RECONSTRUCTION OF NON-$\aleph_0$-CATEGORICAL THEORIES

ITAÏ BEN YAACOV

ABSTRACT. This is work in progress.

The aim is to generalise the correspondence between $\aleph_0$-categorical theories and their automorphism groups (see Ahlbrandt and Ziegler [AZ86] for classical logic and Ben Yaacov and Kaïchouh [BK16] for continuous logic) to arbitrary theories.

CONTENTS

INTRODUCTION

...

We should point out that in the context of categorical logic there exist results which also code a theory with a topological groupoid, e.g. Awodey and Foresell [AF13]. A key difference is that, while we construct our groupoid from models over a fixed set, say **N** (or possibly quotients thereof), Awodey and Forssell consider models over *subsets* of a fixed set. Consequently, Awodey and Forssell's groupoid is merely T_0, and the closure of a singleton representing one model consists of all its sub-models. Our construction, in contrast, associates to a theory T a completely regular Polish groupoid, and (elementary) embeddings of models of T arise as the left-completion of the said groupoid, in close analogy with what happens in the automorphism group of the single model of an $\aleph_0$-categorical theory.

1. GENERAL DEFINITIONS

Let us recall the algebraic definition of a groupoid.

Definition 1.1. A *groupoid* is a set $\mathcal{G}$ equipped with a partial composition law $\cdot\colon \mathcal{G}^2 \dashrightarrow \mathcal{G}$ and an inversion map ${}^{-1}\colon \mathcal{G} \to \mathcal{G}$, such that for all $f, g, h \in \mathcal{G}$:

(i) Composition is associative: $(fg)h = f(gh)$, as soon as one of the two sides is defined (which means that then the other is defined as well).
(ii) The compositions $g^{-1}g$ and gg^{-1} are always defined.
(iii) If fg is defined, then $fgg^{-1} = f$ and $f^{-1}fg = g$.

We call $s_g = g^{-1}g$ the *source* of g and $t_g = gg^{-1}$ its *target*. We call $e \in \mathcal{G}$ *neutral* if $e^2 = e$. The set of neutral elements of $\mathcal{G}$ will be denoted $\mathcal{B}$ or $\mathcal{B}(\mathcal{G})$. We call $\mathcal{B}$ the *base* set of $\mathcal{G}$, and say that $\mathcal{G}$ is a groupoid *over* $\mathcal{B}$.

Let us make a few (standard) observations:

(i) Both s_g and t_g are neutral for all $g \in \mathcal{G}$, defining maps $s, t\colon \mathcal{G} \to \mathcal{B}$.
(ii) The composition fg is defined if and only if $s_f = t_g$. In particular, $s_{fg} = t_{g^{-1}} = s_g$ and $t_{fg} = s_{f^{-1}} = t_f$.
(iii) If e is neutral, then $e = e^{-1}e^2 = e^{-1}e = s_e$, and similarly $e = t_e$. In particular, eg (ge) is defined, necessarily equal to g, if and only if $e = t_g$ ($e = s_g$).
(iv) If fg is neutral, then $f = fgg^{-1} = g^{-1}$ and similarly $g = f^{-1}$. In particular, $(g^{-1})^{-1} = g$ and $(fg)^{-1} = g^{-1}f^{-1}$.

Author supported by ANR project GruPoLoCo (ANR-11-JS01-008).
Revision 3965 *of* 2019-01-08.

From these observations follows the equivalence with the definition of a groupoid as a category all of whose morphisms are invertible, in which case $\mathcal{B}$ is the object set.

Notice that for $A \subseteq \mathcal{G}$ we have $s(A) = A^{-1}A \cap \mathcal{B} = A^{-1}\mathcal{G} \cap \mathcal{B} = \mathcal{G}A \cap \mathcal{B}$ and $t(A) = AA^{-1} \cap \mathcal{B} = A\mathcal{G} \cap \mathcal{B} = \mathcal{G}A^{-1} \cap \mathcal{B}$. Similarly, for $A \subseteq \mathcal{B}$ we have $s^{-1}(A) = \mathcal{G}A$ and $t^{-1}(A) = A\mathcal{G}$.

The advantage of the algebraic definition is that it is easier to cast a topology on top of it.

Definition 1.2. A *topological groupoid* is a groupoid such that $\mathcal{G}$ is a Hausdorff topological space and composition and inversion are continuous (where defined).

A topological groupoid $\mathcal{G}$ with base $\mathcal{B}$ is *open* is the source map $s\colon \mathcal{G} \to \mathcal{B}$ is open.

Clearly, a topological groupoid is open if and only if its target map is open. Every topological group, viewed as a groupoid, is open.

Since the source and target maps are total, the domain of composition, defined by the condition $t(g) = s(f)$, is closed in $\mathcal{G}^2$. It follows that the condition $g^2 = g$ is closed, so the base set $\mathcal{B}$ is a closed subset of $\mathcal{G}$.

Notation 1.3. Let X and Y be sets equipped with maps $\pi\colon X \to \mathcal{B}$ and $\rho\colon Y \to \mathcal{B}$. We then define

$$X \times_{\mathcal{B}} Y = \{(x, y) \in X \times Y : \pi x = \rho y\}.$$

When X is equipped with two maps $s_X, t_X\colon X \to \mathcal{B}$, we implicitly take $\pi = s_X$, and when Y is equipped with $s_Y, t_Y\colon Y \to \mathcal{B}$, we take $\rho = t_Y$.

Definition 1.4. Let $\mathcal{G}$ be a groupoid with base $\mathcal{B}$, and let X be a space, equipped with a map $\pi\colon X \to \mathcal{B}$. A *(left) action* of $\mathcal{G}$ on X, denoted $\mathcal{G} \curvearrowright X$, is a map $\mathcal{G} \times_{\mathcal{B}} X \to X$, sending $(g, x) \mapsto gx$, such that $(gh)x = g(hx)$ whenever either is defined (so $\pi(gx) = t(g)$). It is a *continuous* action if $\mathcal{G}$ is a topological groupoid, X is a topological space, π is continuous, and the product map $\mathcal{G} \times_{\mathcal{B}} X \to X$ is continuous.

A *right action* $X \curvearrowleft \mathcal{G}$ is defined analogously as a map $X \times_{\mathcal{B}} \mathcal{G} \to X$.

In particular, the product map $\mathcal{G} \times_{\mathcal{B}} \mathcal{G} \to \mathcal{G}$ is both a left and a right continuous action of $\mathcal{G}$ on itself.

Lemma 1.5. *The following are equivalent for a topological groupoid $\mathcal{G}$ over a base $\mathcal{B}$:*
- (i) *The groupoid $\mathcal{G}$ is open.*
- (ii) *For any topological space X equipped with a continuous map $\pi\colon X \to \mathcal{B}$, the projection $\mathcal{G} \times_{\mathcal{B}} X \to X$ is open.*
- (iii) *For any continuous action $\mathcal{G} \curvearrowright X$, the action law $\mathcal{G} \times_{\mathcal{B}} X \to X$ is open.*
- (iv) *The groupoid law $\mathcal{G} \times_{\mathcal{B}} \mathcal{G} \to \mathcal{G}$ is open.*

Proof. (i) $\Longrightarrow$ (ii). A basic open set of $\mathcal{G} \times_{\mathcal{B}} X$ is of the form $U \times_{\mathcal{B}} V$, where $U \subseteq \mathcal{G}$ and $V \subseteq X$ are open. Since $\mathcal{G}$ is open, the sets $W = s(U) \subseteq \mathcal{B}$ and $\pi^{-1}(W) \subseteq X$ are open, and the image of $U \times_{\mathcal{B}} V$ in X is the open set $V \cap \pi^{-1}(W)$.

(ii) $\Longrightarrow$ (iii). Compose with $(g, x) \mapsto (g^{-1}, gx)$.

(iii) $\Longrightarrow$ (iv). This is a special case.

(iv) $\Longrightarrow$ (i). If $U \subseteq \mathcal{G}$ is open, then $U^{-1}U \subseteq \mathcal{G}$ is open, and therefore $s(U) = U^{-1}U \cap \mathcal{B}$ is open in $\mathcal{B}$. ■

Definition 1.6. Let $\mathcal{G}$ be a topological groupoid with base $\mathcal{B}$, let $\mathcal{C}$ be some topological space, and let $\theta\colon \mathcal{C} \to \mathcal{B}$ be continuous. We define the *base change* of $\mathcal{G}$ to $\mathcal{C}$, denoted $\mathcal{G}_{\mathcal{C}}$, as follows. The underlying topological space is

$$\mathcal{G}_{\mathcal{C}} = \mathcal{C} \times_{\mathcal{B}} \mathcal{G} \times_{\mathcal{B}} \mathcal{C} = \{(b, g, a) \in \mathcal{C} \times \mathcal{G} \times \mathcal{C} : \theta a = s_g, \ \theta b = t_g\}.$$

For $(c, f, b), (b, g, a) \in \mathcal{G}_{\mathcal{C}}$ define

$$(c, f, b)(b, g, a) = (c, fg, a), \qquad (b, g, a)^{-1} = (a, g^{-1}, b).$$

By abuse of notation (which we justify below), we define $\theta\colon \mathcal{G}_{\mathcal{C}} \to \mathcal{G}$ by $\theta(b, g, a) = g$.

When $\mathcal{C} \subseteq \mathcal{B}$ and θ is the inclusion map, we also denote $\mathcal{G}_{\mathcal{C}}$ by $\mathcal{G}{\restriction}_{\mathcal{C}}$.

