From complexity to algebra and back: digraph classes, collapsibility and the PGP

Catarina Carvalho[1], Florent R. Madelaine[2]*, and Barnaby D. Martin[3]†

[1] School of Physics, Astronomy and Mathematics University of Hertfordshire
c.carvalho2@herts.ac.uk
[2] GREYC, CNRS UMR 6072 Université de Caen Basse Normandie
florent.madelaine@unicaen.fr
[3] School of Science and Technology Middlesex University, London B.Martin@mdx.ac.uk

Abstract

Inspired by computational complexity results for the quantified constraint satisfaction problem, we study the clones of idempotent polymorphisms of certain digraph classes. Our first results are two algebraic dichotomy, even "gap", theorems. Building on and extending [1], we prove that partially reflexive paths bequeath a set of idempotent polymorphisms whose associated clone algebra has: either the polynomially generated powers property (PGP); or the exponentially generated powers property (EGP). Similarly, we build on [2] to prove that semicomplete digraphs have the same property.

These gap theorems are further motivated by new evidence that PGP could be the algebraic explanation that a QCSP is in NP even for unbounded alternation. Along the way we effect also a study of a concrete form of PGP known as *collapsibility*, tying together the algebraic and structural threads from [3], and show that collapsibility is equivalent to its Π_2-restriction. We also give a decision procedure for k-collapsibility from a singleton source of a finite structure (a form of collapsibility which covers all known examples of PGP for finite structures).

Finally, we present a new QCSP trichotomy result, for partially reflexive paths with constants. Without constants it is known these QCSPs are either in NL or Pspace-complete [1], but we prove that with constants they attain the three complexities NL, NP-complete and Pspace-complete.

1 Introduction

A great literature of work exists from the past twenty years on applications of universal algebra in the computational complexity of *constraint satisfaction problems* (CSPs) and a number of celebrated results have been obtained through this method. Each CSP is parameterised by a finite structure $\mathcal{B}$ and asks whether an input sentence φ holds on $\mathcal{B}$, where φ is a primitive positive sentence, that is where only $\exists$ and $\wedge$ may be used. For almost every class of model checking problem induced by the presence or absence of first-order quantifiers and connectors, we can give a complexity classification [4]: the two

*thanks ANR grant ALCOCLAN
†thanks EPSRC grant EP/L005654/1

outstanding classes are CSPs and its popular extension *quantified CSPs* (QCSPs) for positive Horn sentences – where $\forall$ is also present – which is used in Artificial Intelligence to model non-monotone reasoning or uncertainty.

The outstanding conjecture in the area is that all finite-domain CSPs are either in P or are NP-complete, something surprising given these CSPs appear to form a large microcosm of NP, and NP itself is unlikely to have this dichotomy property. This Feder-Vardi conjecture [5], given more concretely in the algebraic language in [6], remains unsettled, but is now known for large classes of structures.

The very useful role of algebra in unlocking the computational complexities of QCSP has also been widely documented (see [7, 8]). Manuel Bodirsky has described the CSP as a *Königsproblem* (king among problems) because it is an important computational problem living at the interface of logic, combinatorics and algebra. The QCSP is a somewhat less important problem, with weaker links outside of the logical, where it is formulated. In particular, its combinatorics are unwieldy – for example a totally satisfactory notion of a core remains elusive [9] – and its algebra is complicated by the fact that the class of surjective polymorphisms is not closed under composition. This perhaps explains why the complexity of QCSPs is classified for rather modest classes of structures, for which only three complexities are observed P, NP-complete and Pspace-complete.

In the case in which only *idempotent polymorphisms* are considered – corresponding relationally to all constants being definable in $\mathcal{B}$ – some better behaviour is restored and it is mostly in this arena that we shall place ourselves. What seems to be a unifying explanation for a complexity in NP is that it suffices to check an instance φ with m universal variables for a small fraction (polynomial in m and $\mathcal{B}$) of all possible choices for these m universal variables. This property can be viewed as a special form of quantifier relativisation in the sense that it suffices to check an instance against restricted Skolem functions. This fits in well with the classification for model checking for other fragments of FO where relativisation also characterises the complexity [4].

In Hubie Chen's [10], a new traverse between algebra and QCSP was discovered. Chen's previous work in QCSP tractability largely involved the special notion of *collapsibility* [3], but in [10] this was extended to a version of the *polynomially generated powers* (PGP) property. This latter ties in with a rich literature of dichotomy ("gap") theorems on growth rate of generating sets of direct powers of algebras. The PGP properly generalises collapsibility and reveals a link to universal algebra that we explore in this paper and we might argue makes QCSP at least a *Fürstenproblem* (prince among problems).

The initial algebraic phenomenon of our study is the growth rate of generating sets for direct powers of an algebra. That is, for an algebra $\mathbb{A}$ we associate a function $f_{\mathbb{A}} : \mathbb{N} \to \mathbb{N}$, giving the cardinality of the minimal generating sets of the sequence $\mathbb{A}, \mathbb{A}^2, \mathbb{A}^3, \ldots$ as $f(1), f(2), f(3), \ldots$, respectively. We may say $\mathbb{A}$ has the *g-generating property* (g-GP for short) if $f(m) \leq g(m)$ for all m. The question then arises as to the growth rate of f and specifically regarding the behaviours constant, logarithmic, linear, polynomial and exponential. Wiegold proved in [11] that if $\mathbb{A}$ is a finite semigroup then $f_{\mathbb{A}}$ is either linear or exponential, with the former prevailing precisely when $\mathbb{A}$ is a monoid. This dichotomy classification may be seen as a gap theorem because no growth rates intermediate between linear and exponential may occur. We say $\mathbb{A}$ enjoys the *polynomially generated powers* property (PGP) if there exists a polynomial p so that $f_{\mathbb{A}} = O(p)$ and the *exponentially generated powers* property (EGP) if there exists a constant b so that $f_{\mathbb{A}} = \Omega(g)$ where $g(i) = b^i$.

The PGP implies that the bounded alternation QCSP is in NP rather than the corresponding level of the polynomial hierarchy one expects in general, provided that generators may be generated effectively, *effective PGP* in Chen's parlance. This should be clear for Π_2-sentences (quantifier prefix of the form $\forall^\star\exists^\star$) as it suffices to solve one CSP per generator, and by induction this holds for bounded alternation. Moreover, for all known examples it also holds for unbounded alternation. In particular, for known examples of finite structures, this drop is witnessed by an operation which characterises a type of collapsibility (from the so-called *singleton source*), which we shall call a *Hubie operation*. When this is present as a polymorphism, it implies a drop to NP also in the unbounded case as it may be composed in a more involved fashion suitable for working with Skolem functions, what Chen terms *reactive composition*.

Hubie Chen proved the first PGP-EGP gap theorem for polymorphism clones in [10]. Namely, let id-Pol($\mathcal{B}$)[1] be the clone of idempotent polymorphisms of a 3-element structure $\mathcal{B}$ such that id-Pol($\mathcal{B}$) does not contain a G-set as a factor[2]. Then either id-Pol($\mathcal{B}$) has PGP or it has EGP. Indeed, this result extended the previous observation of Chen that when id-Pol($\mathcal{B}$) is the clone of idempotent polymorphisms of a 2-element structure $\mathcal{B}$, then either id-Pol($\mathcal{B}$) has PGP or it has EGP. Now, k-Π_2-*collapsibility* (whose naming will be explained in the sequel) can be seen as a special form of the PGP in which the generating set for each $\mathbb{A}^m$ may be taken to be the set of m-tuples which contain the repetition of a single element from a so-called source set at least $m-k$ times, the other at most k positions being arbitrary. k-collapsibility can be seen similarly but manifests slightly differently through the already alluded to *reactive composition* of this set of m-tuples. In the 2-element case, the PGP manifests in the special form of 1-collapsibility, but already in the 3-element case there are algebras with the PGP that are not k-collapsible for any k, though no such example is known for a finite structure (*i.e.* with finitely many relations).

When a structure $\mathcal{H}$ expanded by all constants is so that QCSP($\mathcal{H}$) is Pspace-complete, then (under the complexity-theoretic assumption that NP is different from Pspace) we can assume that id-Pol($\mathcal{H}$) does not have *effective PGP* [8]. Naturally, these are the places to look to prove EGP results. The QCSP complexity classification for 3-element structures is still open, even in the idempotent case, but this paper builds

[1] We will view this as at once a set of polymorphisms on domain B and an algebra of operations over that domain.

[2] This is a technical assumption that we will not define. When there is a G-set as a factor we know the corresponding QCSP is NP-hard [6].

upon Chen's [10] motivated by the extant complexity classifications for the QCSP for partially reflexive trees in [1] and semicomplete digraphs in [2]. Thus, the complexity results lead the algebra, in contrast to the typical modus operandi.

Principal contributions

Complexity to algebra: new PGP-EGP gaps.

For partially reflexive paths we recall the notion of being quasi-loop-connected from [1], and prove the following algebraic gap.

Theorem 1. *Let $\mathcal{H}$ be a partially reflexive path. If $\mathcal{H}$ is quasi-loop-connected, then* id-Pol($\mathcal{H}$) *has the PGP. Otherwise,* id-Pol($\mathcal{H}$) *has the EGP.*

Along the way, we also characterise precisely which partially reflexive paths have only essentially unary polymorphisms.

Building upon and refining [2], we derive a second gap for semicomplete digraphs.

Theorem 2. *Let $\mathcal{H}$ be a semicomplete digraph. If $\mathcal{H}$ has at most one cycle or both a source and a sink, then* id-Pol($\mathcal{H}$) *has the PGP. Otherwise,* id-Pol($\mathcal{H}$) *has the EGP.*

The PGP: collapsibility and beyond.

We prove that when we have a sufficiently uniform form of effective PGP, based on the notion of *projective* sequences of adversaries (an adversary is a set of tuples restricting the tuple of universal variables), then we also have a drop in complexity to NP even in the unbounded case. For such sequences of adversaries, we can show that they are generating iff they are generating via reactive composition. Our proof relies on and adapts the notion of a canonical Π_2-sentence from [12]. The statement of this result, Theorem 36, is somewhat technical so we state here its concrete application to the situation of collapsibility.

Corollary 39 (Part of). Let $\mathcal{A}$ be a structure, $\emptyset \subsetneq B \subseteq A$ and $p > 0$. The following are equivalent.

(i) $\mathcal{A}$ is p-collapsible from source B.

(ii) $\mathcal{A}$ is Π_2-p-collapsible from source B.

(iii) For every m, the structure $\mathcal{A}$ satisfies a canonical Π_2-sentence with $m \cdot |A|$ universal variables.

In the case of a singleton source, which covers all known examples of collapsibility for finite structures (see also Table 1 which recalls the polymorphisms that are known to imply collapsibility), then we can refine this further as follows.

Theorem 44 (Part of). (**p-Collapsibility from a singleton source**). Let $p \geq 1$ and x be a constant in $\mathcal{A}$. The following are equivalent:

(i) $\mathcal{A}$ is p-collapsible from $\{x\}$.

(ii) $\mathcal{A}$ is Π_2-p-collapsible from $\{x\}$.

(iii) $\mathcal{A}$ models a **single** canonical Π_2-sentence which implies that $\mathcal{A}$ admits a Hubie operation as a polymorphism.

This means that we may decide p-collapsibility from a singleton source (the parameter $p > 0$ being part of the input).

Back to complexity.

