}z. (\ulcorner z \urcorner \hookrightarrow \ulcorner x \urcorner, \ulcorner y \urcorner)) \dot{\wedge} (\ulcorner x \urcorner \hookrightarrow \emptyset, \emptyset) \dot{\wedge} (\ulcorner y \urcorner \hookrightarrow \emptyset, \emptyset) \\ \overline{\forall x. \varphi} &:= \dot{\forall}x. (\ulcorner x \urcorner \hookrightarrow \emptyset, \emptyset) \dot{\rightarrow} \overline{\varphi} & \overline{\dot{\exists}x. \varphi} &:= \dot{\exists}x. (\ulcorner x \urcorner \hookrightarrow \emptyset, \emptyset) \dot{\wedge} \overline{\varphi} \end{aligned}$$

and by recursively descending through the remaining logical operations. The next two lemmas verify the correctness of the reduction function.

**Lemma 7.2.** *Given a model  $\mathcal{M}$  over a discrete and finite domain  $D$  exhausted by a duplicate-free list  $l_D$  and coming with a decider  $f_P$  for  $P^{\mathcal{M}}$ , one can compute a functional heap  $h$  and from every environment  $\rho : \mathbb{N} \rightarrow D$  a stack  $s_\rho$  such that  $\mathcal{M} \models_\rho \varphi \leftrightarrow h \models_{s_\rho} \overline{\varphi}$  for all  $\rho$ .*

*Proof.* We encode domain elements  $d : D$  as natural numbers by the unique index  $n_d < |l_D|$  at which  $d$  occurs in  $l_D$  and pairs  $(d, e) : D \times D$  as numbers  $n_{d,e} \geq |l_D|$  by  $n_{d,e} := \pi(n_d, n_e) + |l_D|$  employing an injective pairing function  $\pi : (\mathbb{N} \times \mathbb{N}) \rightarrow \mathbb{N}$ . We then construct a heap  $h$  encoding both the full domain  $D$  and the binary relation  $P^{\mathcal{M}}$  by

$$h := [(n_d, (\emptyset, \emptyset)) \mid d \in l_D] \# [(n_{d,e}, (\ulcorner n_d \urcorner, \ulcorner n_e \urcorner)) \mid f_P d e = \text{tt}]$$



which is functional by the injectivity of the encodings  $n_d$  and  $n_{d,e}$  that are taken care not to overlap by the addition of  $|l_D|$  in the definition of  $n_{d,e}$ .

Given an environment  $\rho$ , we convert it to a stack  $s_\rho x := \ulcorner n_{\rho x} \urcorner$  and prove the claimed equivalence by induction on  $\varphi$  with  $\rho$  generalised. We only discuss the case  $\varphi = P[x; y]$ . Assuming  $\mathcal{M} \models_\rho P[x; y]$  we have  $P^{\mathcal{M}}[d; e]$  and so  $f_P d e = \mathbf{tt}$  for  $d := \rho x$  and  $e := \rho y$ . Therefore  $(n_{d,e}, (\ulcorner n_d \urcorner, \ulcorner n_e \urcorner)) \in h$ . Since by construction also  $(n_d, (\emptyset, \emptyset)) \in h$  and  $(n_e, (\emptyset, \emptyset)) \in h$ , we can conclude  $h \models_{s_\rho} \overline{P[x; y]}$ . Conversely, given  $h \models_{s_\rho} \overline{P[x; y]}$ , we know that  $(a, (\ulcorner n_d \urcorner, \ulcorner n_e \urcorner)) \in h$  at some address  $a$  and hence can deduce that  $f_P d e = \mathbf{tt}$ . Thus  $\mathcal{M} \models_\rho P[x; y]$ .  $\square$

**Lemma 7.3.** *Given a heap  $h$  containing at least one element of the form  $(a_0, (\emptyset, \emptyset))$ , one can compute a decidable model  $\mathcal{M}$  over a discrete and finite domain  $D$  and from every stack  $s$  an environment  $\rho_s : \mathbb{N} \rightarrow D$  such that  $\mathcal{M} \models_{\rho_s} \varphi \leftrightarrow h \models_s \overline{\varphi}$  for all stacks  $s$  that satisfy the condition  $\forall x \in \mathbf{FV}(\varphi). \exists a. s x = \ulcorner a \urcorner \wedge (a, (\emptyset, \emptyset)) \in h$ .*