Lemma 1.7. *Let $\mathcal{G}$ be a topological groupoid with base $\mathcal{B}$, and let $\theta\colon \mathcal{C} \to \mathcal{B}$ be continuous.*
- (i) *The structure $\mathcal{G}_{\mathcal{C}}$ is a topological groupoid. The map $a \mapsto (a, \theta a, a)$ is a homeomorphism between $\mathcal{C}$ and the base of $\mathcal{G}_{\mathcal{C}}$, which we shall henceforth identify. Under this identification, $s(b, g, a) = a$ and $t(b, g, a) = b$.*
- (ii) *The map $\theta\colon \mathcal{G}_{\mathcal{C}} \to \mathcal{G}$, as defined above, is continuous, and extends $\theta\colon \mathcal{C} \to \mathcal{B}$ under the identification of the previous item. The two maps are surjective or open simultaneously.*
- (iii) *If $\mathcal{G}$ is open and θ is open, then $\mathcal{G}_{\mathcal{C}}$ is open.*

Proof. Items (i) and (ii) are clear, except possibly for openness of θ. Assume first that $\theta\colon \mathcal{C} \to \mathcal{B}$ is open. We have a basis of open sets for $\mathcal{G}_{\mathcal{C}}$ of the form $O = U \times_{\mathcal{B}} V \times_{\mathcal{B}} W = (U \times V \times W) \cap \mathcal{G}_{\mathcal{C}}$, where $U, W \subseteq \mathcal{C}$ and $V \subseteq \mathcal{G}$ are open. Then $\theta O = V \cap t^{-1}(\theta U) \cap s^{-1}(\theta W)$ is open in $\mathcal{G}$. Conversely, assume that $\theta\colon \mathcal{G}_{\mathcal{C}} \to \mathcal{G}$ is open and let $U \subseteq \mathcal{C}$ be open. Then $O = U \times_{\mathcal{B}} \mathcal{G} \times_{\mathcal{B}} \mathcal{C} \subseteq \mathcal{G}_{\mathcal{C}}$ is open, and so $\theta U = \theta O \cap \mathcal{B}$ is open in $\mathcal{B}$.

For (iii), we assume that $\mathcal{G}$ and θ are open. Let again $O = U \times_{\mathcal{B}} V \times_{\mathcal{B}} W$ be a basic open set. Since θ is open, we may assume that $V \subseteq t^{-1}(\theta U)$. Then $sO = W \cap \theta^{-1}(sV)$, and it is open since $\mathcal{G}$ is. Therefore $\mathcal{G}_{\mathcal{C}}$ is open. ∎

Definition 1.8. Let $\mathcal{G}$ be a topological groupoid, and let $m \geq 1$. Its *m-fold s-fibred power*

$$\mathcal{G}^{m/s} = \{g \in \mathcal{G}^m : s_{g_0} = s_{g_1} = \cdots\}$$

is its power as a right $\mathcal{G}$-space, and carries a right action $\mathcal{G}^{m/s} \curvearrowleft \mathcal{G}$. We define $\mathcal{P}_m$ (or more precisely, $\mathcal{P}_m(\mathcal{G})$) as the topological quotient

$$\mathcal{P}_m = \mathcal{G}^{m/s}/\mathcal{G} = \{g\mathcal{G} : g \in \mathcal{G}^{m/s}\}.$$

The action $\mathcal{G}^{m/s} \curvearrowleft \mathcal{G}$ commutes with the natural actions $\mathcal{G}^m \curvearrowright \mathcal{G}^{m/s}$ (over the base $\mathcal{B}^m$) and $\mathfrak{S}_m \curvearrowright \mathcal{G}^{m/s}$, as well as with the projection on the first m coordinates $\mathcal{G}^{m+1/s} \to \mathcal{G}^{m/s}$. These give rise to actions and projection map

$$\mathcal{G}^m \curvearrowright \mathcal{P}_m, \qquad \mathfrak{S}_m \curvearrowright \mathcal{P}_m, \qquad \mathcal{P}_{m+1} \to \mathcal{P}_m.$$

We also define $\mathcal{G}^{0/s} = \mathcal{B}$ (the neutral for fibred product over $\mathcal{B}$), with action $\mathcal{B} \curvearrowleft \mathcal{G}$ given by $e \cdot g = s_g$ whenever $t_g = e$.

There three natural groupoid actions on $\mathcal{G}^{m/s}$:

- The left action $\mathcal{G}^m \curvearrowright \mathcal{G}^{m/s}$, relative to $t\colon \mathcal{G}^{m/s} \to \mathcal{B}^m$,
- the right action $\mathcal{G}^{m/s} \curvearrowleft \mathcal{G}$, which can be viewed as an action $\mathcal{G} \curvearrowright (\mathcal{G}^{m/s}, s)$ defined by $h \cdot g = g \cdot h^{-1}$ (for $g \in \mathcal{G}^{m/s}$ and $h \in \mathcal{G}$), relative to $s\colon \mathcal{G}^{m/s} \to \mathcal{B}$,
- and the mixed action $\mathcal{G}^m \curvearrowright \mathcal{G}^{m/s} \curvearrowleft \mathcal{G}$, which can be viewed as an action $\mathcal{G}^{m+1} \curvearrowright (\mathcal{G}^{m/s}, t, s)$ defined by $(f, h) \cdot g = f \cdot g \cdot h^{-1}$ (for $f \in \mathcal{G}^m$, $g \in \mathcal{G}^{m/s}$ and $h \in \mathcal{G}$), relative to $(t, s)\colon \mathcal{G}^{m/s} \to \mathcal{B}^{m+1}$.

Lemma 1.9. *The action $\mathcal{G}^{m+1} \curvearrowright \mathcal{P}_{m+1}$ is homeomorphic to the mixed action on $\mathcal{G}^{m/s}$, for all $m \geq 0$. When $m \geq 1$, this is given by sending*

$$(g_i : i \leq m)\mathcal{G} \mapsto (g_i g_m^{-1} : i < m),$$

and when $m = 0$, this is

$$g\mathcal{G} \mapsto t_g.$$

In particular, the action $\mathcal{G}^m \curvearrowright \mathcal{P}_m$ is continuous for all $m \geq 1$.

Proof. Indeed, in either case ($m \geq 1$ or $m = 0$) the map is well defined and continuous, admitting a continuous inverse map:

$$f \mapsto (f, s_f)\mathcal{G} \qquad m \geq 1,$$
$$e \mapsto e\mathcal{G} \qquad m = 0.$$

One verifies that these maps send one action to the other. ∎

In particular, $\mathcal{G} \curvearrowright \mathcal{P}_1$ is $\mathcal{B} \curvearrowleft \mathcal{G}$, and $\mathcal{G}^2 \curvearrowleft \mathcal{P}_2$ is $\mathcal{G} \curvearrowright \mathcal{G} \curvearrowleft \mathcal{G}$ (identifying the last $\mathcal{G}$ with $\mathcal{G}^{\mathrm{op}}$ in each case). In addition, the action $\mathfrak{S}_2 \curvearrowright \mathcal{P}_2$ coincides with $\{\mathrm{id}, \cdot^{-1}\} \curvearrowright \mathcal{G}$, and the projection $\mathcal{P}_2 \to \mathcal{P}_1$ is $t\colon \mathcal{G} \to \mathcal{B}$.

2. The groupoid associated to a classical theory

In this Section, let T denote a theory, in the sense of classical (i.e., not continuous) first order logic, in a countable language $\mathcal{L}$. We do not assume that T is complete. We consider that by definition of the logic, all structures (so all models of T) are not empty. In order to avoid borderline cases, let us also assume that no model of T is a singleton (or, if T is multi-sorted, that in no model are all sorts singletons).

Definition 2.1. Let T be a classical first-order theory in a countable language. Let $\mathcal{G}_0(T) \subseteq \mathrm{S}_{2\times\mathbf{N}}(T)$ consist of all possible types of a pair of enumerations of a model of T (i.e., any two enumerations of any single countable model). Members of $\mathcal{G}_0(T)$ will be denoted g, h, and so on, or possibly as types $g(x, y)$ where x and y stand for countable tuples of variables. Let $\mathcal{B}_0(T) \subseteq \mathcal{G}_0(T)$ to be the subset defined by the condition $x = y$. We may identify $\mathrm{tp}(a, a) \in \mathcal{B}_0(T)$ with $\mathrm{tp}(a)$, thus identifying $\mathcal{B}_0(T)$ with the subset of $\mathrm{S}_{\mathbf{N}}(T)$ consisting of types of enumerations of models.

If $g = \mathrm{tp}(a, b)$ and $h = \mathrm{tp}(b', c')$, where $b \equiv b'$, then we might as well assume that $b = b'$, in which case $g^{-1} = \mathrm{tp}(b, a)$ and $gh = \mathrm{tp}(a, c')$ depend only on g and h, and belongs to $\mathcal{G}_0(T)$.