As we have argued already, a uniform form of PGP like p-collapsibility might explain when a QCSP is in NP. It is natural in this context to allow constants in the structure not only because it makes things well behaved in the algebra, but also because constants are needed for the natural algorithm which consists in solving a polynomial number of CSP instances induced by replacing all but p variables by a constant. Finally, we apply our earlier results: that collapsibility coincides with its Π_2-restriction and that partially reflexive paths that are not quasi-loop-connected remain Π_2-collapsible in the idempotent case. This morphs the first dichotomy theorem of [1] (cf. Theorem 49.) to become a new trichotomy theorem. Specifically, the NL cases in the absence of constants split to become NL and NP-complete cases in the presence of constants.

Theorem 3. *Let $\mathcal{H}$ be a partially reflexive path expanded with all constants.*

Polymorphism	Arity	Collapsibility
Near unanimity (a.k.a. majority when $k = 3$) satisfies the identities $f(x, y, \ldots, y) = \ldots = f(\ldots, y, x, y \ldots) = f(y, \ldots, y, x) = y$	k	$(k-1)$-collapsibility with source $\{x\}$ for any x.
Dual discriminator. majority acting as a projection when the 3 arguments are distinct	3	1-collapsibility with source A. 2-collapsibility with source $\{x\}$ for any x.
Mal'tsev $m(x, x, y) = m(y, x, x) = y$	3	1-collapsibility with source $\{x\}$ for any x.
Hubie operation : remains surjective when any coordinate is fixed to be x	k	$(k-1)$-collapsibility with source $\{x\}$.
In particular, the case of so-called *semilattice with unit* $\{x\}$: a binary idempotent, associative and commutative polymorphism s that satisfies $s(x, y) = s(y, x) = y$ for any y.	2	1-collapsibility with source $\{x\}$.

Table 1: Some polymorphisms that imply collapsibility.

(i) *If $\mathcal{H}$ is loop-connected, then QCSP($\mathcal{H}$) is in NL.*

(ii) *Else, if $\mathcal{H}$ is quasi-loop-connected, then QCSP($\mathcal{H}$) is NP-complete.*

(iii) *Otherwise, QCSP($\mathcal{H}$) is Pspace-complete.*

Due to space restriction, many proofs have been omitted and can be found in the appendix.

2 Preliminaries

Throughout we consider only finite relational structures possibly with some constants. On first reading, the reader might prefer to assume that all constants are present, for the sake of simplicity; though we can not make this assumption in general as adding all constants may increase the complexity (compare Theorem 3 with Theorem 49). We denote by σ our base signature and hereafter unless otherwise specified, a structure will be a σ-structure. We shall denote by A the domain of a structure $\mathcal{A}$. The *canonical query*[3] of the structure $\mathcal{A}$ is the quantifier-free first-order sentence that has one variable x_a for each element a in A and a conjunction of all the positive facts of $\mathcal{A}$: *e.g.* $R(a_1, a_2, \ldots, a_r)$ holds in $\mathcal{A}$ for some r-ary symbol in σ iff this conjunction contains the conjunct $R(x_{a_1}, x_{a_2}, \ldots, x_{a_r})$. Conversely, given a conjunction of positive atoms φ, we denote by $\mathcal{D}_\varphi$ its *canonical database*, that is the structure with domain the variables of φ and whose tuples are precisely those that are atoms of φ. Let $\mathcal{A}$ and $\mathcal{B}$ be structures. A *homomorphism* h from $\mathcal{A}$ to $\mathcal{B}$ is a map from A to B such that for every relational symbol R of arity r and every r-tuple $(a_1, a_2, \ldots, a_r)$ of elements of A such that $R(a_1, a_2, \ldots, a_r)$ holds in $\mathcal{A}$ we have that $R(h(a_1), h(a_2), \ldots, h(a_r))$ holds in $\mathcal{B}$. The *product* $\mathcal{A} \otimes \mathcal{B}$ is the structure with domain $A \times B$ such that for every relational symbol R of arity r and every r-tuples $(a_1, a_2, \ldots, a_r)$ of elements of A and $(b_1, b_2, \ldots, b_r)$ of elements of B, we have that $R\big((a_1, b_1), (a_2, b_2), \ldots, (a_r, b_r)\big)$ holds in $\mathcal{A} \otimes \mathcal{B}$ iff both $R(a_1, a_2, \ldots, a_r)$ holds in $\mathcal{A}$ and $R(b_1, b_2, \ldots, b_r)$ holds in $\mathcal{B}$. A constant symbol c is interpreted in $\mathcal{A} \otimes \mathcal{B}$ as the element (a, b) where a and b are the interpretation of c in $\mathcal{A}$ and $\mathcal{B}$, respectively. We write $\mathcal{A}^k$ for the product of k copies of $\mathcal{A}$. A *k-ary polymorphism* of $\mathcal{A}$ is a homomorphism f from $\mathcal{A}^k$ to $\mathcal{A}$. We say that f is *idempotent* if for any x in A, $f(x, x, \ldots, x) = x$ holds. Equivalently, f is a polymorphism of an extension of $\mathcal{A}$ with constants symbols naming the elements of $\mathcal{A}$. Let id-Pol($\mathcal{A}$) (resp. sPol($\mathcal{A}$)) denote the set of idempotent (resp. surjective) polymorphisms of $\mathcal{A}$. A *majority* operation is a ternary operation f that satisfies the identities $f(x, x, y) = f(x, y, x) = f(y, x, x) = f(x, x, x) = x$. The *dual discriminator* (*dd*) is the particular majority that satisfies $dd(x, y, z) = x$ when x, y, z are distinct. A *Hubie operation* is a surjective k-ary operation f, on a set $A \ni x$, such that $f(x, x, \ldots, x) = x$ and $f(x, A, \ldots, A) = f(A, x, \ldots, A) = \ldots = f(A, A, \ldots, x) = A$. That is, the restriction of the operation from fixing x in each coordinate position remains

[3] We actually consider the quantifier-free part of the *canonical query*. We depart from the usual definition where an existential sentence is used, as we will often need a different prefix of quantification.

surjective. When we need to specify x, we speak of a *Hubie operation with source* x. A *positive Horn sentence* (pH-sentence for short) is a sentence of first-order logic with equality using both quantifiers $\exists$ and $\forall$ but only the logical connective $\wedge$. We will only consider pH sentences in *prenex form*, that is with all quantifiers in front. In the absence of the universal quantifier, we speak of a *primitive positive sentence* (pp-sentence for short). A Π_2-pH sentence is a pH-sentence with quantifier prefix of the form $\forall^\star\exists^\star$, that is a block of universal variables followed by a block of existential variables. Let $\mathcal{A}$ be a finite relational structure (possibly with constants). The *quantified constraint satisfaction problem* with structure $\mathcal{A}$, denoted $\mathrm{QCSP}(\mathcal{A})$, is the model-checking problem for pH-sentences over $\mathcal{A}$. That is, it takes as input a pH-sentence φ and asks whether $\mathcal{A}$ models φ. When $\mathcal{A}$ is a structure with constants naming its elements, we may write $\mathrm{QCSP}_c(\mathcal{A})$ to stress that all constants are present. Similarly, let $\mathrm{CSP}(\mathcal{A})$ denote the *constraint satisfaction problem* with structure $\mathcal{A}$ defined as above but with pp-sentences. We will denote by $\langle\mathcal{A}\rangle_{\mathrm{pH}}$ the class of relations that are interpretable in $\mathcal{A}$ via some pH-sentence.

Reading the introduction, one could be forgiven for thinking collapsibility is at once a logical property of structures and a property of algebras. Indeed, Chen [3] defines a form of collapsibility for each and shows that the algebraic form implies the logical one (a result reworded here as Theorem 26). One purpose of this paper is to tie these two definitions together and prove the converse. For formal purposes we will define collapsibility only in the logical sense. Let $\mathcal{A}$ be a structure, $B \subseteq A$ and $p \geq 0$. The structure $\mathcal{A}$ is *p-collapsible with source* B when for all $m \geq 1$, for all pH-sentences φ with m universal quantifiers, we have that $\mathcal{A} \models \varphi$ iff $\mathcal{A} \models \psi$, for **all** sentences ψ obtained by instantiating all but p universal variables of φ by some single element $x \in B$. We assume here that $\mathcal{A}$ has all constants from the source set B and will delay to § 4.1 for a more general definition where this assumption is not necessary. $\mathcal{A}$ is *collapsible with source* B if it is p-collapsible with source B for some p. We define similarly the analogous notions for the Π_2-fragment.

3 New PGP-EGP gaps

Let $[n] := \{1, \ldots, n\}$. A digraph $\mathcal{G}$ has vertex set G, of cardinality $|G|$, and edge set $E(\mathcal{G})$. Similarly, an algebra $\mathbb{A}$ has domain A. For a digraph $\mathcal{H}$, the distance between two m-tuples $\overline{s} = (s^1, \ldots, s^m)$ and $\overline{t} = (t^1, \ldots, t^m) \in H^m$ is the minimal r so that there are m-tuples $\overline{z}_1 = (z_1^1, \ldots, z_1^m), \ldots, \overline{z}_{r-1} = (z_{r-1}^1, \ldots, z_{r-1}^m) \in H^m$ such that, for each $i \in [m], j \in [r-2]$, we have $E(s^i, z_1^i)$, $E(z_j^i, z_{j+1}^i)$ and $E(z_{r-1}^i, t^i)$.

3.1 Partially reflexive paths

Henceforth we consider partially reflexive paths, i.e. paths potentially with some loops (we will frequently drop the preface partially reflexive). As we are interested in idempotent polymorphisms these paths come with constants naming each of their vertices. For a sequence $\beta \in \{0,1\}^*$, of length $|\beta|$, let $\mathcal{P}_\beta$ be the undirected path on $|\beta|$ vertices such that the i^{th} vertex has a loop iff the i^{th} entry of β is 1 (we may say that the path $\mathcal{P}$ is *of the form* β). A path $\mathcal{H}$ is *quasi-loop-connected* if it is of either of the forms

(i) $0^a 1^b \alpha$, for $b > 0$ and some α with $|\alpha| = a$, or

(ii) $0^a \alpha$, for some α with $|\alpha| \in \{a, a-1\}$.

Where a path satisfies both (i) and (ii), we use formulation (i) preferentially. A path whose self-loops induce a connected component is further said to be *loop-connected.* We will usually envisage the domain of a path with n vertices to be $[n]$, where the vertices appear in the natural order (and a good behaviour brought by the absence of self-loops of the quasi-loop connected case is exhibited in the lower numbers). The *centre* of a path is either the middle vertex, if there is an odd number of vertices, or between the two middle vertices, otherwise. The main result of this section was stated as Theorem 1.

Proof of Theorem 1. The PGP cases follow from Lemmas 4, 6 and 7. The EGP cases follow from Proposition 10. □

3.1.1 Partially reflexive paths with the PGP

The loop-connected case is well understood.

Lemma 4. *Let $\mathcal{H}$ be a partially reflexive path that is loop-connected. Then* id-Pol($\mathcal{H}$) *has the PGP.*

Proof. $\mathcal{H}$ admits a majority polymorphism (see Lemma 3 of [1]). This is a Hubie polymorphism of $\mathcal{G}$ (where the single element can be chosen arbitrarily), whereupon the result follows from [3] (see our forthcoming Lemma 42 together with Corollary 39). □

The quasi-loop connected case is more technical. Due to space restriction, we will only present in full half of this case, which will suffice to illustrate the proof principle. First, we are able to exhibit specific binary idempotent polymorphisms.