*Proof.* As domain  $D$  we take the type of all addresses  $a$  such that  $(a, (\emptyset, \emptyset)) \in h$ , formally defined as  $D := \{a \mid (a, (\emptyset, \emptyset)) \in h\}$  using the Boolean counterpart of list membership for unicity of proofs. By this construction, elements  $a$  and  $a'$  of  $D$  are equal iff they are equal as addresses in  $\mathbb{N}$  and so  $D$  is discrete and, since it is bounded by  $h$ , also finite. To turn  $D$  into a model  $\mathcal{M}$ , we set

$$P^{\mathcal{M}}[a_1; a_2] := \exists a. (a, (\ulcorner a_1 \urcorner, \ulcorner a_2 \urcorner)) \in h$$

which is decidable since it expresses a bounded quantification over a list.

Given the dummy element  $a_0$  of  $D$ , we can convert a stack  $s$  into an environment  $\rho_s$  mapping  $x$  to  $a$  if  $s x = \ulcorner a \urcorner$  with  $(a, (\emptyset, \emptyset)) \in h$ , and to  $a_0$  in any other case. With these constructions in place, the claim is established by induction on  $\varphi$  with  $s$  generalised, we again just discuss the case  $\varphi = P[x; y]$ . First suppose  $\mathcal{M} \models_{\rho_s} P[x; y]$ , then since  $x, y \in \mathbf{FV}(P[x; y])$  we know that  $\rho_s$  is well-defined on  $x, y$  by the condition on  $s$  and obtain corresponding  $a_1, a_2 : D$ . From  $P^{\mathcal{M}}[a_1; a_2]$  we obtain  $a$  with  $(a, (\ulcorner a_1 \urcorner, \ulcorner a_2 \urcorner)) \in h$  and therefore conclude  $h \models_s \overline{P[x; y]}$ . Conversely, starting with  $h \models_s \overline{P[x; y]}$  straightforwardly yields  $\mathcal{M} \models_{\rho_s} P[x; y]$ .  $\square$

Since Lemma 7.3 imposes a condition on the free variables of the input formula, it is convenient to start from the restriction  $\mathbf{FSAT}'_c$  of  $\mathbf{FSAT}'$  to closed formulas  $\varphi$ . To connect  $\mathbf{FSAT}'_c$  to BPCP, one could verify that the reduction chain  $\mathbf{BPCP} \preceq \mathbf{FSAT}'$  given in the previous sections actually yields closed formulas, hence witnessing  $\mathbf{BPCP} \preceq \mathbf{FSAT}'_c$  directly. However, we prefer to implement the general (and straightforward) reduction  $\mathbf{FSAT}' \preceq \mathbf{FSAT}'_c$ . The following theorem then summarises the three parts comprising the reduction  $\mathbf{FSAT}' \preceq \mathbf{SLSAT}$ :

**Theorem 7.4.** *We have reductions as follows:*

1.  $\mathbf{FSAT}'(\Sigma) \preceq \mathbf{FSAT}'_c(\Sigma)$
2.  $\mathbf{FSAT}'_c(0; \{P^2\}) \preceq \mathbf{MSLSAT}$
3.  $\mathbf{MSLSAT} \preceq \mathbf{SLSAT}$

*Proof.* We establish each reduction separately.

1. Given a formula  $\varphi$  over signature  $\Sigma$ , we first compute a bound  $N$  such that  $x < N$  for all  $x \in \mathbf{FV}(\varphi)$ . Then it is easy to verify that  $\varphi' := \exists^N \varphi$  prefixing  $\varphi$  with  $N$  existential quantifiers is closed and has a (finite and discrete) model iff  $\varphi$  does.