Lemma 2.2. *As defined above, $\mathcal{G}_0(T)$ is a Polish open topological groupoid with base $\mathcal{B}_0(T)$. If $g = \mathrm{tp}(a, b) \in \mathcal{G}_0(T)$, then $t_g = \mathrm{tp}(a)$ and $s_g = \mathrm{tp}(b)$.*

Proof. It is easy to check that $\mathcal{G}_0(T)$ is indeed a topological groupoid. Let us prove that $\mathcal{G}_0(T)$ is open, i.e., that the map $s\colon \mathrm{tp}(a, b) \mapsto \mathrm{tp}(b)$ is open. A basic open set $U \subseteq \mathcal{G}_0(T)$ is defined by a formula $\varphi(x, y)$ (in which only finitely many variables actually appear). We claim that $s(U)$ is defined by $\exists x\, \varphi(x, y)$ (quantifying only over those x_i that appear in φ). Indeed, let $\mathrm{tp}(b) \in \mathcal{B}_0(T)$, so b enumerate some $M \vDash T$. If $g = \mathrm{tp}(a, b) \in U$, then a also enumerates M, $s(g) = \mathrm{tp}(b)$, and $\vDash \varphi(a, b)$ implies $\vDash \exists x\, \varphi(x, b)$. Conversely, if $\vDash \exists x\, \varphi(x, b)$, then there exists a tuple a in M such that $\vDash \varphi(a, b)$. Since only finitely many variables actually appear in φ, we may replace a tail of a with an enumeration of M, so still $\vDash \varphi(a, b)$, and now $g = \mathrm{tp}(a, b) \in U$. Thus $s(U)$ is indeed defined by $\exists x\, \varphi(x, y)$. ■

Our goal is to associate to each theory T a groupoid $\mathcal{G}(T)$ such that for any theory T' we have $\mathcal{G}(T) \cong \mathcal{G}(T')$ as topological groupoids if and only if T and T' are bi-interpretable. In fact, we desire a seemingly stronger version of the left-to-right implication, to which we refer as *reconstruction*: a procedure by which we obtain, from $\mathcal{G}(T)$, a theory bi-interpretable with T, in a (reasonably) constructive fashion. While $\mathcal{G}_0(T)$ may seem natural, neither implication seems to hold for it, nor, *a fortiori*, reconstruction. Indeed, naïve attempts at reconstruction quickly run into obstacles that seem to arise from the fact that the base $\mathcal{B}_0(T)$ is not compact. The following definition was originally an attempt to remedy this. It turned out, quite surprisingly (to the author, in any case), to provide the opposite implication as well, together with a "clean" relation to the automorphism group associated to an $\aleph_0$-categorical theory.

We assume throughout that T is in a single-sorted language. The definitions and arguments adapt in an obvious manner to the multi-sorted case, with additional bookkeeping that we prefer to avoid.

Definition 2.3. Assume T is a theory in classical logic. Fix an enumeration of formulas $\Phi = \big(\varphi_n(x_{<n}, y) : n \in \mathbf{N}\big)$, such that every formula of the form $\varphi(x_{<k}, y)$ appears as φ_n for some $n \geq k$ (with dummy variables). Define

$$\varphi'_n(x_{\leq n}) = \big(\exists y \varphi_n(x_{<n}, y)\big) \to \varphi_n(x_{<n}, x_n), \qquad \varphi''_n(x_{<n}) = \bigwedge_{k<n} \varphi'_k, \qquad \varphi''(x) = \bigwedge_{k \in \mathbf{N}} \varphi'_k.$$

Each φ''_n is a formula, φ'' is a partial type.

Let D_Φ denote the type-definable set of sequences satisfying φ''. Let $D_{\Phi,n}$ denote the definable set of n-tuples satisfying φ''_n, and let $D_{m\Phi} = D_\Phi^m$.

We define $\mathrm{S}_{m\Phi}(T) \subseteq \mathrm{S}_{m\times\mathbf{N}}(T)$ to be the (compact) set of possible types of members of the type-definable set $D_{m\Phi}$. We define $\mathcal{B}_\Phi(T) = \mathrm{S}_\Phi(T) = \mathrm{S}_{1\Phi}(T)$, so $\mathcal{B}_\Phi(T) \subseteq \mathcal{B}_0(T)$ (any member of Φ must satisfy the Tarski-Vaught test), and $\mathcal{G}_\Phi(T) = \mathcal{G}_0(T)\restriction_{\mathcal{B}_\Phi(T)}$.

In other words, $\mathcal{G}_\Phi(T) = \mathcal{G}_0(T) \cap \mathrm{S}_{2\Phi}(T)$ consists of all $\mathrm{tp}(a, b)$ where $a, b \in \Phi$ enumerate the same set. Since φ'' imposes that each element is repeated infinitely often, this implies that each of a and b is a sub-sequence of the other.

The following is immediate from the definitions, but deserves nonetheless to be stated explicitly:

Lemma 2.4. *Any member of $D_{\Phi,n}$ in a countable model $M \vDash T$ can be extended to a member of D_Φ that moreover enumerates M.*

Lemma 2.5. *As defined above, $\mathcal{G}_\Phi(T)$ is open, and its base $\mathcal{B}_\Phi(T)$ is the Cantor set.*

Proof. The base $\mathcal{B}_\Phi(T)$ is totally disconnected, compact and second-countable by construction. We have agreed to assume that no model of T is a singleton, so no sentence that implies, modulo T, that the model is a singleton. This excludes the possibility of isolated points in $\mathcal{B}_\Phi(T)$, which is therefore the Cantor set. Let $U \subseteq \mathcal{G}_\Phi(T)$ be a basic open set, say defined by a formula $\chi(x_{<n}, y_{<n})$. We claim that sU is defined by $\exists x_{<n}\, \big(x_{<n} \in D_{\Phi,n} \wedge \chi(x_{<n}, y_{<n})\big)$. One inclusion is as for Lemma 2.2, while the other follows from Lemma 2.4. ■

This construction raises several questions. The easiest one is that of dependence on the choice of Φ.

Proposition 2.6. *Let T be a classical theory, and let Φ and Ψ be enumerations as in Definition 2.3. Then the groupoids $\mathcal{G}_\Phi(T)$ and $\mathcal{G}_\Psi(T)$ are isomorphic (as topological groupoids).*

Proof. Let us define functions $m, n\colon \mathbf{N} \to \mathbf{N}$ by induction. If there exists $j < i$ such that $n(j) = i$, then we define $m(i)$ to be the least $k > m(i-1)$ such that φ_k is $x_j = y$. Otherwise, we choose the least $k > m(i-1)$ (where $m(-1) = -1$) where

$$\varphi_k(x_{<k}, y) = \psi_i(x_{m(0)}, \ldots, x_{m(i-1)}, y).$$

We choose $n(i) > n(i-1)$ in the mirror-image fashion.

For $p = \mathrm{tp}(a,b) \in S_{2\times\mathbf{N}}(T)$, let us define $m^*(a) = (a_{m(i)} : i \in \mathbf{N})$ and $m^*(p) = \mathrm{tp}\big(m^*(a), m^*(b)\big)$, and similarly for n.

Assume now that $a \in D_\Phi$. Then $a_{m\circ n(j)} = a_j$ for all j, so $n^* \circ m^*(a) = a$. In addition, we claim that $\psi'_i(a_{m(0)}, \ldots, a_{m(i)})$ holds. Indeed, if i is not in the range of n, then

$$\psi'_i(a_{m(0)}, \ldots, a_{m(i)}) = \varphi'_{m(i)}(a_{\leq m(i)}),$$

which holds. If, on the other hand, i is in the range of n, then $i = n(j)$ for some $j \leq i$. Then $a_{m(i)} = a_j$, $\psi_i(x_{<i}, y) = \varphi_j(x_{n(0)}, \ldots, x_{n(j-1)}, y)$, and so

$$\psi'_i(a_{m(0)}, \ldots, a_{m(i)}) = \varphi'_j(a_{m\circ n(0)}, \ldots, a_{m\circ n(j)}) = \varphi'(a_{\leq j}),$$

and the latter holds. Therefore, $m^*(a) \in D_\Psi$.

We obtain continuous morphisms $m^*\colon \mathcal{G}_\Phi(T) \to \mathcal{G}_\Psi(T)$ and, by the same reasoning, $n^*\colon \mathcal{G}_\Psi(T) \to \mathcal{G}_\Phi(T)$. They are inverses, so $\mathcal{G}_\Phi(T) \cong \mathcal{G}_\Psi(T)$. ■

Therefore, we may write $\mathcal{G}(T)$, omitting Φ. Moreover, it does not depend on a the choice of language: if we name definable predicates or functions by new symbols, adding to T the corresponding axioms, then $\mathcal{G}(T)$ does not change. The same argument can be improved to show that $\mathcal{G}(T)$ only depends on T up to bi-interpretation.

Let us first consider the case of a plain interpretation. One can always break it down into two steps:

(i) adjoining a definable set (without parameters) in an imaginary sort as a new sort, and
(ii) forgetting the original sort, keeping some of the induced structure on the new sort.

Lemma 2.7. *Let T' be obtained from T by adjoining a definable subset (without parameters) of an imaginary sort as a new sort. Then $\mathcal{G}(T) \cong \mathcal{G}(T')$.*

Proof. We may assume, for simplicity, that T has a single sort S, and T' has an additional sort $S' = D/E$, where $D \subseteq S^d$ is definable and E a definable equivalence relation. Let $x = (x_n : n \in \mathbf{N})$ be an enumeration of variables of T', where $x_{(d+2)n}$ are in S' and all others in S.