Lemma 5. *Let $\mathcal{P}_{0^a1^b\alpha}$, with $b > 0$, be a quasi-loop-connected path on vertices $[n]$. For each $y \in [n]$ there is a binary idempotent polymorphism f_y of $\mathcal{P}_{0^a1^b\alpha}$ so that $f_y(1,x) = x$ (for all x) and $f_y(n,1) = y$.*

Next, we exhibit specific linear generating set for the powers.

Lemma 6. *Let $\mathcal{P}_{0^a1^b\alpha}$, for $b > 0$, be a quasi-loop-connected path on vertices $[n]$. Let $\mathbb{A}$ be the algebra specified by* id-Pol($\mathcal{P}_{0^a1^b\alpha}$)*. For each m, $\mathbb{A}^m$ is generated from the $n+1$ m-tuples* $(1,1,\dots,1)$, $(n,1,\dots,1), (1,n,\dots,1), \dots, (1,1,\dots,n)$.

Proof. We will make use of the polymorphisms f_y guaranteed to exist by Lemma 5. Firstly, from $(n, 1, \ldots, 1)$ and $(1, 1, \ldots, 1)$ we can, for each y, use f_y to generate $(y, 1, \ldots, 1)$. And we can similarly build all co-ordinate permutations of this. We now have the base case in an inductive proof, where our inductive hypothesis will be that for all k we can build the tuple which has entries $y_1, \ldots, y_k$ with the remaining entries being 1. The result for $k = m$ implies the lemma, so it remains only to test the inductive step where we will assume $y_1, \ldots, y_k, y_{k+1}$ are the first $k+1$ entries of a tuple continued by $1, \ldots, 1$ (of course we can build the rest through co-ordinate permutation). From $(1, \ldots, 1, n, 1, \ldots, 1)$ and $(y_1, \ldots, y_k, 1, \ldots, 1)$ (where n is in the $k+1$st position) we can use $f_{y_{k+1}}$ to build $(y_1, \ldots, y_k, y_{k+1}, 1, \ldots, 1)$. This proves the claim. □

Lemma 5 fails for the other type of quasi-loop-connected paths, essentially when $b = 0$. This is easily seen to be the case when we take an irreflexive path on an odd number n of vertices (for an example on paths with an even number of vertices ≥ 4, take an irreflexive path leading to a single looped vertex at the end). Then no idempotent polymorphism f may have $f(n, 1) = 2$ for parity reasons, since odd and even vertices must be at odd distance in the square of the graph. In fact, Lemma 5 does hold for quite a few of the remaining cases (e.g. for $\mathcal{P}_{0^a\alpha}$ when $|\alpha| = a$ and the first entry of α is 1), but the proof requires an alternative construction. This alternative construction and a proof in the spirit of that of Lemma 6 yields the following result which deals at once with all the outstanding cases.

Lemma 7. *Let $\mathcal{P}_{0^a\alpha}$, for $|\alpha| \in \{a, a-1\}$, be a quasi-loop-connected path on vertices $[n]$ (that is not of the form $\mathcal{P}_{0^a1^b\alpha}$ with $|\alpha| = a$). Let $\mathbb{A}$ be the algebra specified by* id-Pol$(\mathcal{P}_{0^a\alpha})$. *For each m, $\mathbb{A}^m$ is generated from the $2n+2$ m-tuples* $(1, 1, \ldots, 1)$, $(2, 2, \ldots, 2), (n, 1, \ldots, 1), (1, n, \ldots, 1), \ldots, (1, 1, \ldots, n), (n, 2, \ldots, 2), (2, n, \ldots, 2), \ldots, (2, 2, \ldots, n)$.

We remark that if we were not in the idempotent situation (i.e. without constants in the structure) then the lemmas could have been proved from observations about the so-called Q-core [9] via the main result of [12] (see Application 41).

3.1.2 Partially reflexive paths with the EGP

By induction on the arity, we prove the following.

Lemma 8. *Let α be any sequence of zeros and ones. All idempotent polymorphisms of $\mathcal{P}_{10\alpha01}$ are projections.*

This will suffice to derive EGP for all non-quasi loop connected graphs as we will be able to pinpoint a suitable copy of $\mathcal{P}_{10\alpha01}$ in all such graphs. But first we need to appeal to another ingredient, namely the well-known Galois correspondence $\mathrm{Inv}(\mathrm{sPol}(\mathcal{B})) = \langle \mathcal{B} \rangle_{\mathrm{pH}}$ holding for finite structures $\mathcal{B}$ [13], which can be used to derive the following.

Corollary 9. *Suppose* $\mathbb{A} = \text{id-Pol}(\mathcal{B})$*, for some finite structure* $\mathcal{B}$*, and* Γ *is a generating set for* $\mathbb{A}^m$*. Let* $\varphi(v_1, \dots, v_m)$ *be a formula from* $\langle \mathcal{B} \rangle_{\text{pH}}$*. If* $\mathcal{B} \models \varphi(x_1, \dots, x_m)$ *for all* $(x_1, \dots, x_m) \in \Gamma$*, then* $\mathcal{B} \models \forall v_1, \dots, v_m \; \varphi(v_1, \dots, v_m)$*.*

We are now ready to conclude our proof of the PGP/EGP gap for p.r. paths and establish EGP for the remaining cases.

Proposition 10. *Let* $\mathcal{G}$ *be a p.r. path that is not quasi-loop connected. Then* id-Pol($\mathcal{G}$) *has the EGP.*

Proof. Number the vertices of $\mathcal{G}$ left-to-right over $[n]$ and let p be the leftmost loop and let q be the rightmost loop. Since $\mathcal{G}$ is not quasi-loop connected, p will be to the left of the centre and q will be to the right of centre. Let μ be $\max\{p, n-q, \lfloor \frac{q-p-1}{2} \rfloor\}$. Let P and Q be the sets of vertices at distance $\leq \mu$ from p and q, respectively.

A word $\tau \in ((P \setminus Q) \cup (Q \setminus P))^m$ is a *cousin* of a word $\sigma \in \{p, q\}^m$ if τ can be obtained by some local substitutions of $p \mapsto x \in (P \setminus Q)$ and $q \mapsto y \in (Q \setminus P)$. A word $\tau \in G^m$ is a *friend* of a word $\sigma \in \{p, q\}^m$ if τ can be obtained by some local substitutions of $p \mapsto 1, \dots, p, \dots, p + \mu$ and $q \mapsto q - \mu, \dots, q, \dots, n$. The relations friend and cousin are symmetric. If $\max\{p, n-q\} > q - p - 1$ then a situation can arise in which all words $\{p, q\}^m$ are friends of each other (this will not be a problem). However, it is not hard to see that every word in G^m has a friend in $\{p, q\}^m$ and one can walk to this friend pointwise in at most μ steps. Further,

$$(\dagger) \begin{cases} \cdot \text{ each word in } ((P \setminus Q) \cup (Q \setminus P))^m \text{ has a unique cousin in } \{p,q\}^m\text{; and,} \\ \cdot \text{ every word in } G^m \setminus ((P \setminus Q) \cup (Q \setminus P))^m \text{ has more than one friend in } \{p,q\}^m. \end{cases}$$

Note that it is possible that $G^m \setminus ((P \setminus Q) \cup (Q \setminus P))^m$ is empty. So let m be given and suppose there exists a generating set Γ for G^m of size $< 2^m$. It follows from ($\dagger$) that, for some $\tau \in \{p, q\}^m$, Γ omits τ and all of τ's cousins (though it may contain some of τ's non-cousin friends). We will prove that Γ does not generate G^m, by assuming otherwise and reaching a contradiction using Corollary 9. Let R_Γ be the subset of $\{p, q\}^m$ induced by $\{p, q\}^m \setminus \{\tau\}$. Note

$$(*) \quad \text{that every element } \sigma \in \Gamma \text{ has a friend in } R_\Gamma.$$

Note also that R_Γ is pp-definable since $\{p, \dots, q\}$ is pp-definable and all polymorphisms of the induced sub-structure given by$\{p, \dots, q\}$ are projections (this was Lemma 8).

Consider the pH-formula $\varphi(x_1, \dots, x_n) :=$

$$\exists x_1^1, \dots, x_1^{\mu-1}, \dots\dots, \exists x_n^1, \dots, x_n^{\mu-1} R_\Gamma(x_1', \dots, x_n') \wedge \Big(\bigwedge_{i \in [n]} E(x_i, x_i^1) \wedge E(x_i^1, x_i^2) \wedge \dots \wedge E(x_i^{\mu-2}, x_i^{\mu-1}) \wedge E(x_i^{\mu-1}, x_i^{\mu-1}) \Big).$$

The sentence $\forall x_1, \ldots, x_n\ \varphi(x_1, \ldots, x_n)$ is false and can be witnessed as false by taking $\overline{x}$ to be that word in $\{1, n\}^m$ derived from τ by substituting $p \mapsto 1$ and $q \mapsto n$. However, consider now that $\varphi(y_1, \ldots, y_n)$ is true for all $(y_1, \ldots, y_n) \in \Gamma$, precisely because of property $(*)$, i.e. when $(x_1, \ldots, x_n)$ is evaluated as σ, choose $(x_1^{\mu-1}, \ldots, x_n^{\mu-1})$ to be evaluated as σ's friend in R_Γ. $\Box$

3.2 Semicomplete digraphs

Recall that a digraph $\mathcal{G}$ is *semicomplete* if it is irreflexive and for each $x \neq y \in G$ we have either $E(x, y)$ or $E(y, x)$, or both. We will often abuse of the substantive and speak of semicompletes rather than semicomplete graphs. If we always have precisely one of $E(x, y)$ or $E(y, x)$, then the digraph is additionally a *tournament*. In a digraph, a *source* (resp., *sink*) is a vertex of in-degree (resp., out-degree) zero. A digraph is *smooth* if it has neither a source nor a sink. For a digraph $\mathcal{G}$ we define $\mathcal{G}^+$ to be $\mathcal{G}$ augmented with a new sink to which all other vertices have a directed edge. Let y^- be the set $\{x \in G : E(x, y) \in \mathcal{G}\}$ and y^+ be the set $\{x \in G : E(y, x) \in \mathcal{G}\}$. In the sequel we use the notation $x_j^{i'}$ to indicate the prime of x_j^i (i.e., the prime does not modify just the i).

The main result of this section is the gap theorem stated as Theorem 2.

Proof of Theorem 2. The PGP cases follow from Propositions 11 and 12. The EGP cases follow from Corollary 24. $\Box$

3.2.1 Semicomplete graphs with the PGP

Proposition 11. *Let $\mathcal{G}$ be a semicomplete graph with exactly one cycle and either a source or a sink, or none, then* id-Pol($\mathcal{G}$) *has the PGP.*

Proof. If $\mathcal{G}$ has neither source nor sink, then it is either the directed 3-cycle $\mathcal{DC}_3$ or $\mathcal{K}_2$. Let $\mathbb{A}$:=id-Pol($\mathcal{DC}_3$) or id-Pol($\mathcal{K}_2$). Both of these have the dual discriminator for a polymorphism which witnesses, for each a in the domain, that $\mathbb{A}^m$ can be generated from tuples, for all $x \in A$, of the form $(a, a, \ldots, a)$, $(x, a, \ldots, a)$, $(a, x, \ldots, a)$, $\ldots$, $(a, a, \ldots, x)$ (this latter appears in [3]).