2. Given a closed formula  $\varphi$  over the binary signature  $(0; \{P^2\})$ , we define  $\varphi' : \text{MSL}$  by

$$\varphi' := (\dot{\exists}(\ulcorner 0^\top \leftrightarrow \emptyset, \emptyset)) \dot{\wedge} \bar{\varphi}$$

and show that  $\varphi$  has a finite and discrete model iff  $\varphi'$  is MSL-satisfiable.

First, if  $\mathcal{M} \models_\rho \varphi$  over a finite and discrete domain  $D$ , we can apply Lemma 7.2 to obtain  $h \models_{s_\rho} \bar{\varphi}$  since over discrete types any list can be turned duplicate-free. Moreover,  $D$  at least contains the element  $d := \rho 0$  and hence by construction of  $h$  we have  $(n_d, (\emptyset, \emptyset)) \in h$ , establishing the guard  $\dot{\exists}(\ulcorner 0^\top \leftrightarrow \emptyset, \emptyset)$ . So in total  $h \models_{s_\rho} \varphi'$ .

Secondly, from  $h \models_s \varphi'$  we obtain  $h \models_s \bar{\varphi}$  and some  $a_0$  with  $(a_0, (\emptyset, \emptyset)) \in h$ . Since  $\varphi$  is closed, the condition in Lemma 7.3 holds vacuously and thus we obtain  $\mathcal{M} \models_{\rho_s} \varphi$ .

3. We embed MSL into SL by the map sending the sole deviating assertion  $(t \leftrightarrow t_1, t_2)$  to  $(t \mapsto t_1, t_2) * \dot{\top}$  where  $\dot{\top} := \dot{\perp} \dot{\rightarrow} \dot{\perp}$ . To verify this reduction, it suffices to establish that  $h \models_s (t \leftrightarrow t_1, t_2)$  iff  $h \models_s (t \mapsto t_1, t_2) * \dot{\top}$ , which follows by straightforward list manipulation.  $\square$

**Corollary 7.5.** *FSAT'( $\Sigma$ ) reduces to SLSAT, therefore also BPCP reduces to SLSAT.*

*Proof.* By composing the three parts of Theorem 7.4 with Theorems 6.3 and 4.7.  $\square$

As in the case of FSAT and backed by the explanation in Section 2.2, we may interpret this reduction as an undecidability result:

**Observation 7.6.** *SLSAT is undecidable and, more specifically, not co-enumerable.*

In comparison to [CYO01], our reduction is formulated for satisfiability problems instead of the dual validity problems. However, this change is inessential since the models are transformed pointwise as visible in Lemma 7.2 and Lemma 7.3 and so the only consequence is a flipped quantifier in the proof of (2) of Theorem 7.4. More importantly, the formal setting forced us to be more explicit about the handling of addresses, in particular the encoding of a given finite first-order interpretation. For instance, the way chosen in Lemma 7.2 to start with an abstract domain  $D$  and encode both elements and pairs over  $D$  as numbers is not the only alternative but allowed us to maintain the explicit representation of the address space as  $\mathbb{N}$ . Moreover, our syntax fragments differ slightly since we don't need equality in MSL as it is not a primitive in Form but on the other hand keep all logical connectives and not just the classically sufficient base  $\dot{\rightarrow}, \dot{\forall}, \dot{\perp}$  as this is (in the general case) constructively insufficient.

We end this section with the remark that the reduction given in [CYO01] and adapted here of course crucially relies on the binary pointers  $(t \mapsto t_1, t_2)$  as a language primitive. As discussed in [BDL12], with a less explicit memory structure, the considered fragment of separation logic is decidable and only turns undecidable on addition of separating implication.

## 8. DISCUSSION

**8.1. Code overview, implementation, and design choices.** The main part of our Coq development directly concerned with the classification of finite satisfiability consists of a bit more than 10k loc, in addition to 3k loc of (partly reused) utility libraries. Most of the code comprises the signature transformations, with more than 4k loc for reducing discrete signatures to membership. Comparatively, the initial reduction from BPCP to

FSATEQ( $\Sigma_{\text{BPCF}}$ ) takes less than 500 loc. The application to the undecidability of separation logic also amounts to roughly 500 loc.