We will now choose an enumeration Φ of formulae $\varphi_n(x_{<n}, y)$ of T', as is Definition 2.3, where y is in S' if and only if $d+2$ divides n. For $i < d$, let:

$$\chi_i(v, u_{\leq i}) = (\exists u_{i+1,\ldots,d-1})\,\big(u_{<d} \in D \,\&\, v = [u_{<d}]_E\big).$$

For n divisible by $d+2$ and $i < d$, we choose φ_{n+i+1} to be

$$\chi_i(x_{n,\ldots,n+i}, y).$$

This leaves $\varphi_{m(d+2)}$ (with y in S') and $\varphi_{m(d+2)-1}$ (with y in S) without constraints, so we can choose the entire family Φ to be as in Definition 2.3. If $a \in D_\Phi$ and $d+2$ divides n, then $a_n = [a_{n+1}, \ldots, a_{n+d}]_E$.

Now let $\mathbf{N}' = \mathbf{N} \smallsetminus (d+2)\mathbf{N}$, and let us try to restrict everything to $x' = (x_n : n \in \mathbf{N}')$, i.e., to those x_n that are in S. Let $x'_{<n}$ denote the tuple of x'_i such that $i < n$ (i.e., x_i such that $i < n$ and $i \in \mathbf{N}'$). When $d+2 \mid n+1$, let $\psi_n(x'_{<n}, y)$ be the formula obtained from $\varphi(x_{<n}, y)$ by substituting $[x_{i+1,\ldots,i+d}]_E$ for each x_i with $d+2 \mid i$. In other words, we replace each reference to a variable in S' with a reference to its representative, obtaining a formula in the sorts of T.

When $d+2 \mid n$, the variable y is in S'. Doing the same, and in addition replacing y with $[u_{<d}]_E$ and adding the requirement that $u_{<d} \in D$, we obtain a formula $\psi_n(x'_{<n}, u_{<d})$. We let ψ_{n+i+1}, for $i < d$, be:

$$(\exists u_{i+1,\ldots,d-1})\,\psi_n(x'_{<n+i+1}, y, u_{i+1,\ldots,d-1}).$$

(In particular, the tuple substituted for $u_{<d}$ is $x'_{n+1,\ldots,n+i}, y, u_{i+1,\ldots,d-1}$.)

Let $\Psi = (\psi_n : n \in \mathbf{N}')$. Then Ψ consists only of formulas with variables in S. Moreover, if we take a realisation $a \in D_\Phi$ and let $a' = (a_n : n \in \mathbf{N}')$, then $a' \in D_\Psi$, and conversely, any $a' \in D_\Psi$ can be extended in a unique fashion to $a \in D_\Phi$, by letting $a_n = [a_{n+1,\ldots,n+d}]_E$ for $d+2 \mid n$.

Finally, the sequence Ψ satisfies the hypotheses of Definition 2.3, since Φ does (any formula that needs to appear only refers to variables in S, and already appears in the sequence Φ). We conclude that $\mathcal{G}(T) \cong \mathcal{G}_\Psi(T) \cong \mathcal{G}_\Phi(T') \cong \mathcal{G}(T')$. ■

Lemma 2.8. *Let T' be a reduct of T, i.e., obtained by forgetting some sorts, and some of the structure on the remaining sorts. Then there exists a continuous morphism $\mathcal{G}(T) \to \mathcal{G}(T')$.*

Proof. Easy. ■

One usually defines a *bi-interpretation* between T and T' as a pair of interpretations of one in the other, such that, when composed to yield an interpretation of T or of T' in itself, the models are uniformly definably isomorphic to their interpreted copies. It is however fairly easy to check that this is equivalent to the property that the theory obtained by adjoining to T the sort of T' (without forgetting anything), and that that is obtained by adjoining to T' the sort of T, are the same up to a change of language.

Theorem 2.9. *If T and T' are bi-interpretable, then $\mathcal{G}(T) \cong \mathcal{G}(T')$. If T' is merely interpretable in T, then there exists a continuous morphism $\mathcal{G}(T) \to \mathcal{G}(T')$.*

Proof. By Lemma 2.7, Lemma 2.8 and the preceding paragraph. ∎

When T is $\aleph_0$-categorical (so, in particular, complete), another issue arises, namely, that we have already associated to T a different object, the topological *group* $G(T) = \mathrm{Aut}(M)$, where M is any countable model of T. Viewing $G(T)$ as a topological groupoid, it is distinct from $\mathcal{G}(T)$, since the base of $G(T)$ is a singleton (its identity). Our next result says that this is the only difference between the two.

Let G be a topological group and $\mathcal{B}$ an arbitrary topological space. Let us consider G as a groupoid over a singleton $*$, and let $\mathcal{B} \twoheadrightarrow *$ be the unique map. Then $G_{\mathcal{B}\twoheadrightarrow *}$ is a groupoid based over $\mathcal{B}$.

Definition 2.10. Say that a topological groupoid is *trivially based* if it is isomorphic, as a topological groupoid, to a groupoid of the form $G_{\mathcal{B}\twoheadrightarrow *}$.

Fact 2.11. *Let $\mathcal{G}$ be a topological groupoid with base $\mathcal{B}$. Then the following are equivalent:*

(i) *The groupoid $\mathcal{G}$ is trivially based.*
(ii) *For all (at least one) $e \in \mathcal{B}$ there exists continuous map $\rho\colon \mathcal{B} \to \mathcal{G}$ such that $s \circ \rho \equiv e$ and $t \circ \rho = \mathrm{id}$.*

In this case $\mathcal{G} \cong G_{\mathcal{B}\twoheadrightarrow *}$, where $G \cong \mathcal{G}_e = \{g \in \mathcal{G} : s_g = t_g = e\}$.

Proof. Assume first that is trivially based, say $\mathcal{G} = G_{X\twoheadrightarrow *}$, which, as a topological space, is just $X \times G \times X$. Let $e \in X$. Then $X = \mathcal{B}$, $G \cong \mathcal{G}_e$, and the map $\rho(e') = (e', 1, e)$ is as desired. Conversely, assume that ρ exists for some e, and let $G = \mathcal{G}_e$. Then $\psi(g) = \rho(t_g)^{-1} g \rho(s_g) \in G$ for all $g \in \mathcal{G}$, and

$$g \mapsto \big(t_g, \psi(g), s_g\big)$$

is the desired isomorphism $\mathcal{G} \cong G_{\mathcal{B}\twoheadrightarrow *}$. ∎

Proposition 2.12. *Let T be $\aleph_0$-categorical. Then $\mathcal{G}(T)$ is trivially based over the Cantor set $\mathcal{C}$.*

Proof. Let Φ be as in Definition 2.3, and let $q(x) \in \mathcal{B}_\Phi(T)$. We have already observed that $\mathcal{B}_\Phi(T) \cong \mathcal{C}$. Let $q_n(x_{<n})$ be the restriction of q to $x_{<n}$.

We choose an increasing sequence $(N_k : k \in \mathbf{N})$, with $N_0 = 0$, such that if $a \vDash q$, then any 1-type over $a_{<N_k}$ is realised in $\{a_i : N_k \leq i < N_{k+1}\}$. Then, if $N_k \leq n < N_{k+1}$, we choose $m(n) > m(n-1)$ such that $\varphi_{m(n)}(x_{<m(n)}, y)$ is the formula saying that $q_{n+1}(x_{m(0),\ldots,m(n-1)}, y)$ holds, and in addition, if $q_{n+1}(x_{m(0),\ldots,m(n-1)}, x_k)$ holds, then $y = x_k$.

If $a \in D_\Phi$, then $m^*(a)$ enumerates the same set as a, and $m^*(a) \vDash q$. Therefore, if $p = \mathrm{tp}(a) \in \mathcal{B}(T)$, then $\rho(p) = \mathrm{tp}\big(a, m^*(a)\big) \in \mathcal{G}(T)$, and $\rho\colon \mathcal{B}(T) \to \mathcal{G}(T)$ is as in Fact 2.11. It is easy to check that $\mathcal{G}(T)_q \cong G(T)$, concluding the proof. ∎

3. Reconstructing a theory

We turn to reconstruction, namely, recovering T, up to bi-interpretation, from the topological groupoid $\mathcal{G} = \mathcal{G}(T) = \mathcal{G}_\Phi(T)$, for some (any) choice of Φ. Let us first give a general construction, and then explain why it works. Recall the actions $\mathcal{G}^m \curvearrowright \mathcal{G}^{m/s} \curvearrowleft \mathcal{G}$ and $\mathcal{G}^m \curvearrowright \mathcal{P}_m = \mathcal{G}^{m/s}/\mathcal{G}$ defined in Section 1. Given a subset $X \subseteq \mathcal{P}_m$ and a tuple $S = (S_i : i < m)$, where $S_i \subseteq \mathcal{G}$, let

$$SX = \left(\prod_i S_i\right) X = \big\{(s_i x_i) : \text{whenever defined for } (x_i) \in X \text{ and } s_i \in S_i\big\}.$$

Definition 3.1. Let $\mathcal{G}$ be an open groupoid with compact base. Let $\mathscr{H}$ be the collection of clopen sub-groupoids of $\mathcal{G}$ which contain $\mathcal{B}$.

$$\mathscr{H} = \big\{\mathcal{H} \subseteq \mathcal{G} \text{ clopen} : \mathcal{H} = \mathcal{H}\mathcal{H}^{-1} \supseteq \mathcal{B}\big\}.$$

For each m and sequence $\mathcal{H} \in \mathscr{H}^m$, let $\mathcal{L}_\mathcal{H}$ be the Boolean algebra of clopen, $\mathcal{H}$-invariant subsets of $\mathcal{P}_m$:

$$\mathcal{L}_\mathcal{H} = \big\{X \subseteq \mathcal{P}_m \text{ clopen} : \mathcal{H}X = X\big\}.$$

Define the language $\mathcal{L}(\mathcal{G})$ as follows. It contains a sort $D_{\mathcal{H}}$ for each $\mathcal{H} \in \mathscr{H}$. For each $\mathcal{H} \in \mathscr{H}^m$ and $X \in \mathcal{L}_{\mathcal{H}}$, $\mathcal{L}(\mathcal{G})$ contains a predicate symbol $Q_X(u_0, \ldots, u_{m-1})$, where u_i is in the sort $D_{\mathcal{H}_i}$.