Let us suppose $\mathcal{G}$ has a sink but no source (the alternative being a symmetric proof). Then $\mathcal{G}$ was built from $\mathcal{DC}_3$ or $\mathcal{K}_2$ by the iterative addition of sinks $t_1, \ldots, t_k$, where t_k is the sink of $\mathcal{G}$. Define $f(x, y, z)$ to be the ternary operation on $\mathcal{G}$ that acts as dual discriminator in the subgraph $\mathcal{DC}_3$ or $\mathcal{K}_2$ and returns the element t_i with the highest index i whenever the triple (x, y, z) contains an element from $\{t_1, \ldots, t_k\}$. It is straightforward to verify that f is a polymorphism of $\mathcal{G}$. Further, it is a Hubie polymorphism as is witnessed by any element z in the subgraph $\mathcal{DC}_3$ or $\mathcal{K}_2$; that is $f(z, G, G) = f(G, z, G) = f(G, G, z) = G$. The result follows from [3] (that we will quote as Lemma 42). $\Box$

Proposition 12. *Let $\mathcal{G}$ be a semicomplete graph with both a source and a sink, then* id-Pol($\mathcal{G}$) *has the PGP.*

Proof. We will give a Hubie polymorphism of $\mathcal{G}$ whereupon the result follows from [3] (that we will quote as Lemma 42).

Let x, y, z be elements of $\mathcal{G}$ distinct from s and t which are the source and sink, respectively, of $\mathcal{G}$. Define the ternary operation f so that $f(\{\{x, s, t\}\}) = x$ (we use multiset notation to indicate any coordinate permutation) extended as a projection on its first coordinate otherwise (e.g. $f(s, t, s) = f(s, t, t) = s$ and $f(x, y, z) = x$). It is easy to see this is a polymorphism, once one notes that in $\mathcal{G}^3$ all vertices of the form $\{\{x, s, t\}\}$ are isolated. Furthermore, f is a Hubie operation in both the single elements s and t. □

We will shortly need to talk about variables that are indexed individually over two dimensions and use overbar to denote columns (top index vary) and underbar to denote rows (bottom index vary). Suppose id-Pol($\mathcal{A}$) has the $f(m)$-GP. Then we are saying, for each $m \in \mathbb{N}$, that there exist $k = f(m)$ tuples $\overline{x}_1 = (x_1^1, x_1^2, \ldots, x_1^m)$, $\ldots$, $\overline{x}_k = (x_k^1, x_k^2, \cdots, x_k^m)$ so that, for each $\underline{y} = (y^1, y^2, \ldots, y^m)$ there is a k-ary polymorphism $f_{\underline{y}}$ of $\mathcal{A}$ so that

$$\underline{y} = (y^1, y^2, \ldots, y^m) = (f_{\underline{y}}(x_1^1, \ldots, x_k^1), \ldots, f_{\underline{y}}(x_1^m, \ldots, x_k^m)).$$

This can be presented by the following picture for $f := f_{\underline{y}}$,

$$\begin{array}{cccc} f & f & \cdots & f \\ \frown & \frown & \cdots & \frown \\ x_1^1 & x_1^2 & \cdots & x_1^m \\ x_2^1 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & & \vdots \\ x_k^1 & x_k^2 & \cdots & x_k^m \\ \smile & \smile & \cdots & \smile \\ \| & \| & & \| \\ y^1 & y^2 & \cdots & y^m \end{array}$$

which indicates that f is a homomorphism from $(\mathcal{A}^k; \underline{x}^1, \ldots, \underline{x}^m)$ to $(\mathcal{A}; y^1, \ldots, y^m)$. It follows of course that all pp-formulas that are true on $(\mathcal{A}^k; \underline{x}^1, \ldots, \underline{x}^m)$ are also true on $(\mathcal{A}; y^1, \ldots, y^m)$.

The following well-known model-theoretic lemma is in some sense trivial for finite structures.

Lemma 13. *Let $\mathcal{A}$ and $\mathcal{B}$ be finite structures. If all pp-sentences that are true $(\mathcal{A}; a_1, \ldots, a_m)$ are true on $(\mathcal{B}; b_1, \ldots, b_m)$, then there is a homomorphism f from $\mathcal{A}$ to $\mathcal{B}$ so that $f(a_j) = (b_j)$ for each $j \in [m]$.*

3.2.2 Semicompletes with more than one cycle but without sources

It is known from [2] that a smooth semicomplete digraph $\mathcal{H}$ with more than one cycle has only essentially unary polymorphisms, since these are also cores we can immediately say in this case that id-Pol($\mathcal{H}$) has the EGP. What remains is to classify semicompletes with more than one cycle but without sources, and semicompletes with more than one cycle but without sinks. These situations are symmetric so we will address directly only the former. We begin with some simple results.

Lemma 14. *Let $\mathcal{G}$ be a digraph. Let* id-Pol$(\mathcal{G}^{++})$ *have the $f(m)$-GP, for some $f(m)$. Then* id-Pol$(\mathcal{G}^{+})$ *has the $f(m)$-GP.*

Proof. Let t be the sink in $\mathcal{G}^{++}$ and let t' be the sink in $\mathcal{G}^{+} \subseteq \mathcal{G}^{++}$. Let m be given and set $k = f(m)$. Let $\overline{x}_1 = (x_1^1, x_1^2, \ldots, x_1^m)$, $\ldots$, $\overline{x}_k = (x_k^1, x_k^2, \cdots, x_k^m)$ be a set of generators for id-Pol$(\mathcal{G}^{++})$. Set $\overline{x}'_1, \ldots, \overline{x}'_k$ to be the tuples obtained from $\overline{x}_1, \ldots, \overline{x}_k$ by substituting t by t' and leaving everything else unchanged. We claim that $\overline{x}'_1, \ldots, \overline{x}'_k$ is a set of generators for id-Pol$(\mathcal{G}^{+})$. To prove this then, let $\underline{y} = (y^1, y^2, \ldots, y^m) \in (G^+)^m$ be given. We need to prove there is $f' \in$ id-Pol$(\mathcal{G}^{+})$ so that we have the following.

$$
\begin{array}{cccc}
f' & f' & \cdots & f' \\
\frown & \frown & \cdots & \frown \\
x_1^{1'} & x_1^{2'} & \cdots & x_1^{m'} \\
x_2^{1'} & x_2^{2'} & \cdots & x_2^{m'} \\
\vdots & \vdots & & \vdots \\
x_k^{1'} & x_k^{2'} & \cdots & x_k^{m'} \\
\smile & \smile & \cdots & \smile \\
\| & \| & & \| \\
y^1 & y^2 & \cdots & y^m
\end{array}
$$

Let $1, \ldots, n, n+1, n+2$ enumerate the elements of $\mathcal{G}^{++}$ with $\mathcal{G}^{+}$ being induced on the subset $\{1, \ldots, n, n+1\}$. For $i \in [n+2]$, let i^k denote the k-tuple of is.

By Lemma 13, it is sufficient to show that all pp-formulas that are true on $((\mathcal{G}^+)^k; 1^k, \ldots, (n+1)^k, \underline{x}^{1'}, \ldots, \underline{x}^{m'})$ are also true on $(\mathcal{G}^+; 1, \ldots, n+1, y^1, \ldots, y^m)$.

Let $\varphi = \exists\, \overline{w} \varphi(\overline{w}, \overline{v})$ be a pp-formula that is true on $((\mathcal{G}^+)^k; 1^k, \ldots, (n{+}1)^k, \underline{x}^{1'}, \ldots, \underline{x}^{m'})$, that is, for each $j \in [k]$, it is true on $(\mathcal{G}^+; 1, \ldots, n+1, x_j^{1'}, \ldots, x_j^{m'})$. Let $\overline{w}_j$ be the witnesses for the existential variables of φ on this latter structure. Since for all $x \in G^{++}$ we have $E(x, t')$ implies $E(x, t)$, we deduce that φ is also true on $(\mathcal{G}^{++}; 1, \ldots, n+1, x_j^1, \ldots, x_j^m)$, using the same witnesses $\overline{w}_0$. Now it follows from $f_{\underline{y}}$ that φ is true on $(\mathcal{G}^{++}; 1, \ldots, n+1, y^1, \ldots, y^m)$, by mapping the tuples $\overline{w}_0, \ldots, \overline{w}_k$ under $f_{\underline{y}}$ to obtain the witness for $\overline{w}$ in $(\mathcal{G}^{++}; 1, \ldots, n+1, y^1, \ldots, y^m)$. But, the idempotent $f_{\underline{y}}$ preserves the set $\{1, \ldots, n, n+1\}$, which is pp-definable in $\mathcal{G}^{++}$, so this shows that the same witnesses show φ is also true on $(\mathcal{G}^+; 1, \ldots, n+1, x_j^1, \ldots, x_j^m)$. The result follows. $\square$

Corollary 15. *Let $\mathcal{G}$ be a digraph. If* id-Pol$(\mathcal{G}^+)$ *has the EGP then so does* id-Pol$(\mathcal{G}^{++})$.

Let $\mathcal{G}$ be a semicomplete digraph with more than one cycle and no source. We say $\mathcal{G}$ has the *Novi Sad property* if there exist vertices $p, q \in G$ so that

- for all $v \in G$ there is the edge $E(v, p)$ or $E(v, q)$.

Note that the Novi Sad property implies a double edge between p and q, hence this fails on all tournaments. Importantly for our uses, on irreflexive graphs this property implies that (picking $p' := q$ and $q' := p$):

- exists $p' \in G$ so that $E(p', p)$ but not $E(p', q)$,
- exists $q' \in G$ so that $E(q', q)$ but not $E(q', p)$.

The Novi Sad property does not feature in [2].
Specific results imported from [2]. We now need to borrow some definitions and results from [2]. In that paper the authors usually refer to Pol instead of id-Pol, but the the objects are always cores expanded by constants, so the two coincide.