Our mechanisation of first-order logic in principle follows previous developments [FKS19, FKW20, FKW21] but also differs in a few aspects that were picked up by follow-up work [KH21]. Notably, we had to separate function from relation signatures to be able to express distinct signatures that agree on one sort of symbols computationally. This mostly avoids wandering in “setoid hell,” i.e. the painful manipulation of cumbersome type castings on formulæ, of which the type *depends* on signatures. Moreover, we found it favorable to abstract over the logical connectives in form of  $\dot{\square}$  and  $\dot{\nabla}$  to shorten purely structural definitions and proofs. Most of the signature reductions proceed in such structural approaches so this is a real relief for propositional connectives. For existential quantifiers  $\dot{\exists}$  and universal quantifiers  $\dot{\forall}$ , it is however less frequent that they can be managed in a unified way.

The quantifiers  $\dot{\exists}$  and  $\dot{\forall}$  are also binders, and to deal with these binders, we used standard unscoped de Bruijn syntax. Notice that it is much easier to implement de Bruijn for first-order logic than it is for higher-order logics or even just lambda calculus, the reason being that (first-order) terms used for substitutions do not contain binders. It could have been possible to use the Autosubst 2 [SSK19] support for de Bruijn syntax, but we refrained from this choice because of its current (technical but not fundamental) dependency on the axiom of functional extensionality. We remind the reader that in the context of synthetic undecidability, axioms cannot be freely added without risking breaking the requisite of computability of the terms constructed in the underlying type theory.

So far, we discussed technical implementation choices that have little or no impact on the meaning or provability of the reduction results that we implement in this paper. Other choices are clearly possible, but some could lead to considerably more complicated proofs, e.g. setoid hell if dependent types are managed too naively.

At the other end are choices that could/would impact the provability of our reduction results, and paramount to them, the notions involved in the definition of what is a finite model of first-order logic. In classical settings, there is usually no discussion on how to interpret function symbols: as set-theoretic functions in the model, i.e. total and functional binary relations. However in a constructive setting, already the notion of finiteness has several non-equivalent implementations. We choose to define finiteness by basing on the inductive type of lists, used to enumerate members of finite types. Hence finiteness of  $D$  just requires that the terms of type  $D$  can be collected into a list, i.e. there is  $l : \mathbb{L}D$  such that  $\forall x : D. x \in l$ .

We could have required the domain  $D$  of the model to also be discrete (i.e. equipped with a computable way to discriminate elements which are not identical), but we devoted Section 4.1 to establish that this requirement is not necessary and does not impact the reduction results (see Theorem 4.7). While finiteness implies discreteness in a purely predicative setting like the one of Agda where  $\in : D \rightarrow \mathbb{L}D \rightarrow \mathbb{T}$ , it does not imply discreteness when the membership predicate  $\in : D \rightarrow \mathbb{L}D \rightarrow \mathbb{P}$  is typed in the impredicative sort  $\mathbb{P}$ .<sup>27</sup>

This illustrates that finiteness alone does not imply computability in the impredicative constructive setting of Coq. Even on the finite type  $\mathbb{F}_n = \{0, \dots, n-1\}$ , relations  $\mathbb{F}_n \rightarrow \mathbb{P}$  are not necessarily decidable. We choose to assume that the intended meaning of a finite

<sup>27</sup>because the membership predicate typed  $\in : D \rightarrow \mathbb{L}D \rightarrow \mathbb{P}$  cannot be used to computationally recover the position of an element in a list, whereas it can when typed as  $\in : D \rightarrow \mathbb{L}D \rightarrow \mathbb{T}$ .

model of first-order logic is that it can be described *in extension* by tables. Hence, we found appropriate to interpret a function symbol  $f$  by a term of type  $D^{|f|} \rightarrow D$ , that if applied to the list enumerating  $D$  and then reduced on each instance, would yield the intended table. For relation symbols  $P$ , we used a term  $P^{\mathcal{M}}$  of type  $D^{|f|} \rightarrow \mathbb{P}$ , *moreover assuming that  $P^{\mathcal{M}}$  is decidable* (or equivalently Boolean). This ensures that not only functions but also relations can be described by tables.