For each $e \in \mathcal{B}$ we define a $\mathcal{L}(\mathcal{G})$-structure M_e. For each $\mathcal{H} \in \mathscr{H}$, we interpret the sort $D_{\mathcal{H}}$ in M_e as the set $\{\mathcal{H}g : g \in \mathcal{G}e\}$ (where $\mathcal{G}e = \{g \in \mathcal{G} : s_g = e\}$). For each predicate Q_X on $\prod_{i<m} D_{\mathcal{H}_i}$, and each $g \in (\mathcal{G}e)^m \subseteq \mathcal{G}^{m/s}$, we define $Q_X(\ldots, \mathcal{H}_i g_i, \ldots)$ to hold if $g\mathcal{G} \in X$.

Finally, we define $T(\mathcal{G})$ to be the $\mathcal{L}(\mathcal{G})$-theory of the family $\{M_e : e \in \mathcal{B}\}$.

Now let T be a theory and $\mathcal{G} = \mathcal{G}(T) = \mathcal{G}_\Phi(T)$ for some Φ. We shall prove that the theory $T(\mathcal{G})$, as defined in Definition 3.1, is bi-interpretable with T.

Lemma 3.2. *Let $g = (g_i : i < m) \in \mathcal{G}^{m/s}$, with common source $e = \mathrm{tp}(a)$. Then each g_i can be written in a unique fashion as $\mathrm{tp}(b_i, a)$, and the map $g \mapsto \mathrm{tp}(b_i : i < m)$ is a dense topological embedding $\mathcal{P}_m \hookrightarrow \mathrm{S}_{m\Phi}(T)$.*

Identifying $\mathcal{P}_m$ with its image in this fashion, it consists of all types of m sequences in D_Φ, all of which enumerate the same model (so in particular, $\mathcal{P}_1 = \mathcal{B}$ and $\mathcal{P}_2 = \mathcal{G}$, which we already knew).

Proof. It is easy to check that the map is continuous, and to construct a continuous inverse via a section into $\mathcal{G}^{m/s}$. ■

Once we have described the sets $\mathcal{P}_m$ in terms of T, we can describe the language.

Definition 3.3. A *definable equivalence relation on D_Φ* is, as its name suggests, an equivalence relation on D_Φ defined by a formula $E(x, y)$, in which x and y are infinite tuples. We let $D_E = D_\Phi / E$ be the quotient space. Given several such relations $(E_i : i < m)$, their product $E = \prod_i E_i$ will be called an *m-fold definable equivalence relation on D_Φ* (which is more restrictive than merely a definable equivalence relation on $D_{m\Phi}$), and again we let $D_E = D_{m\Phi}/E = \prod_i D_{E_i}$.

If E is a definable equivalence relation on D_Φ, then only finitely many variables actually occur in the formula E, say all among $x_{<n}, y_{<n}$, and the formula E also defines an equivalence relation on $D_{\Phi,n}$. It follows from Lemma 2.4 that the map $[a]_E \mapsto [a_{<n}]_E$ is a bijection between D_E and $D_{\Phi,n}/E$. Moreover, this bijection does not depend on the choice of n, up to a natural definable bijection between $D_{\Phi,n}/E$ and, say, $D_{\Phi,n'}/E$. Therefore D_E can be considered to be an interpreted sort (a non-empty definable subset of an imaginary sort). When E is an n-fold definable equivalence relation on D_Φ, then D_E is a product of interpreted sorts (so again an interpreted sort).

Definition 3.4. When E is an m-fold equivalence relation on D_Φ, we define $\mathrm{S}_E(T)$ to be the space of types in the sort D_E, giving rise to natural maps

$$\chi_E \colon \mathrm{S}_{m\Phi}(T) \to \mathrm{S}_E(T) \qquad \text{and its restriction} \qquad \chi'_E \colon \mathcal{P}_m \to \mathrm{S}_E(T).$$

Notation 3.5. Let E be an m-fold equivalence relation on D_Φ. For $i < m$, let x_i be an infinite sequence representing an element of D_Φ, and let u_i be in the sort D_{E_i}.

(i) For a formula $\psi(x_0, \ldots, x_{m-1})$, we let $[\psi] \subseteq \mathrm{S}_{m\Phi}(T)$ be the clopen subset defined by ψ, and $\psi_E(u_0, \ldots, u_{m-1})$ the expression

$$(\exists x_0, \ldots, x_{m-1} \in D_\Phi) \left(\psi(x_0, \ldots, x_{m-1}) \wedge \bigwedge_i u_i = [x_i]_{E_i} \right).$$

(ii) For a formula $\rho(u_0, \ldots, u_{m-1})$, we let $[\rho] \subseteq \mathrm{S}_E(T)$ be the corresponding subset, and $\rho^*(x_0, \ldots, x_{m-1})$ the expression

$$\rho([x_0]_{E_0}, \ldots, [x_{m-1}]_{E_{m-1}}).$$

Lemma 3.6. *With the hypotheses of Notation 3.5, the expressions ψ_E and ρ^* are (equivalent to) formulas, and*

$$\chi_E([\psi]) = \chi_E([\psi] \cap \mathcal{P}_m) = [\psi_E] \subseteq \mathrm{S}_E(T), \qquad \chi_E^{-1}([\rho]) = [\rho^*] \subseteq \mathrm{S}_{m\Phi}(T).$$

Proof. We may fix n big enough that only $x_{<m,<n}$ appear freely in ψ, and only $x_{i,<n}, y_{i,<n}$ appear freely in E_i. We may then identify D_{E_i} with $D_{\Phi,n}/E_i$. We may now replace the x_i in the definition of ψ_E with variables in $D_{\Phi,n}$, showing that ψ_E is a formula. For ρ^* this is even more direct, producing a formula with free variables $x_{<m,<n}$.

Clearly, $\chi_E^{-1}([\rho]) = [\rho^*]$ and $\chi_E([\psi]) \subseteq [\psi_E]$. On the other hand, if $\psi_E(b_0, \ldots, b_{m-1})$ holds, then each b_i is of the form $[a_{i,<n}]_{E_i}$, where $a_{i,<n} \in D_{\Phi,n}$. Letting all happing in some countable model $M \vDash T$, each $a_{i,<n}$ can be extended to $a_i \in D_\Phi$ that enumerates M, so $p = \mathrm{tp}(a_0, \ldots, a_{n-1}) \in [\psi] \cap \mathcal{P}_m$ and $\chi_E(p) = \mathrm{tp}(b_0, \ldots, b_{m-1})$ It follows that $[\psi_E] \subseteq \chi_E([\psi] \cap \mathcal{P}_m)$, concluding the proof. ■

Lemma 3.7. *The maps χ_E and χ'_E are continuous, open and surjective.*

Proof. The map χ_E is clearly continuous, and so is its restriction. By definition of $\mathrm{S}_{m\Phi}(T)$, the range of χ_E, and *a fortiori* of its restriction to $\mathcal{P}_m$, lies inside $\mathrm{S}_E(T)$. Applying Lemma 3.6 to a basic open subset of $\mathrm{S}_{m\Phi}(T)$ (defined by a formula), we see that both maps are open. Applying the same when ψ is the tautology, and so is ψ_E, we see that the maps are surjective. ■

Lemma 3.8. *Let $E = \prod E_i$ be an m-fold definable equivalence relation on D_Φ.*

(i) *Relative to $\mathcal{G}^m$, E defines a clopen sub-groupoid $\mathcal{H}_E = \prod \mathcal{H}_{E_i} \in \mathscr{H}_m$.*
(ii) *With the hypotheses of Notation 3.5:*

$$\chi_E'^{-1}([\psi_E]) = [(\psi_E)^*] \cap \mathcal{P}_m = \mathcal{H}_E([\psi] \cap \mathcal{P}_m) = \mathcal{H}_E[\psi] \cap \mathcal{P}_m,$$

where the right hand side is to be interpreted in the sense of the action $\mathrm{S}_{m\Phi}(T) \curvearrowleft \mathcal{G}^m$.