Definition 16 (Definition 6 in [2])**.** *Let $\mathcal{G}$ be a directed graph. We define the relation $\preceq_{\mathcal{G}}$ on V by $x \preceq_{\mathcal{G}} y$ iff $x^- \subseteq y^-$.*

Proposition 17 (Proposition 9 in [2])**.** *Assume that $\mathcal{G}$ is semicomplete. Then $\preceq_{\mathcal{G}}$ is a partial order, $\preceq_{\mathcal{G}}$ has the largest element t iff t is a sink, and dually for least elements and sources.*

Definition 18 (Definition 7 in [2])**.** *Let $\mathcal{G}$ be a digraph. We define the partition of the vertex set V into $V^{\mathcal{G}}_{min}$, $V^{\mathcal{G}}_{max}$, $V^{\mathcal{G}}_{both}$ and $V^{\mathcal{G}}_{none}$ so that all vertices in $V^{\mathcal{G}}_{max}$ are maximal, but not minimal, in the order $\preceq_{\mathcal{G}}$, all vertices in $V^{\mathcal{G}}_{min}$ are minimal, but not maximal, in the order $\preceq_{\mathcal{G}}$, all vertices in $V^{\mathcal{G}}_{both}$ are both minimal and maximal in the order $\preceq_{\mathcal{G}}$, while vertices in $V^{\mathcal{G}}_{none}$ are neither minimal nor maximal in the order $\preceq_{\mathcal{G}}$. When the digraph $\mathcal{G}$ is understood, we will omit the superscript $^{\mathcal{G}}$.*

Definition 19 (Definition 8 in [2])**.** *Let $\mathcal{G}$ be a digraph. We define the irreflexive digraph $\mathcal{S}(\mathcal{G})$ by:*

1. *For all $x, y \in V_{max} \cup V_{both}$, $(x, y), (y, x) \in E(\mathcal{S}(\mathcal{G}))$,*
2. *For all $x, y \in V_{min}$, $(x, y), (y, x) \in E(\mathcal{S}(\mathcal{G}))$,*
3. *For all $x, y \in V_{none}$, $(x, y) \in E(\mathcal{S}(\mathcal{G}))$ iff $(x, y) \in E(\mathcal{G})$.*
4. *For all $x \in V_{min}$ and $y \in V_{none} \cup V_{max}$, $(x, y) \in E(\mathcal{S}(\mathcal{G}))$, but not $(y, x) \in E(\mathcal{S}(\mathcal{G}))$,*
5. *For all $x \in V_{none}$ and $y \in V_{max}$, $(x, y) \in E(\mathcal{S}(\mathcal{G}))$, but not $(y, x) \in E(\mathcal{S}(\mathcal{G}))$,*
6. *For all $x \in V_{both}$ and $y \in V_{none} \cup V_{min}$, $(x, y) \in E(\mathcal{S}(\mathcal{G}))$, but not $(y, x) \in E(\mathcal{S}(\mathcal{G}))$.*

Proposition 20 (Proposition 10 in [2])**.** *$V^{\mathcal{S}(\mathcal{G})}_{min} = V^{\mathcal{G}}_{min}$, $V^{\mathcal{S}(\mathcal{G})}_{max} = V^{\mathcal{G}}_{max}$, $V^{\mathcal{S}(\mathcal{G})}_{both} = V^{\mathcal{G}}_{both}$ and $V^{\mathcal{S}(\mathcal{G})}_{none} = V^{\mathcal{G}}_{none}$. Consequently, $\mathcal{S}(\mathcal{S}(\mathcal{G})) = \mathcal{S}(\mathcal{G})$.*

Corollary 21 (Corollary 6 in [2]). *Let $\mathcal{G}$ be a smooth semicomplete digraph which is not a cycle. Then* id-Pol$(\mathcal{G}^+) \subseteq$ id-Pol$(\mathcal{S}(\mathcal{G})^+)$.

Applications of results imported from [2].

Theorem 22. *Let $\mathcal{G}$ be a smooth semicomplete with more than one cycle. There exists a smooth semicomplete with more than one cycle $\mathcal{H}$ so that* id-Pol$(\mathcal{G}^+) \subseteq$ id-Pol$(\mathcal{H}^+)$ *and $\mathcal{H}^+$ has the Novi Sad property.*

Proof. Note that $|V^{\mathcal{G}}_{both} \cup V^{\mathcal{G}}_{max}| \geq 2$, so we can apply Corollary 21, choosing $\mathcal{H} = \mathcal{S}(\mathcal{G})$, with $p \neq q$ chosen as follows: If $V^{\mathcal{G}}_{max} = \emptyset$, this implies that $V^{\mathcal{G}}_{min} = V^{\mathcal{G}}_{none} = \emptyset$ and $V = V^{\mathcal{G}}_{both}$, and p, q can be chosen arbitrarily; If $V^{\mathcal{G}}_{max} \neq \emptyset$, then we choose $p \in V^{\mathcal{G}}_{max}$ and $q \in V^{\mathcal{G}}_{max} \cup V^{\mathcal{G}}_{both}$. Then $(p, q), (q, p) \in E(\mathcal{S}(\mathcal{G}))$ (and this graph has no loops), and there is an edge from all vertices of $\mathcal{S}(\mathcal{G})$, except p, to p. $\square$

Main EGP result for semicompletes.

Proposition 23. *Let $\mathcal{G}$ be a semicomplete digraph with more than one cycle, no source, and the Novi Sad property. Then* id-Pol$(\mathcal{G})$ *has the EGP.*

Proof. Let p and q, together with p' and q', be as guaranteed to exist by the Novi Sad property. Let U be the unary relation specifying the domain of the smooth semicomplete digraph with more than one cycle which is obtained from $\mathcal{G}$ by removing sinks repeatedly.

A word $\tau \in G^m$ is said to be a *sub-predecessor* of a word $\sigma \in \{p, q\}^m$ if τ can be obtained by some local substitutions of $p \mapsto x \in p^-$ and $q \mapsto x \in q^-$. If τ is a sub-predecessor of σ then we may say σ is a *sub-successor* of τ. Note that every word $\tau \in G^m$ has a sub-successor in $\sigma \in \{p, q\}^m$, by the Novi Sad property. A word $\tau \in G^m$ is said to be a *predecessor* of a word $\sigma \in \{p, q\}^m$ if τ can be obtained by some local substitutions of $p \mapsto x \in p^- \setminus q^-$ and $q \mapsto x \in q^- \setminus p^-$. If τ is a predecessor of σ then we may say σ is a *successor* of τ. Note that predecessor (resp., successor) imply sub-predecessor (resp., sub-successor). Now,

$$(\dagger) \begin{cases} \text{. each word in } ((p^- \setminus q^-) \cup (q^- \setminus p^-))^m \text{ has a unique successor in } \{p, q\}^m\text{; and} \\ \text{. every word in } G^m \setminus ((p^- \setminus q^-) \cup (q^- \setminus p^-))^m \text{ has more than one sub-successor in } \{p, q\}^m. \end{cases}$$

In analogy to the proof of Proposition 10, predecessor/ successor play the role of cousin and sub-predecessor/ sub-successor play the role of friend.

Let m be given and suppose there exists a generating set Γ for G^m of size $< 2^m$. It follows from (†) that, for some $\tau \in \{p, q\}^m$, Γ omits τ and all of τ's predecessors.

We will prove that Γ does not generate G^m, by assuming otherwise and reaching a contradiction. Let R_Γ be the subset $\{p, q\}^m \setminus \{\tau\}$. Note

$$(*) \quad \text{that every } \sigma \in \Gamma \text{ has a sub-successor in } R_\Gamma.$$

Note also that R_Γ is pp-definable since U is pp-definable and all polymorphisms of the sub-structure induced by U are projections (see [2]).

Consider the pH-formula $\varphi(x_1, \ldots, x_n) :=$

$$\exists x'_1, \ldots, x'_n \left(\bigwedge_{i \in [n]} E(x_i, x'_i) \right) \wedge R_\Gamma(x'_1, \ldots, x'_n).$$

The sentence $\forall x_1, \ldots, x_n\ \varphi(x_1, \ldots, x_n)$ is false and can be witnessed as false by taking $\overline{x}$ to be that word in $\{p', q'\}^m$ derived from τ by substituting $p \mapsto p'$ and $q \mapsto q'$. However, consider now that $\varphi(y_1, \ldots, y_n)$ is true for all $(y_1, \ldots, y_n) \in \Gamma$, precisely because of property $(*)$, i.e. when $(x_1, \ldots, x_n)$ is evaluated as σ, choose $(x'_1, \ldots, x'_n)$ to be evaluated as σ's sub-successor in R_Γ. □

Corollary 24. *Let $\mathcal{G}$ be a semicomplete digraph with more than one cycle and either no source or no sink. Then* id-Pol$(\mathcal{G})$ *has the EGP.*

Proof. From [2] we know that semicomplete digraphs $\mathcal{H}$ with more than one cycle and neither a source nor a sink (smooth) have only essentially unary polymorphisms. It follows of course that id-Pol$(\mathcal{H})$ has the EGP. The result now follows from Proposition 23 (and its symmetric dual). □

4 The PGP: collapsibility and beyond

Throughout this section, we shall be concerned with a relational structure $\mathcal{A}$ over a finite set A of size n. In the few cases when we will require $\mathcal{A}$ to have specific constants, we shall state it explicitly.

4.1 Games, adversaries and reactive composition

We recall some terminology due to Chen [3, 10], for his natural adaptation of the model checking game to the context of pH-sentences. We shall not need to explicitly play these games but only to handle strategies for the existential player. An *adversary* $\mathscr{B}$ of length $m \geq 1$ is an m-ary relation over A. When $\mathscr{B}$ is precisely the set $B_1 \times B_2 \times \ldots \times B_m$ for some non-empty subsets $B_1, B_2, \ldots, B_m$ of A, we speak of a *rectangular adversary.* Let φ have universal variables $x_1, \ldots, x_m$ and quantifier-free part ψ. We write $\mathcal{A} \models \varphi_{\restriction \mathscr{B}}$ and say that *the existential player has a winning strategy in the* $(\mathcal{A}, \varphi)$*-game against adversary* $\mathscr{B}$ iff there exists a set of Skolem functions $\{\sigma_x : \text{‘}\exists x\text{’} \in \varphi\}$ such that for any assignment π of the universally quantified variables of φ to A, where $\big(\pi(x_1), \ldots, \pi(x_m)\big) \in \mathscr{B}$, the map h_π is a homomorphism from $\mathcal{D}_\psi$ (the canonical database) to $\mathcal{A}$, where

$$h_\pi(x) := \begin{cases} \pi(x) & \text{, if } x \text{ is a universal variable; and,} \\ \sigma_x(\pi|_{Y_x}) & \text{, otherwise.} \end{cases}$$

(Here, Y_x denotes the set of universal variables preceding x and $\pi|_{Y_x}$ the restriction of π to Y_x.) Clearly, $\mathcal{A} \models \varphi$ iff the existential player has a winning strategy in the $(\mathcal{A}, \varphi)$-game against the so-called *full (rectangular) adversary* $A \times A \times \ldots \times A$ (which we will denote hereafter by A^m). We say that an adversary $\mathscr{B}$ of length m *dominates* an adversary $\mathscr{B}'$ of length m when $\mathscr{B}' \subseteq \mathscr{B}$. Note that $\mathscr{B}' \subseteq \mathscr{B}$ and $\mathcal{A} \models \varphi_{\restriction \mathscr{B}}$ implies $\mathcal{A} \models \varphi_{\restriction \mathscr{B}'}$. We will also consider sets of adversaries of the same length, denoted by uppercase greek letters as in Ω_m; and, sequences thereof, which we denote with bold uppercase greek letters as in $\mathbf{\Omega} = \left(\Omega_m\right)_{m\in\mathbb{N}}$. We will write $\mathcal{A} \models \varphi_{\restriction \Omega_m}$ to denote that $\mathcal{A} \models \varphi_{\restriction \mathscr{B}}$ holds for every adversary $\mathscr{B}$ in Ω_m. We call *width* of Ω_m and write $\text{width}(\Omega_m)$ for $\sum_{\mathscr{B}\in\Omega_m} |\mathscr{B}|$. We say that $\mathbf{\Omega}$ is *polynomially bounded* if there exists a polynomial $p(m)$ such that for every $m \geq 1$, $\text{width}(\Omega_m) \leq p(m)$. We say that $\mathbf{\Omega}$ is *effective* if there exists a polynomial $p'(m)$ and an algorithm that outputs Ω_m for every m in total time $p'(\text{width}(\Omega_m))$.

Let f be a k-ary operation of $\mathcal{A}$ and $\mathscr{A}, \mathscr{B}_1, \ldots, \mathscr{B}_k$ be adversaries of length m. We say that $\mathscr{A}$ is *reactively composable* from the adversaries $\mathscr{B}_1, \ldots, \mathscr{B}_k$ via f, and we write $\mathscr{A} \trianglelefteq f(\mathscr{B}_1, \ldots, \mathscr{B}_k)$ iff there exist partial functions $g_i^j : A^i \to A$ for every i in $[m]$ and every j in $[k]$ such that, for every tuple $(a_1, \ldots, a_m)$ in adversary $\mathscr{A}$ the following holds.