However, dropping the decidability of  $P^{\mathcal{M}}$ , or else interpreting function symbols by functional and weakly total<sup>28</sup> relations of type  $D^{|f|} \rightarrow D \rightarrow \mathbb{P}$ , would certainly break some of our reductions, again because of the impredicativity of sort  $\mathbb{P}$ . The question of whether this one relaxation or other relaxations would still be acceptable renderings of the notion of finite model in constructive type theory could be debated informally, but formally they could change the meaning of Trakhtenbrot's theorem up to a breaking point.

**8.2. Future work.** We refrained from additional axioms since we included our development in the growing Coq library of synthetic undecidability proofs [FLWD<sup>+</sup>20]. In this context, we plan to generalise some of the intermediate signature reductions so that they become reusable for other undecidability proofs concerning first-order logic over arbitrary models. For these general reductions however, additional assumptions like unique choice for the encoding of functions as relations or the requirements for a model of set theory as described in [KS19] to compress to the binary signature will play a role.

As further future directions, we want to explore whether a more direct reduction into the binary signature can be given, circumventing the intermediate signature transformations. Also possible, though rather ambitious, would be to mechanise the classification of first-order satisfiability with regards to the quantifier prefix as comprehensively developed in [BGG97].

## REFERENCES

- [BDL12] Rémi Brochenin, Stéphane Demri, and Etienne Lozes. On the almighty wand. *Information and Computation*, 211:106–137, 2012. doi:<https://doi.org/10.1016/j.ic.2011.12.003>.
- [BGG97] Egon Börger, Erich Grädel, and Yuri Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic. Springer-Verlag Berlin Heidelberg, 1997.
- [BP10] Thomas Braibant and Damien Pous. An efficient Coq tactic for deciding Kleene algebras. In *International Conference on Interactive Theorem Proving*, pages 163–178. Springer, 2010.
- [Coh13] Cyril Cohen. Pragmatic quotient types in Coq. In Sandrine Blazy, Christine Paulin-Mohring, and David Pichardie, editors, *Interactive Theorem Proving*, pages 213–228, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [CYO01] Cristiano Calcagno, Hongseok Yang, and Peter W. O’Hearn. Computability and complexity results for a spatial assertion language for data structures. In Ramesh Hariharan, V. Vinay, and Madhavan Mukund, editors, *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science*, pages 108–119, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.
- [dB72] Nicolaas G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Indagationes Mathematicae (Proceedings)*, 75(5):381–392, January 1972. doi:[10.1016/1385-7258\(72\)90034-0](https://doi.org/10.1016/1385-7258(72)90034-0).
- [Dud20] Andrej Dudenhefner. Undecidability of semi-unification on a napkin. In *5th International Conference on Formal Structures for Computation and Deduction (FSCD 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.
- [FHS18] Yannick Forster, Edith Heiter, and Gert Smolka. Verification of PCP-related computational reductions in Coq. In *International Conference on Interactive Theorem Proving*, pages 253–269. Springer, 2018.

<sup>28</sup>where  $R : X \rightarrow Y \rightarrow \mathbb{P}$  is *weakly total* if  $\forall x \exists y R x y$  holds.