Proof. The first item is immediate. In the second item, the first and last equalities are immediate, as is the inclusion $[(\psi_E)^*] \cap \mathcal{P}_m \supseteq \mathcal{H}_E[\psi] \cap \mathcal{P}_m$. For the opposite inclusion, let n be as in the proof of Lemma 3.6, and say that $p = \mathrm{tp}(a_0, \ldots, a_{m-1}) \in [(\psi_E)^*] \cap \mathcal{P}_m$. Then all the sequences a_i enumerate some model $M \vDash T$, and for each i, there exist $b_{i,<n} \in [a_{i,<n}]_{E_i}$ such that $\psi(b_{0,<n}, \ldots, b_{m-1,<n})$ holds. We may take the finite tuples $b_{i,<n}$ to lie inside M, and by Lemma 2.4, we may extend each of them to $b_i \in D_\Phi$ that enumerates M. Let

$$q = \mathrm{tp}(b_0, \ldots, b_{m-1}), \qquad g_i = \mathrm{tp}(a_i, b_i), \qquad \text{and} \qquad g = (g_0, \ldots, g_{m-1}).$$

Then

$$q \in [\psi] \cap \mathcal{P}_m, \qquad g \in \mathcal{H}_E, \qquad \text{and} \qquad p = gq \in \mathcal{H}_E[\psi] \cap \mathcal{P}_m.$$

This concludes the proof. ■

Proposition 3.9. *Let E be an m-fold definable equivalence relation on D_Φ. Then the map $\rho \mapsto [\rho^*] \cap \mathcal{P}_m$ is a bijection between formulae $\rho(u_0, \ldots, u_{m-1})$ in the sort $D_E = D_{E_0} \times \cdots \times D_{E_{m-1}}$ (up to equivalence) and $\mathcal{H}_E$-invariant clopen $X \subseteq \mathcal{P}_m$.*

Proof. Clearly, if ρ is a formula in D_E, then $X = [\rho^*] \cap \mathcal{P}_m$ is clopen in $\mathcal{P}_m$ and $\mathcal{H}_E$-invariant. By Lemma 3.6, $\chi'_E(X) = [(\rho^*)_E] = [\rho]$, so the map is injective.

Now, let $X \subseteq \mathcal{P}_m$ be any clopen $\mathcal{H}_E$-invariant set, and let $Y = \chi'_E(X)$. If $V \subseteq X$ is any basic open subset of $\mathcal{P}_m$ (i.e., defined by a formula), and $W = \chi'_E(V)$, then $\chi_E'^{-1}(W) = \mathcal{H}_E V \subseteq X$, by Lemma 3.8. It follows that $\chi_E'^{-1}(Y) = X$. On the other hand, if $X' = \mathcal{P}_m \smallsetminus X$ and $Y' = \chi'_E(X')$, then X' is $\mathcal{H}_E$-invariant as well, so $\chi_E'^{-1}(Y') = X'$. It follows that Y and Y' are disjoint open sets, so Y is clopen in $\mathrm{S}_E(T)$, and therefore equal to $[\rho]$ for some formula $\rho(u_0, \ldots, u_{m-1})$. It follows that $[\rho^*] \cap \mathcal{P}_m = \chi_E'^{-1}(Y) = X$. This proves that our map is surjective. ■

Corollary 3.10. (i) *Let $X \subseteq \mathcal{P}_m$. Then there exists a formula $\psi(x_0, \ldots, x_{m-1})$ such that $X = \mathcal{P}_m \cap [\psi]$ if and only if X is clopen, and there exists a neighbourhood U of $\mathcal{B}$ such that $X \cdot U^m = X$ (here $U^m \subseteq \mathcal{G}^m$ is the Cartesian power, and not iterated multiplication).*
(ii) *In particular, $\mathcal{H} \subseteq \mathcal{G}$ is the trace of a definable equivalence relation on D_Φ if and only if $\mathcal{H} \in \mathscr{H}$.*

Proof. Let $E_n(x, y)$ be the definable equivalence relation $x_{<n} = y_{<n}$, which defines a clopen sub-groupoid $\mathcal{H}_n \subseteq \mathcal{G}$. If $X = \mathcal{P}_m \cap [\psi]$, where only $x_{<m,<n}$ are free in ψ, then X is clopen in $\mathcal{P}_m$ and $\mathcal{H}_n^m \cdot X = X$. Conversely, assume that X is clopen and $X \cdot U^m = X$ for some neighbourhood U of $\mathcal{B}$. By compactness of $\mathcal{B}$, there exists n such that $U \supseteq \mathcal{H}_n$, so $\mathcal{H}_n^m \cdot X = X$. By Proposition 3.9, X is defined, relative to $\mathcal{P}_m$, by a formula (with free variables among $x_{<m,<n}$).

For the second item, let us use the fact that $\mathcal{G}$ can be identified with $\mathcal{P}_2$, and the action $\mathcal{G}^2 \curvearrowright \mathcal{P}_2$ with the pair of actions $\mathcal{G} \curvearrowright \mathcal{G} \curvearrowleft \mathcal{G}$ (the right action by $\mathcal{G}$ is a left action by the opposite groupoid $\mathcal{G}^{\mathrm{op}}$, which is isomorphic to $\mathcal{G}$). Assume that $\mathcal{H} \subseteq \mathcal{G}$ is clopen sub-groupoid and contains $\mathcal{B}$. Then it is an open neighbourhood of $\mathcal{B}$, and $\mathcal{H} \cdot \mathcal{H} \cdot \mathcal{H} = \mathcal{H}$ in $\mathcal{G}$. In terms of the action $\mathcal{G}^2 \curvearrowright \mathcal{P}_2$, this translates to $\mathcal{H}^2 \cdot \mathcal{H} = \mathcal{H}$, so $\mathcal{H}$ is defined by a formula $E(x, y)$. The relation E is reflexive since $\mathcal{H}$ contains $\mathcal{B}$, it is symmetric since $\mathcal{H}$ is closed under inverses, and transitive since $\mathcal{H}$ is closed under products. It is therefore an equivalence relation. ■

This allows us to recover the sorts and language. The last step is to recover the actual structures.

Lemma 3.11. *Let $\mathcal{G} = \mathcal{G}_\Phi(T)$ as usual. Let $a \in D_\Phi$ enumerate some countable model $M \vDash T$, let $e = \mathrm{tp}(a) \in \mathcal{B}$, and let E be a definable equivalence relation on Φ. Then the sets $D_E(M)$ (members of the sort D_E in M) and $\{\mathcal{H}_E g : g \in \mathcal{G}e\}$ (where $\mathcal{G}e = s^{-1}(e)$) are in a natural bijection, which sends $[b]_E$, where $b \in D_\Phi$ enumerates M, to $\mathcal{H}_E\, \mathrm{tp}(b, a)$.*

Proof. We have already seen that $D_E(M)$ consists of all $[b]_E$, where $b \in D_\Phi$ enumerates M (possibly with repetitions). If b is such and $g = \mathrm{tp}(b,a)$, then $g \in \mathcal{G}e$. If $c \in D_\Phi$ is yet another enumeration of M, with $f = \mathrm{tp}(c,a)$, then

$$[b]_E = [c]_E \iff E(b,c) \iff gf^{-1} = \mathrm{tp}(b,c) \in \mathcal{H}_E \iff \mathcal{H}_E g = \mathcal{H}_E f.$$

This shows that the map is well defined and injective. Since every $g \in \mathcal{G}e$ can be written (in a unique fashion) as $\mathrm{tp}(b,a)$, where $b \in D_\Phi$ enumerates M, the map is surjective. ■

Theorem 3.12. *Let T be a classical theory and $\mathcal{G} = \mathcal{G}(T)$. Then $T(\mathcal{G})$, as constructed in Definition 3.1, is bi-interpretable with T. Up to a change of language, its sorts consist of all interpretable sorts in T, with the full induced structure.*

In particular, if T' is another theory and $\mathcal{G}(T) \cong \mathcal{G}(T')$, then T and T' are bi-interpretable.

Proof. We have already seen that the sorts of $\mathcal{L}(\mathcal{G})$ are exactly those to interpreted sorts in T, and that the predicate symbols in $\mathcal{L}(\mathcal{G})$, on these sorts, correspond to the full induced structure on the corresponding interpreted sorts of T. This gives an interpretation of an $\mathcal{L}(\mathcal{G})$-theory in T. We have also seen that if $e = \mathrm{tp}(a) \in \mathcal{B}$, where a enumerates M, then M_e is the $\mathcal{L}(\mathcal{G})$-structure interpreted, in this fashion, in M. Since every countable model of T can be enumerated in this fashion, the interpreted $\mathcal{L}(\mathcal{G})$-theory is indeed $T(\mathcal{G})$.

Each sort of the language of T is in a definable bijection with some D_E, and therefore appears as a sort of $T(\mathcal{G})$: just choose any n such that x_n is in the desired sort and φ_n is a tautology, and let E be the relation $x_n = y_n$. It follows that T is bi-interpretable with $T(\mathcal{G})$ (in one direction we add interpreted sorts, in the other we forget them).

Finally, every interpreted sort of T can be adjoined to T, yielding a theory T', which is bi-interpretable with T. In this case $\mathcal{G} = \mathcal{G}(T) = \mathcal{G}(T')$, so all interpreted sorts of T appear as sorts of $T(\mathcal{G})$. ■

4. A GROUPOID WHICH MAY BE ASSOCIATED TO A THEORY IN CONTINUOUS LOGIC

We can, without much trouble, repeat the construction of $\mathcal{G}_0(T)$ for a continuous theory, i.e., a theory in the sense of continuous logic [BU10, BBHU08]. By a *dense enumeration* of a topological (or metric) set let us mean an enumeration of a dense subset.

Definition 4.1. Let T be a continuous first-order theory in a countable language. We define $\mathcal{G}_0(T) \subseteq \mathrm{S}_{2\times\mathbf{N}}(T)$ to consist of all possible types of a pair of dense enumerations of a model of T (i.e., any two dense enumerations of any single separable model). Everything else can be defined as in Definition 2.1.