- for every j in $[k]$, the values $g_1^j(a_1), g_2^j(a_1, a_2), \ldots, g_m^j(a_1, a_2, \ldots, a_m)$ are defined and the tuple $\left(g_1^j(a_1), g_2^j(a_1, a_2), \ldots, g_m^j(a_1, a_2, \ldots, a_m)\right)$ is in adversary $\mathscr{B}_j$; and,
- for every i in $[m]$, $a_i = f\left(g_i^1(a_1, a_2, \ldots, a_i), g_i^2(a_1, a_2, \ldots, a_i), \ldots, g_i^k(a_1, a_2, \ldots, a_i)\right)$.

We write $\mathscr{A} \trianglelefteq \{\mathscr{B}_1, \ldots, \mathscr{B}_k\}$ if there exists a k-ary operation f such that $\mathscr{A} \trianglelefteq f(\mathscr{B}_1, \ldots, \mathscr{B}_k)$

Remark 25. *We will never show reactive composition by exhibiting a polymorphism f and partial functions g_j^i that depend on all their arguments. We will always be able to exhibit partial functions that depend only on their last argument.*

Reactive composition allows to interpolate complete Skolem functions from partial ones.

Theorem 26 ([10, Theorem 7.6])**.** *Let φ be a pH-sentence with m universal variables. Let $\mathscr{A}$ be an adversary and Ω_m a set of adversaries, both of length m.*

If $\mathcal{A} \models \varphi_{\restriction \Omega_m}$ and $\mathscr{A} \trianglelefteq \Omega_m$ then $\mathcal{A} \models \varphi$.

As a concrete example of an interesting sequence of adversaries, consider the adversaries for the notion of *p-collapsibility*, which we introduced in a purely logical fashion in the introduction. Let $p \geq 0$ be some fixed integer. For x in A, let $\Upsilon_{m,p,x}$ be the set of all rectangular adversaries of length m with p coordinates that are the set A and all the other that are the fixed singleton $\{x\}$. For $B \subseteq A$, let $\Upsilon_{m,p,B}$ be the union of $\Upsilon_{m,p,x}$ for all x in B. Let $\mathbf{\Upsilon}_{p,B}$ be the sequence of adversaries $\left(\Upsilon_{m,p,B}\right)_{m\in\mathbb{N}}$. We will define a structure $\mathcal{A}$ to be *p-collapsible from source B* iff for every m and for all pH-sentence φ with m universal variable, $\mathcal{A} \models \varphi_{\restriction \Upsilon_{m,p,B}}$ implies $\mathscr{A} \models \varphi$.

4.2 The Π_2-case

For a Π_2-pH sentence, the existential player knows the values of all universal variables beforehand, and it suffices for her to have a winning strategy for each instantiation (and perhaps no way to reconcile them as should be the case for an arbitrary sentence). This also means that considering a set of adversaries of same length is not really relevant in this Π_2-case as we may as well consider the union of these adversaries or the set of all their tuples (see also statement of Corollary 9).

Lemma 27 (**principle of union**). *Let Ω_m be a set of adversaries of length m and φ a Π_2-sentence with m universal variables. Let $\mathscr{O}_{\cup\Omega_m} := \bigcup_{\mathscr{O}\in\Omega_m} \mathscr{O}$ and $\Omega_{tuples} := \{\{t\}|t \in \mathscr{O}_{\cup\Omega_m}\}$. We have the following equivalence.*

$$\mathcal{A} \models \varphi_{\restriction\Omega_m} \quad\Longleftrightarrow\quad \mathcal{A} \models \varphi_{\restriction\mathscr{O}_{\cup\Omega_m}} \quad\Longleftrightarrow\quad \mathcal{A} \models \varphi_{\restriction\Omega_{tuples}}$$

Let $\mathscr{A}$ be an adversary and Ω_m a set of adversaries, both of length m. We say that Ω_m *generates* $\mathscr{A}$ iff for any tuple t in $\mathscr{A}$, there exists a k-ary polymorphism f_t of $\mathcal{A}$ and tuples $t_1, \ldots, t_k$ in Ω_{tuples} such that $f_t(t_1, \ldots, t_k) = t$. We have the following analogue of Theorem 26.

Proposition 28. *Let φ be a Π_2-pH-sentence with m universal variables. Let $\mathscr{A}$ be an adversary and Ω_m a set of adversaries, both of length m.*

If $\mathcal{A} \models \varphi_{\restriction\Omega_m}$ and Ω_m generates $\mathscr{A}$ then $\mathcal{A} \models \varphi_{\restriction\mathscr{A}}$.

We will construct a *canonical* Π_2*-sentence to assert that an adversary is generating.* Let $\mathscr{O}$ be some adversary of length m. Let $\sigma^{(m)}$ be the signature σ expanded with a sequence of m constants. For a map μ from $[m]$ to A, we write $\mu \in \mathscr{O}$ as shorthand for $(\mu(1), \mu(2), \ldots, \mu(m)) \in \mathscr{O}$. For some set Ω_m of adversaries of length m, we consider the following $\sigma^{(m)}$-structure:

$$\bigotimes_{\mathscr{O}\in\Omega_m} \bigotimes_{\mu\in\mathscr{O}} \mathfrak{A}_\mu$$

where the $\sigma^{(m)}$-structure $\mathfrak{A}_\mu$ denotes the expansion of $\mathcal{A}$ by m constants as given by the map μ. Let $\varphi_{\Omega_m,\mathcal{A}}$ be the Π_2-pH-sentence[4] created from the canonical query of the σ-reduct of this $\sigma^{(m)}$-structure with the m constants c_j becoming variables w_j, universally quantified outermost, when all *constants are pairwise distinct.* Otherwise, we will say that Ω_m is *degenerate*, and not define the canonical sentence.

Note that adversaries such as $\Upsilon_{m,p,B}$ corresponding to p-collapsibility are not degenerate for $p > 0$, and degenerate for $p = 0$.

Proposition 29. *Let Ω_m be a set of adversaries of length m that is not degenerate. The following are equivalent.*

[4]For two structures $\mathcal{A}$ and $\mathcal{B}$, when Ω_m is A^m and m is $|A|^B$, $\mathcal{B}$ models this canonical sentence iff $\text{QCSP}(\mathcal{A}) \subseteq \text{QCSP}(\mathcal{B})$ [12]

(i) *for any* Π_2*-pH sentence* ψ, $\mathcal{A} \models \psi_{\restriction \Omega_m}$ *implies* $\mathcal{A} \models \psi$.

(ii) *for any* Π_2*-pH sentence* ψ, $\mathcal{A} \models \psi_{\restriction \mathscr{O}_{\cup\Omega}}$ *implies* $\mathcal{A} \models \psi$.

(iii) *for any* Π_2*-pH sentence* ψ, $\mathcal{A} \models \psi_{\restriction \Omega_{tuples}}$ *implies* $\mathcal{A} \models \psi$.

(iv) $\mathcal{A} \models \varphi_{\mathscr{O}_{\cup\Omega},\mathcal{A}}$

(v) $\mathcal{A} \models \varphi_{\Omega_{tuples},\mathcal{A}}$

(vi) Ω_m *generates* A^m.

4.3 The unbounded case

Let n denote the number of elements of the structure $\mathcal{A}$. Let $\mathscr{B}$ be an adversary from $\Omega_{n\cdot m}$. We will denote by $\mathrm{Proj}\mathscr{B}$ the set of adversaries of length m induced by projecting over some arbitrary choice of m coordinates, one in each block of size n; that is $1 \leq i_1 \leq n, n+1 \leq i_2 \leq 2\cdot n, \ldots, n\cdot(m-1)+1 \leq i_m \leq n\cdot m$. Of special concern to us are *projective sequences of adversaries* $\boldsymbol{\Omega}$ satisfying the following for every $m \geq 1$,

$$\forall \mathscr{B} \in \Omega_{n.m}\ \exists \mathscr{A} \in \Omega_m \bigwedge_{\widetilde{\mathscr{B}} \in \mathrm{Proj}\mathscr{B}} \widetilde{\mathscr{B}} \subseteq \mathscr{A} \quad (m\text{-}\textbf{projectivity})$$

As an example, consider the adversaries for collapsibility.

Fact 30. *Let* $B \subseteq A$ *and* $p \geq 0$. *The sequence of adversaries* $\boldsymbol{\Upsilon}_{p,B}$ *are projective.*

Example 31. *For a concrete illustration consider* $A = \{0,1,2\}$ *(thus* $n = 3$*). We illustrate the fact that* $\boldsymbol{\Upsilon}_{p=2,B=\{0\}}$ *is projective for* $m = 4$ *and some adversary* $\mathscr{B} \in \Omega_{n\cdot m} = \Upsilon_{p=2,B=\{0\},3\cdot 4=12}$. *Adversaries are depicted vertically with horizontal lines separating the blocks.*

$\mathscr{B} \in \Omega_{n\cdot m}$	*Proj*$\mathscr{B}$				$\mathscr{A} \in \Omega_m$
A	A	A		$\cancel{A}$	
0	$\cancel{0}$	$\cancel{0}$	$\ldots$	$\cancel{0}$	A
0	$\cancel{0}$	$\cancel{0}$		0	
0	0	0		$\cancel{0}$	
0	$\cancel{0}$	$\cancel{0}$	$\ldots$	$\cancel{0}$	0
0	$\cancel{0}$	$\cancel{0}$		0	
0	0	0		$\cancel{0}$	
0	$\cancel{0}$	$\cancel{0}$	$\ldots$	$\cancel{0}$	0
0	$\cancel{0}$	$\cancel{0}$		0	
0	0	$\cancel{0}$		$\cancel{0}$	
A	$\cancel{A}$	A	$\ldots$	$\cancel{A}$	A
0	$\cancel{0}$	$\cancel{0}$		0	

The adversary $\mathscr{A}$ *dominates any adversary obtained by projecting the original larger adversary* $\mathscr{B}$ *by keeping a single position per block.*

We could actually consider w.l.o.g. sequences of *singleton* adversaries.

Fact 32. *If* $\boldsymbol{\Omega}$ *is projective then so is the sequence* $\left(\bigcup_{\mathscr{O}\in\Omega_m} \mathscr{O}\right)_{m\in\mathbb{N}}$.

A *canonical sentence for composability for arbitrary pH-sentences* with m universal variables may be constructed similarly to the canonical sentence for the Π_2 case, except that it will have $m.n$ universal variables, which we view as m blocks of n variables, where n is the number of elements of the structure $\mathcal{A}$. Let $\mathscr{O}$ be some adversary of length m. Let $\sigma^{(n\cdot m)}$ be the signature σ expanded with a sequence of $n.m$ constants $c_{1,1},\ldots,c_{n,1},c_{1,2}\ldots,c_{n,2},\ldots c_{1,m}\ldots,c_{n,m}$. We say that a map μ from $[n]\times[m]$ to A is *consistent* with $\mathscr{O}$ iff for every $(i_1,i_2,\ldots,i_m)$ in $[n]^m$, the tuple $(\mu(i_1,1),\mu(i_2,2),\ldots,\mu(i_m,m))$ belongs to the adversary $\mathscr{O}$. We write $A^{[n.m]}_{\restriction\mathscr{O}}$ for the set of such consistent maps. For some set Ω_m of adversaries of length m, we consider the following $\sigma^{(n.m)}$-structure:

$$\bigotimes_{\mathscr{O}\in\Omega_m}\ \bigotimes_{\mu\in A^{[n.m]}_{\restriction\mathscr{O}}} \mathfrak{A}_{\mathscr{O},\mu}$$

where the $\sigma^{(n\cdot m)}$-structure $\mathfrak{A}_{\mathscr{O},\mu}$ denotes the expansion of $\mathcal{A}$ by $n.m$ constants as given by the map μ. Let $\varphi_{n,\Omega_m,\mathcal{A}}$ be the Π_2-pH-sentence created from the canonical query of the σ-reduct of this $\sigma^{(n.m)}$ product structure with the $n.m$ constants c_{ij} becoming variables w_{ij}, universally quantified outermost. As for the canonical sentence of the Π_2-case, this sentence is not well defined if constants are not pairwise distinct, which occurs precisely for degenerate adversaries.