- [FK19] Yannick Forster and Fabian Kunze. A Certifying Extraction with Time Bounds from Coq to Call-By-Value Lambda Calculus. In John Harrison, John O’Leary, and Andrew Tolmach, editors, *10th International Conference on Interactive Theorem Proving*, volume 141 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:19, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.ITP.2019.17.
- [FKS19] Yannick Forster, Dominik Kirst, and Gert Smolka. On synthetic undecidability in Coq, with an application to the Entscheidungsproblem. In *International Conference on Certified Programs and Proofs*, pages 38–51. ACM, 2019.
- [FKW20] Yannick Forster, Dominik Kirst, and Dominik Wehr. Completeness Theorems for First-Order Logic Analysed in Constructive Type Theory. In *Symposium on Logical Foundations Of Computer Science, 2020, Deerfield Beach, Florida, U.S.A.*, Jan 2020.
- [FKW21] Yannick Forster, Dominik Kirst, and Dominik Wehr. Completeness theorems for first-order logic analysed in constructive type theory: Extended version. *Journal of Logic and Computation*, 31(1):112–151, 2021.
- [FLW19] Yannick Forster and Dominique Larchey-Wendling. Certified undecidability of intuitionistic linear logic via binary stack machines and minsky machines. In *Proceedings of the 8th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 104–117. ACM, 2019.
- [FLWD<sup>+</sup>20] Yannick Forster, Dominique Larchey-Wendling, Andrej Dudenhefner, Edith Heiter, Dominik Kirst, Fabian Kunze, Gert Smolka, Simon Spies, Dominik Wehr, and Maximilian Wuttke. A Coq Library of Undecidable Problems. In *CoqPL 2020*, New Orleans, LA, United States, 2020. URL: <https://github.com/uds-psl/coq-library-undecidability>.
- [IO01] Samin S. Ishtiaq and Peter W. O’Hearn. BI as an assertion language for mutable data structures. In *Proceedings of the 28th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, pages 14–26, 2001.
- [KH21] Dominik Kirst and Marc Hermes. Synthetic undecidability and incompleteness of first-order axiom systems in Coq. In *12th International Conference on Interactive Theorem Proving (ITP 2021)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2021.
- [KLW20] Dominik Kirst and Dominique Larchey-Wendling. Trakhtenbrot’s theorem in Coq, a constructive approach to finite model theory. In *International Joint Conference on Automated Reasoning*, pages 79–96. Springer, 2020.
- [KS19] Dominik Kirst and Gert Smolka. Categoricity results and large model constructions for second-order ZF in dependent type theory. *Journal of Automated Reasoning*, 63(2):415–438, 2019.
- [Lib10] Leonid Libkin. *Elements of Finite Model Theory*. Springer Publishing Company, Incorporated, 1st edition, 2010.
- [LWF19] Dominique Larchey-Wendling and Yannick Forster. Hilbert’s tenth problem in Coq. In *4th International Conference on Formal Structures for Computation and Deduction*, volume 131 of *LIPIcs*, pages 27:1–27:20, Feb 2019.
- [Lö15] Leopold Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76:447–470, 1915. URL: <http://eudml.org/doc/158703>.
- [MS15] Petar Maksimović and Alan Schmitt. HOCore in Coq. In *International Conference on Interactive Theorem Proving*, pages 278–293. Springer, 2015.
- [Rey02] John C. Reynolds. Separation logic: A logic for shared mutable data structures. In *Proceedings 17th Annual IEEE Symposium on Logic in Computer Science*, pages 55–74. IEEE, 2002.
- [SF20] Simon Spies and Yannick Forster. Undecidability of higher-order unification formalised in Coq. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 143–157, 2020.
- [SS16] Gert Smolka and Kathrin Stark. Hereditarily finite sets in constructive type theory. In *Interactive Theorem Proving - 7th International Conference, ITP 2016, Nancy, France, August 22-27, 2016*, volume 9807 of *LNCS*, pages 374–390. Springer, 2016.
- [SSK19] Kathrin Stark, Steven Schäfer, and Jonas Kaiser. Autosubst 2: reasoning with multi-sorted de Bruijn terms and vector substitutions. In *International Conference on Certified Programs and Proofs*, pages 166–180. ACM, 2019.

- [SST15] Steven Schäfer, Gert Smolka, and Tobias Tebbi. Completeness and decidability of de Bruijn substitution algebra in Coq. In *Proceedings of the 2015 Conference on Certified Programs and Proofs*, pages 67–73. ACM, 2015.
- [Tra50] Boris Abramovich Trakhtenbrot. The impossibility of an algorithm for the decidability problem on finite classes. *Dokl. Akad. Nauk. SSSR*, 70(4):569–572, 1950.
- [VW96] Wim Veldman and Frank Waaldijk. Some elementary results in intuitionistic model theory. *The Journal of Symbolic Logic*, 61(3):745–767, 1996.
- [Wer97] Benjamin Werner. Sets in types, types in sets. In *Theoretical Aspects of Computer Software*, pages 530–546. Springer, Berlin, Heidelberg, September 1997. doi:10.1007/BFb0014566.