In particular, $\mathcal{G}_0(T)$ is an open topological groupoid (by pretty much the same argument as for Lemma 2.2). Identifying a type $\mathrm{tp}(a,a)$ with $\mathrm{tp}(a)$ (as in Section 2), we identify the base of $\mathcal{G}_0(T)$, denoted $\mathcal{B}_0(T)$, with a subset of $\mathrm{S}_{\mathbf{N}}(T)$ consisting of all types of dense enumerations of models of T.

One may now try to repeat the definition of $\mathcal{G}(T)$ by taking a sufficiently rich sequence Φ of formulas $\varphi_n(x_{<n}, y)$, and restrict the base to those types which satisfy for all n

$$\varphi_n(x_{<n}, x_n) = \inf_y \varphi_n(x_{<n}, y). \tag{1}$$

However, unlike the situation in classical logic, any such condition does not necessarily define a definable set in the sense of continuous logic (i.e., the logic is not closed under quantification over sequences satisfying one, or all, of these conditions). The resulting groupoid need not be open, and our argument from Section 2 that it does not depend on Φ fails. If we allow any fixed error in (1), the same problems persist. One possible solution, which we pursue here, is to allow a variable error in (1).

Let us first see how we construct the base set.

Definition 4.2. Let X be a set and $n \geq 1$. Let $J_n = [0,1)^n \cup \{1\}^n \subseteq [0,1]^n$, and define $F_n X$ (F stands for "fan") to be the space

$$(X \times J_n)/\sim,$$

where $(x,\alpha) \sim (y,\beta)$ if and only if equality holds or $\alpha = \beta = 1$ (and then x and y may be arbitrary). Points of $F_n X$ will be denoted $[x,\alpha]$, and the point $[x,1]$ (which does not depend on x) will also be denoted $*$. We also define a map $\lambda\colon F_n X \to [0,1]$, sending $[x,\alpha] \mapsto \alpha$.

When X is a topological space, we equip $F_n X$ with the minimal topology in which $\{*\}$ is closed, and both maps $F_n X \smallsetminus \{*\} \to X$ and $\lambda\colon F_n X \to [0,1]^n$ are continuous.

If $f\colon X \to Y$ is a map, we define $F_n f\colon F_n X \to F_n Y$ by $F_n f([x,\alpha]) = [f(x),\alpha]$.

If $A \subseteq X$ and $B \subseteq [0,1]^n$, let $[A \times B] = (A \times (B \cap J_n))/\sim \subseteq F_n X$.

When $n = 1$, we have $J_n = [0,1]$, and we drop the subscript n everywhere.

The following is immediate.

Fact 4.3. (i) *Given a basis for the topology of X, the topology of $F_n X$ admits a basis consisting of $[U \times V]$ where $V \subseteq [0,1]$ is open, and*
- *either $U \subseteq X$ is basic open and $V \subseteq [0,1-\varepsilon]^n$ for some $\varepsilon > 0$,*
- *or $U = X$.*

(ii) *If X is compact, then the topology on $F_n X$ is the quotient topology (otherwise it merely weaker than the quotient topology), and FX (i.e., $F_1 X$) is compact.*
(iii) *If X is Hausdorff, or completely regular, then so is $F_n X$.*
(iv) *If $f\colon X \to Y$ is continuous or open, then so is $F_n f\colon F_n X \to F_n Y$. In particular, F_n is functor of topological spaces.*
(v) *Given two spaces X and Y, the space $F_n(X \times Y)$ is naturally homeomorphic to the fibred product $F_n X \times_\lambda F_n Y$, via $[x,y,\alpha] \mapsto \big([x,\alpha],[y,\alpha]\big)$.*

Definition 4.4. Assume T is a theory in continuous logic. Fix an enumeration of $[0,1]$-valued formulas $\Phi = \big(\varphi_n(x_{<n},y) : n \in \mathbf{N}\big)$, such that every $[0,1]$-valued formula of the form $\varphi(x_{<k},y)$ can be approximated arbitrarily well by φ_n for $n \geq k$ (with dummy variables). Define

$$\varphi'_n(x_{\leq n}) = \varphi_n(x_{<n},x_n) - \inf_y \varphi_n(x_{<n},y), \qquad \varphi''_n(x_{<n}) = \bigvee_{k<n} \varphi'_k, \qquad \varphi''(x) = \bigvee_{k\in\mathbf{N}} \varphi'_k.$$

Each φ''_n is a formula, while φ'' is a construct that has some truth value, defining a lower semi-continuous function on $\mathrm{S}_{\mathbf{N}}(T)$ (but it is not necessarily even a formula in continuous $\mathcal{L}_{\omega_1,\omega}$, as defined in Ben Yaacov and Iovino [BI09]). We define

$$\mathcal{B}^F_\Phi(T) = \big\{[p,\alpha] \in F\,\mathrm{S}_{\mathbf{N}}(T) : \varphi''(p) \leq \alpha\big\}.$$

As in Section 2, assume that T has no singleton models.

Lemma 4.5. *As defined above, $\mathcal{B}^F_\Phi(T) \subseteq F\mathcal{B}_0(T)$ is compact. Moreover, for $0 \leq \alpha < 1$ let*

$$S_\alpha = \lambda^{-1}_{\mathcal{B}^F_\Phi(T)}(\alpha) = \big\{p \in \mathrm{S}_{\mathbf{N}}(T) : [p,\alpha] \in \mathcal{B}^F_\Phi(T)\big\}.$$

Then each S_α is compact and totally disconnected, and the family (S_α) is increasing and right-continuous ($S_\alpha = \bigcap_{\beta>\alpha} S_\beta$).

Proof. Compactness is easy, and $\mathcal{B}^F_\Phi(T) \subseteq F\mathcal{B}_0(T)$ holds by the Tarski-Vaught test. For the moreover part, we need only check that S_α is totally disconnected. Let $p,q \in S_\alpha$ be distinct, so there exists a formula $\chi(x_{<n})$ such that $\chi(p) = 0$ and $\chi(q) = 1$.

Since T has no singleton models, the diameter of every model of T is at least some $r > 0$. By Urysohn's Lemma, there exists a continuous function $t\colon [0,1]^2 \to [0,1]$ such that:
- $t(u,v)$ vanishes on $\{0\} \times [0,1/2]$ and on $[r/2,1] \times [1/2,1]$,
- and equals one on $\{r/4\} \times [0,1]$, on $[0,r/4] \times \{1\}$, and on $[r/4,1] \times \{0\}$.

Such a function exists by Urysohn's Lemma. Consider the formula $\psi(x_{<n},y) = t\big(d(x_0,y),\chi(x_{<n})\big)$. Let $0 < \varepsilon < (1-\alpha)/2$, and find k such that $\sup|\varphi_k - \psi| < \varepsilon$.

It is clear that $\inf_y \psi(x_{<n},y)$ always vanishes (just take y equal to x_0 or sufficiently distant, according to the value of $\chi(x_{<n})$), so in S_α we always have

$$\psi(x_{<n},x_k) \leq \varphi(x_{\leq k}) + \varepsilon \leq \inf_y \varphi(x_{<k},y) + \alpha + \varepsilon \leq \inf_y \psi(x_{<n},y) + \alpha + 2\varepsilon < 1.$$

In particular, the condition $d(x_0,x_k) \leq r/4$ defines a clopen set in S_α, which cannot contain q. Its complement, defined by $d(x_0,x_k) \geq r/4$, cannot contain p, providing the desired separation. ■

Fact 4.6. *Any increasing sequence of metrisable compact totally disconnected spaces $X_0 \subseteq X_1 \subseteq \cdots$ can be realised as a sequence of subsets of the Cantor space $\mathcal{C}$.*

Proof. Let $\mathscr{B}_0 \twoheadleftarrow \mathscr{B}_1 \twoheadleftarrow \cdots$ be the corresponding inverse system of Boolean algebras, and let $\mathscr{A} = \varprojlim \mathscr{B}_n$. This algebra $\mathscr{A}$ need not be countable, even though each $\mathscr{B}_n$ is. However, one can choose a countable sub-algebra $\mathscr{A}_0 \subseteq \mathscr{A}$ such that for each n, the image of $\mathscr{A}_0$ under the projection $\mathscr{A} \twoheadrightarrow \mathscr{B}_n$ is all of $\mathscr{B}_n$. In other words, each $X_n = S(\mathscr{B}_n)$ can be embedded in the Stone space $S(\mathscr{A})$ in a manner compatible with the inclusions, and we can always embed $S(\mathscr{A})$ in the Cantor space. ■

Lemma 4.7. *The space $\mathcal{B}^F_\Phi(T)$ can be embedded topologically in $F\mathcal{C}$ in a manner which respects λ.*

Proof. Let $S = \bigcup_{\alpha<1} S_\alpha \subseteq \mathrm{S}_{\mathbf{N}}(T)$. By Lemma 4.5 and Fact 4.6, there exists a map $\theta \colon S \to \mathcal{C}$ which is an embedding on each S_α, giving rise to a map $F\theta \colon FS \to F\mathcal{C}$. Recall that $\mathcal{B}^F_\Phi(T) \subseteq FS$, and let us show that $F\theta$ is an embedding of $\mathcal{B}^F_\Phi(T)$ (it respects λ by construction). Since θ is injective, so is $F\theta$. Let $[U \times V] \subseteq F\mathcal{C}$ be a basic open set as per Fact 4.3, so $U \subseteq \mathcal{C}$ and $V \subseteq [0,1]$ are open, and either $U = \mathcal{C}$ or $V \subseteq [0,\alpha]$ for some $\alpha < 1$. In the first case, $(F\theta)^{-1}[\mathcal{C} \times V] = [S \times V]$ which is open. In the second case, $F\theta$ is an embedding of $\big[S_\alpha \times [0,\alpha]\big]$, and therefore in particular of $[S \times V] \cap \mathcal{B}^F_\Phi(T)$. Since the latter is an open subset of $\mathcal{B}^F_\Phi(T)$, we conclude that $(F\theta)^{-1}[\mathcal{C} \times V] \cap \mathcal{B}^F_\Phi(T)$ is open in $\mathcal{B}^F_\Phi(T)$. We have shown that $F\theta$ is continuous on the compact space $\mathcal{B}^F_\Phi(T)$, which is enough. ■

Let us recall a few definitions and facts from Charatonik [Cha89]. The definition of a *smooth fan* is fairly technical, suffice it to say that the *Cantor fan* $F\mathcal{C}$ is a smooth fan, as well as any closed connected subset thereof. An *endpoint* of a fan F is a point $x \in F$ such that $F \smallsetminus \{x\}$ is path-connected. The set of endpoints is denoted $E(F)$. Finally, there exists a unique (up to homeomorphism) smooth fan F such that $E(F)$ is dense in F, called the *Lelek fan*.