Lemma 33. *Let Ω_m be a set of adversaries of length m that is not degenerate. Let $\mathcal{A}$ be a structure of size n. If $\mathcal{A}$ models $\varphi_{n,\Omega_m,\mathcal{A}}$ then the full adversary A^m is reactively composable from Ω_m. That is, $\mathcal{A}\models\varphi_{n,\Omega_m,\mathcal{A}} \implies A^m \trianglelefteq \Omega_m$*

Proof. We let each block of n universal variables of the canonical sentence $\varphi_{n,\Omega_m,\mathcal{A}}$ enumerate the elements of A. That is, given an enumeration $a_1,a_2,\ldots,a_n$ of A, we set $w_{i,j}=a_i$ for every j in $[m]$ and every i in $[n]$.

The assignment to the existential variables provides us with a k-ary polymorphism (the sentence being built as the conjunctive query of a product of k copies of $\mathcal{A}$) together with the desired partial maps. A coordinate r in $[k]$ corresponds to a choice of some adversary $\mathscr{O}$ of Ω_m and some map μ_r from $[n]\times[m]$ to A, consistent with this adversary. The partial map $g^r_\ell : A^\ell \to A$ with ℓ in $[m]$ (and r in $[k]$) is given by μ_r as follows: $g^r_\ell(a_{i_1},\ldots,a_{i_\ell})$ depends only on the last coordinate a_{i_ℓ} and takes value $\mu(i,\ell)$ if $a_{i_\ell}=a_i$. By construction of the sentence and the property of consistency of such μ_r with the adversary $\mathscr{O}$, these partial functions satisfy the properties as given in the definition of reactive composition. □

Lemma 34. *Let $\mathbf{\Omega}$ be a sequence of sets of adversaries that has the m-projectivity property for some $m\geq 1$ such that $\Omega_{n\cdot m}$ is not degenerate. The following holds.*

(i) $\mathcal{A}\models\psi_{\restriction\mathbf{\Omega_{n.m}}}$, *where* $\psi=\varphi_{n,\mathbf{\Omega_m},\mathcal{A}}$

(ii) *If for every Π_2-sentence ψ with $m.n$ universal variables, it holds that $\mathcal{A}\models\psi_{\restriction\mathbf{\Omega_{m.n}}}$ implies $\mathcal{A}\models\psi$, then $\mathcal{A}\models\varphi_{n,\mathbf{\Omega_m},\mathcal{A}}$.*

Theorem 35. *Let* $\mathbf{\Omega}$ *be a sequence of sets of adversaries that has the* m*-projectivity property for some* $m \geq 1$ *such that* $\Omega_{n.m}$ *is not degenerate. The following chain of implications holds*

$$\text{(i)} \implies \text{(ii)} \implies \text{(iii)} \implies \text{(iv)}$$

where,

(i) *For every* Π_2*-pH-sentence* ψ *with* $m.n$ *universal variables,* $\mathcal{A} \models \psi_{\restriction \Omega_{m.n}}$ *implies* $\mathcal{A} \models \psi$.

(ii) $\mathcal{A} \models \varphi_{n,\Omega_m,\mathcal{A}}$.

(iii) $A^m \trianglelefteq \Omega_m$.

(iv) *For every pH-sentence* ψ *with* m *universal variables,* $\mathcal{A} \models \psi_{\restriction \Omega_m}$ *implies* $\mathcal{A} \models \psi$.

Proof. The first implication holds by the previous lemma (second item of Lemma 34, this is the step where we use projectivity). The second implication is Lemma 33. The last implication is Theorem 26. □

Thus, in the projective case, when an adversary is good enough in the Π_2-case, it is good enough in general. This can be characterised logically via canonical sentences or "algebraically" in terms of reactive composition or the weaker and more usual composition property (see (vi) below).

Theorem 36 (In abstracto). *Let* $\mathbf{\Omega} = \big(\Omega_m\big)_{m\in\mathbb{N}}$ *be a projective sequence of adversaries, none of which are degenerate. The following are equivalent.*

(i) *For every* $m \geq 1$*, for every pH-sentence* ψ *with* m *universal variables,* $\mathcal{A} \models \psi_{\restriction \Omega_m}$ *implies* $\mathcal{A} \models \psi$.

(ii) *For every* $m \geq 1$*, for every* Π_2*-pH-sentence* ψ *with* m *universal variables,* $\mathcal{A} \models \psi_{\restriction \Omega_m}$ *implies* $\mathcal{A} \models \psi$.

(iii) *For every* $m \geq 1$*,* $\mathcal{A} \models \varphi_{n,\Omega_m,\mathcal{A}}$.

(iv) *For every* $m \geq 1$*,* $\mathcal{A} \models \varphi_{\mathscr{O}_{\cup\Omega},\mathcal{A}}$.

(v) *For every* $m \geq 1$*,* $A^m \trianglelefteq \Omega_m$.

(vi) *For every* $m \geq 1$*,* Ω_m *generates* A^m.

Remark 37. *The above equivalences can be read along two dimensions:*

	general	Π_2
logical interpolation	(i)	(ii)
canonical sentences	(iii)	(iv)
algebraic interpolation	(v)	(vi)

In [10], Chen introduces effective PGP and shows that it entails a QCSP to CSP reduction, for the bounded alternation QCSP. For concrete examples, such as collapsibility and switchability, he shows a QCSP to CSP reduction even in the unbounded case [10, Theorem 7.11]. As a second corollary, we can generalise this last result to effective and "projective" PGP, though we formulate this in terms of sequence of adversaries.

Corollary 38. *Let $\mathcal{A}$ be a structure. Let $\mathbf{\Omega}$ be a sequence of non degenerate adversaries that is effective, projective and polynomially bounded such that Ω_m generates A^m for every $m \geq 1$.*

Let $\mathcal{A}'$ be the structure $\mathcal{A}$, possibly expanded with constants, at least one for each element that occurs in $\mathbf{\Omega}$. The problem QCSP($\mathcal{A}$) reduces in polynomial time to CSP($\mathcal{A}'$). In particular, if $\mathcal{A}$ has all constants, the problem QCSP$_c$($\mathcal{A}$) reduces in polynomial time to CSP$_c$($\mathcal{A}$).

4.4 Studies of Collapsibility

Let $\mathcal{A}$ be a structure, $B \subseteq A$ and $p \geq 0$. Recall the structure $\mathcal{A}$ is *p-collapsible with source B* when for all $m \geq 1$, for all pH-sentences φ with m universal quantifiers, $\mathcal{A} \models \varphi$ iff $\mathcal{A} \models \varphi_{|\Upsilon_{m,p,B}}$. Collapsible structures are very important: to the best of our knowledge, they are in fact the only examples of structures that enjoy a form of polynomial QCSP to CSP reduction. This is different if one considers structures with infinitely many relations where the more general notion of *switchability* crops up [10]. Our abstract results of the previous section apply to both switchability and collapsibility but we concentrate here on the latter. This result applies since the underlying sequence of adversaries are projective (see Fact 30), as long as $p > 0$ (non degenerate case).

Corollary 39 (In concreto). *Let $\mathcal{A}$ be a structure, $\emptyset \subsetneq B \subseteq A$ and $p > 0$. The following are equivalent.*

(i) *$\mathcal{A}$ is p-collapsible from source B.*

(ii) *$\mathcal{A}$ is Π_2-p-collapsible from source B.*

(iii) *For every m, the structure $\mathcal{A}$ satisfies the canonical Π_2-sentence with $m \cdot |A|$ universal variables $\varphi_{n,\Upsilon_{m,p,B},\mathcal{A}}$.*

(iv) *For every m, the structure $\mathcal{A}$ satisfies the canonical Π_2-sentence with m universal variables $\varphi_{\mathscr{U},\mathcal{A}}$, where $\mathscr{U} = \bigcup_{\mathcal{O} \in \Upsilon_{m,p,B}} \mathcal{O}$.*

(v) *For every m, there exists a polymorphism f of $\mathcal{A}$ witnessing that $A^m \trianglelefteq \Upsilon_{m,p,B}$.*

(vi) *For every m, for every tuple t in A^m, there is a polymorphism f_t of $\mathcal{A}$ of arity k at most $\binom{m}{p}.|B|$ and tuples $t_1, t_2, \ldots, t_k$ in $\Upsilon_{m,p,B}$ such that $f_t(t_1, t_2, \ldots, t_k) = t$.*

Remark 40. *When $p = 0$, we obtain degenerate adversaries and this is due to the fact that if a QCSP is permitted equalities, then 0-collapsibility can never manifest (think of $\forall x, y\ x = y$).*

In [3], Case (v) of Corollary 39 is equivalent to id-Pol($\mathcal{A}$) being p-collapsible (in the algebraic sense). It is proved in [3] that if id-Pol($\mathcal{A}$), is k-collapsible (in the algebraic sense), then $\mathcal{A}$ is k-collapsible. We note that Corollary 39 proves the converse, finally tying together the two forms of collapsibility.

A fun application of Corollary 39 is an alternative proof of Proposition 12. It is easy to see that a semicomplete digraph with both a source and a sink is 1-collapsible with any singleton source. This is because any input sentence for QCSP($\mathcal{G}$), involving a universal variable v in an edge relation E, is false (evaluate as either the source or the sink, depending on whether v appears as the second or first entry of E, respectively). The statement of the proposition now follows from Corollary 39, via (i) $\Rightarrow$ (vi).

Another application of Corollary 39 is the following (compare with § 3.1.1).

Application 41. *A partially reflexive path $\mathcal{A}$ (no constants are present) that is quasi-loop connected has the PGP.*

The last two conditions of Corollary 39 provide us with a semi-decidability result: for each m, we may look for a particular polymorphism (v) or several polymorphisms (vi). Instead of a sequence of polymorphisms, we now strive for a better algebraic characterisation. We will only be able to do so for the special case of a singleton source, but this is the only case hitherto found in nature.

Chen uses the following lemma to show 4-collapsibility of bipartite graphs and disconnected graphs [8, Examples 1 and 2]. Though, we know via a direct argument [14] that these examples are in fact 1-collapsible from a singleton source.

Lemma 42 (Chen's lemma [3, Lemma 5.13])**.** *Let $\mathcal{A}$ be a structure with a constant x. If there is a k-ary polymorphism of $\mathcal{A}$ such that f is surjective when restricted at any position to $\{x\}$, then $\mathcal{A}$ is $(k-1)$-collapsible from source $\{x\}$ (i.e. $\mathcal{A}$ has a k-ary Hubie polymorphism).*

An interesting consequence of last section's formal work is a form of converse of Chen's Lemma, which allows us to give an algebraic characterisation of collapsibility from a singleton source.