Lemma 4.8. *The space $\mathcal{B}^F_\Phi(T)$ is homeomorphic to the Lelek fan.*

Proof. We have already seen that it is a smooth fan. Let us consider a non-empty basic open set $W = [U \times V] \cap \mathcal{B}^F_\Phi(T)$. In either of the two possibilities for a basic open set, we may assume that U is defined by a condition $\chi(x_{<n}) < 1$, and there exists $[p,\alpha] \in W$ with $\alpha < 1$. Since $\varphi''(p) \leq \alpha < 1$, we can construct q which agrees with p on $x_{<n}$ and such that $\varphi''(q) = \alpha$. Then $[q,\alpha] \in W$ is an endpoint of $\mathcal{B}^F_\Phi(T)$, proving that the endpoints are dense. ■

If $\mathcal{G}$ is a topological groupoid with base $\mathcal{B}$, we can equip $F\mathcal{G}$ with a structure of topological groupoid, with base $F\mathcal{B}$, by applying the functor F to the composition law and inverse map. More precisely, if $D \subseteq \mathcal{G}^2$ denotes the domain of composition, then $FD \subseteq F(\mathcal{G}^2) \cong F\mathcal{G} \times_\lambda F\mathcal{G} \subseteq F\mathcal{G}^2$, so $F(\cdot) \colon FD \to F\mathcal{G}$ can be viewed as a partial map $F\mathcal{G}^2 \dashrightarrow F\mathcal{G}$. The source and target maps of $F\mathcal{G}$ are Fs and Ft, respectively, and we always have $\lambda \circ Fs = \lambda \circ Ft$. We want the two values to be able to vary independently, so let us propose a slightly different construction.

Definition 4.9. Let $\mathcal{G}$ be a topological groupoid with base $\mathcal{B}$. Let us write points of $F_2\mathcal{G}$ as $[\alpha, g, \beta]$ (where $\alpha, \beta \in [0,1]$ are either both equal or both unequal to one). When gf exists, define

$$[\alpha, g, \beta][\beta, f, \gamma] = [\alpha, gf, \gamma], \qquad [\alpha, g, \beta]^{-1} = [\beta, g^{-1}, \alpha].$$

In particular, $*^2 = *^{-1} = *$.

We shall also tend to view $F\mathcal{B}$ as a subset of $F_2\mathcal{G}$, identifying $[\alpha, e]$ with $[\alpha, e, \alpha]$.

Lemma 4.10. *Assume $\mathcal{G}$ is a topological groupoid with base $\mathcal{B}$. Then the identification of $F\mathcal{B} \subseteq F_2\mathcal{G}$ is a topological embedding, and $F_2\mathcal{G}$ is a topological groupoid with base $F\mathcal{B}$. If $\mathcal{G}$ is open, then so is $F_2\mathcal{G}$.*

In particular, the source and target of $[\alpha, g, \beta]$ are $[s_g, \beta]$ and $[t_g, \alpha]$, respectively.

Proof. This is fall fairly straightforward. For openness, observe that the map $F_2\mathcal{B} \to F\mathcal{B}$ sending $[\alpha, e, \beta] \mapsto [e, \beta]$ is open. ■

Definition 4.11. Continuing Definition 4.4, and recalling that $\mathcal{B}^F_\Phi(T) \subseteq F\mathcal{B}_0(T)$, we define

$$\mathcal{G}^F_\Phi(T) = F_2\mathcal{G}_0(T)\restriction_{\mathcal{B}^F_\Phi(T)}.$$

Lemma 4.12. *As defined above, $\mathcal{G}^F_\Phi(T)$ is a open topological groupoid.*

Proof. Let us show that $s \colon \mathcal{G}^F_\Phi(T) \to \mathcal{B}^F_\Phi(T)$ is open. Indeed, it is enough to consider a basic open subset of $\mathcal{G}^F_\Phi(T)$, namely $W = [V_1 \times U \times V_2] \cap \mathcal{G}^F_\Phi(T)$, where $U \subseteq \mathcal{G}_0(T)$ and $V_i \subseteq [0,1]$ are open, and either $U = \mathcal{G}_0(T)$ or $V_i \subseteq [0, 1-\varepsilon]$. In the first case, $s(W) = \big(\mathcal{B}_0(T) \times V_2\big) \cap \mathcal{B}^F_\Phi(T)$, which is again open. In the second case, let us consider $[\alpha, g, \beta] \in W$. We may assume that U is defined by $\chi(x_{<n}, y_{<n}) < 1$, where $\chi(g) = 0$, and that $[\alpha, \gamma] \subseteq V_1$, where $\gamma = \alpha + 2^{-n}$. Consider the formula

$$\chi'(y_{<n}) = \inf_{x_{<n}} \Big[\chi(x_{<n}, y_{<n}) \vee 2^n \big(\varphi''_n(x_{<n}) - \alpha\big)\Big].$$

Let $U' \subseteq \mathcal{B}_0(T)$ be defined by $\chi' < 1$, and let $W' = [U' \times V_2] \cap \mathcal{B}^F_\Phi(T)$. Then W' is open in $\mathcal{B}^F_\Phi(T)$, and $[s_g, \beta] \in W'$. Now let $[q, \delta] \in W'$, and say $b \vDash q$. Then b is a dense enumeration of some $M \vDash T$. Since $\chi'(b_{<n}) < 1$, there exist $a_{<n}$ in M such that $\chi(a_{<n}, b_{<n}) < 1$ and $\varphi''_n(a_{<n}) < \gamma$. This can be extended to a dense enumeration a of M such that $\varphi''(a) \leq \gamma$. Let $f = \operatorname{tp}(a,b)$. Then $f \in \mathcal{G}_0(T)$, and $[\gamma, f, \delta] \in W$.

Therefore $s_{[\alpha,g,\beta]} \in W' \subseteq s(W)$, which completes the proof. ■

REFERENCES

[AF13] Steve AWODEY and Henrik FORSSELL, *First-order logical duality*, Annals of Pure and Applied Logic **164** (2013), no. 3, 319–348, doi:10.1016/j.apal.2012.10.016.

[AZ86] Gisela AHLBRANDT and Martin ZIEGLER, *Quasi-finitely axiomatizable totally categorical theories*, Annals of Pure and Applied Logic **30** (1986), no. 1, 63–82, Stability in model theory (Trento, 1984), doi:10.1016/0168-0072(86)90037-0.

[BBHU08] Itaï BEN YAACOV, Alexander BERENSTEIN, C. Ward HENSON, and Alexander USVYATSOV, *Model theory for metric structures*, Model theory with applications to algebra and analysis. Vol. 2, London Math. Soc. Lecture Note Ser., vol. 350, Cambridge Univ. Press, Cambridge, 2008, pp. 315–427, doi:10.1017/CBO9780511735219.011.

[BI09] Itaï BEN YAACOV and José IOVINO, *Model theoretic forcing in analysis*, Annals of Pure and Applied Logic **158** (2009), no. 3, 163–174, doi:10.1016/j.apal.2007.10.011, arXiv:0810.4068.

[BK16] Itaï BEN YAACOV and Adriane KAÏCHOUH, *Reconstruction of separably categorical metric structures*, Journal of Symbolic Logic **81** (2016), no. 1, 216–224, doi:10.1017/jsl.2014.80, arXiv:1405.4177.

[BU10] Itaï BEN YAACOV and Alexander USVYATSOV, *Continuous first order logic and local stability*, Transactions of the American Mathematical Society **362** (2010), no. 10, 5213–5259, doi:10.1090/S0002-9947-10-04837-3, arXiv:0801.4303.

[Cha89] Włodzimierz J. CHARATONIK, *The Lelek fan is unique*, Houston Journal of Mathematics **15** (1989), no. 1, 27–34.

ITAÏ BEN YAACOV, UNIVERSITÉ CLAUDE BERNARD – LYON 1, INSTITUT CAMILLE JORDAN, CNRS UMR 5208, 43 BOULEVARD DU 11 NOVEMBRE 1918, 69622 VILLEURBANNE CEDEX, FRANCE

URL: http://math.univ-lyon1.fr/~begnac/