Proposition 43. *Let x be a constant in $\mathcal{A}$. The following are equivalent:*

(i) *$\mathcal{A}$ is collapsible from $\{x\}$.*

(ii) *$\mathcal{A}$ has a Hubie polymorphism with source x.*

In the proof of the above, for *(i)* $\Rightarrow$ *(ii)* $\Rightarrow$ *(i)*, we no longer control the collapsibility parameter as the arity of our polymorphism is larger than the parameter we start with. By inspecting more carefully the properties of the polymorphism f we get as a witness that $\mathcal{A}$ models a canonical sentence, we may derive in fact p-collapsibility by an argument akin to the one used above in the proof of Chen's Lemma. We obtain this way a nice concrete result to counterbalance the abstract Theorem 36.

Theorem 44 (**p-Collapsibility from a singleton source**)**.** *Let x be a constant in $\mathcal{A}$ and $p > 0$. The following are equivalent:*

(i) *$\mathcal{A}$ is p-collapsible from $\{x\}$.*

(ii) *For every $m \geq 1$, the full adversary A^m is reactively composable from $\Upsilon_{m,p,x}$.*

(iii) *$\mathcal{A}$ is Π_2-p-collapsible from $\{x\}$.*

(iv) *For every* $m \geq 1$, $\Upsilon_{m,p,x}$ *generates* A^m.

(v) $\mathcal{A}$ *models* $\varphi_{n,\Upsilon_{p+1,p,x},\mathcal{A}}$ *(which implies that* $\mathcal{A}$ *admits a particularly well behaved Hubie polymorphism with source* x *of arity* $(p+1)n^p$*).*

Corollary 45. *Given* $p \geq 1$*, a structure* $\mathcal{A}$ *and* x *a constant in* $\mathcal{A}$*, we may decide whether* $\mathcal{A}$ *is* p*-collapsible from* $\{x\}$*.*

Remark 46. *We say that a structure* $\mathcal{A}$ *is* B-conservative *where* B *is a subset of its domain iff for any polymorphism* f *of* $\mathcal{A}$ *and any* $C \subseteq B$*, we have* $f(C, C, \ldots, C) \subseteq C$*. Provided that the structure is conservative on the source set* B*, we may prove a similar result for* p*-Collapsibility from a conservative source.*

Expanding on Remark 40, we note that if we forbid equalities in the input to a QCSP, then we can observe the natural case of 0-collapsibility, to which now we turn. This is not a significant restriction in a context of complexity, since in all but trivial cases of a one element domain, one can propagate equality out through renaming of variables.

We investigated a similar notion in the context of positive equality free first-order logic, the syntactic restriction of first-order logic that consists of sentences using only $\exists, \forall, \wedge$ and $\vee$. For this logic, relativisation of quantifiers fully explains the complexity classification of the model checking problem (a tetrachotomy between Pspace-complete, NP-complete, Co-NP-complete and Logspace) [15]. In particular, a complexity in NP is characterised algebraically by the preservation of the structure by a *simple A-shop* (to be defined shortly), which is equivalent to a strong form of 0-collapsibility since it applies not only to pH-sentences but also to sentences of positive equality free first-order logic. We will show that this notion corresponds in fact to 0-collapsibility from a singleton source. Let us recall first some definitions.

A *shop* on a set B, short for surjective hyper-operation, is a function f from B to its powerset such that $f(x) \neq \emptyset$ for any x in B and for every y in B, there exists x in B such that $f(x) \ni y$. An *A-shop*[5] satisfies further that there is some x such that $f(x) = B$. A *simple A-shop* satisfies further that $|f(x')| = 1$ for every $x' \neq x$. We say that a shop f is a *she* of the structure $\mathcal{B}$, short for *surjective hyper-endomorphism*, iff for any relational symbol R in σ of arity r, for any elements $a_1, a_2 \ldots, a_r$ in B, if $R(a_1, \ldots, a_r)$ holds in $\mathcal{B}$ then $R(b_1, \ldots, b_r)$ holds in $\mathcal{B}$ for any $b_1 \in f(a_1), \ldots, b_r \in f(a_r)$. We say that $\mathcal{B}$ *admits a (simple) A-she* if there is a (simple) A-shop f that is a she of $\mathcal{B}$.

Theorem 47. *Let* $\mathcal{B}$ *be a finite structure. The following are equivalent.*

(i) $\mathcal{B}$ *is* 0*-collapsible from source* $\{x\}$ *for some* x *in* B *for equality-free pH-sentences.*

(ii) $\mathcal{B}$ *admits a simple A-she.*

(iii) $\mathcal{B}$ *is* 0*-collapsible from source* $\{x\}$ *for some* x *in* B *for sentences of positive equality free first-order logic.*

[5]The A does not stand for the name of the set, it is short for *All*.

The above applies to singleton source only, but up to taking a power of a structure (which satisfies the same QCSP), we may always place ourselves in this singleton setting for 0-collapsibility.

Theorem 48. *Let $\mathcal{B}$ be a structure. The following are equivalent.*

(i) *$\mathcal{B}$ is 0-collapsible from source C*

(ii) *$\mathcal{B}^{|C|}$ is 0-collapsible from some (any) singleton source x which is a (rainbow) $|C|$-tuple containing all elements of C.*

5 Back to Complexity

The trichotomy of Theorem 3 should be seen as a companion to the following *dichotomy* result.

Theorem 49 (Theorem 1 of [1])**.** *Let $\mathcal{H}$ be a p.r. path.*

(i) *If $\mathcal{H}$ is quasi-loop-connected, then QCSP($\mathcal{H}$) is in NL.*

(ii) *Otherwise, QCSP($\mathcal{H}$) is Pspace-complete.*

Case (i) is proved in 2 steps : a loop connected p.r. path is known to be in NL via a majority polymorphism and a quasi-loop connected p.r. path is shown to have the same QCSP via some surjective homomorphisms from powers (via the methodology from [12]). This means that we can build a Hubie polymorphism for a quasi-loop connected p.r. path (see Application 41). However, this polymorphism need not be idempotent and the argument does not extend to p.r. paths with constants.

Using results from both of the previous sections we can now give a proof of Theorem 3.

Proof of Theorem 3. For Cases (i) and (ii), NP membership follows from Corollary 38 as we established suitable forms of PGP in Lemmas 4, 6 and 7. More specifically, the Ptime membership of Case (i) is established by the majority polymorphism mentioned in the proof of Lemma 4 (via [3]). As for Case (ii), we note in passing that collapsibility follows from Lemmas 6 and 7 which establish item (vi) of Corollary 39. More importantly, NP-hardness follows from the classification of [16].

For Case (iii), we observe from [1] that we are Pspace-hard even without constants. □

We note that the complexity classification for semicomplete digraphs from [2] is unchanged regardless of whether all constants are present (since semicompletes are cores).

6 Conclusion

One important application of our abstract investigation of PGP yields a nice characterisation in the concrete case of collapsibility, in particular in the case of a singleton source which we now know can be equated with preservation under a single polymorphism, namely a Hubie polymorphism. So far, this is the only known explanation for a complexity of a QCSP in NP which provokes the following question.

Question 1. *For a structure $\mathcal{A}$, is it the case that QCSP($\mathcal{A}$) is in NP iff $\mathcal{A}$ admits a Hubie polymorphism?*

In the literature, it is common to study the case of non finite constraint languages. This means that for an infinite set of relations over the same finite domain Γ we study the uniform problem QCSP(Γ) which covers all problems QCSP($\mathcal{A}$) where $\mathcal{A}$ is a structure with relations from Γ.

Typically Γ is taken to be the invariant of some algebra. There is an example of such a problem QCSP(Γ) with a complexity in NP that is provably not collapsible but enjoys a property similar to p-collapsibility, namely *p-switchability* [10], which is a special form of PGP.

For $m \geq 1$ and $\bar{\imath} = (i_1, i_2, \ldots, i_p)$ a strictly increasing sequence in $[m-1]^p$, let $\mathscr{S}_{\bar{\imath},p}$ be the adversary that consists of tuples $t \in A^m$ such that each of the following sets contain a single element: $\{t[j] \in A | 1 \leq j \leq i_1\}$, $\{t[j] \in A | i_1 + 1 \leq j \leq i_2\}$, ..., $\{t[j] \in A | i_p + 1 \leq j \leq m\}$. Let $\Sigma_{m,p}$ be the set of all such adversaries $\mathscr{S}_{\bar{\imath},p}$. Let $\boldsymbol{\Sigma}_p$ be the sequence of adversaries $\left(\Sigma_{m,p}\right)_{m \in \mathbb{N}}$.

We say that a structure A is *p-switchable* iff for every m and for all pH-sentence φ with m universal variable, $\mathcal{A} \models \varphi_{\restriction \Sigma_{m,p}}$ implies $\mathscr{A} \models \varphi$.

We say that a set of relations Γ is *p-switchable* iff every structure $\mathcal{A}$ with relations from Γ is p-switchable.

Our definition of switchability is not exactly the same as that of Hubie Chen who uses instead a single adversary $\cup \mathscr{S}_{\bar{\imath}}$ for each arity. It is a simple exercise to show that both sequences of adversaries satisfy the hypotheses of Theorem 36. Since the two notions are of course equivalent in the Π_2 case via the principle of union (Lemma 27), they are therefore equivalent in general. Thus not only we can equate switchability with its Π_2 analogue but we can also give a purely syntactic definition of switchability as follows. A structure $\mathcal{A}$ is p-switchable iff, for all m and for all pH formula φ with m universal variables $x_1, x_2, \ldots, x_m$ (in this order), $\mathcal{A} \models \varphi$ iff for all $\bar{\imath} = (i_1, i_2, \ldots, i_p)$ a strictly increasing sequence in $[m-1]^p$, $\mathcal{A} \models \varphi \wedge \eta_{\bar{\imath}}$ where $\eta_{\bar{\imath}}$ is $\bigwedge_{0 \leq \ell_1 < \ell_2 \leq p} \bigwedge_{i_{\ell_1} \leq j < k \leq i_{\ell_2}} x_j = x_k$.

However, there are two limitations to our result on switchability. Firstly, we do not have a crisp candidate for a single polymorphism or even a sequence of polymorphisms that would endow switchability. Secondly, our findings only hold for finite structures, where it is unclear that switchability plays a natural role. This provokes the following question.

Question 2. *For every infinite set of relations Γ, is it the case that Γ is switchable iff it is Π_2-switchable?*

Going back to collapsibility, regarding the meta-question of deciding whether a structure is collapsible, one can wonder if the parameter p of collapsibility depends on the size of the structure $\mathcal{A}$. In particular, this would provide a positive answer to the following.

Question 3. *Given a structure $\mathcal{A}$, can we decide if it is p-collapsible for some p?*

A tantalising question remains.

Question 4. *Are there any finite algebras, minimal generating sets for whose powers grow sub-exponentially (e.g. $\Theta(2^{\sqrt{i}})$)?*

The alternative is that finite algebras exhibit a PGP-EGP gap in general. In a sequence of three papers *Growth rates of algebras*, Kearnes, Kiss and Szenderei explore this question, demonstrating all polynomial growth rates are possible.

Finally, let us return to the *foundation for Fürstenproblem* and contemplate the complexity of the QCSP. Let $\mathcal{B}$ be a finite structure. At present it is not conjectured where one might seek to prove the boundary between QCSP($\mathcal{B}$) being in P and QCSP($\mathcal{B}$) being NP-hard, even in the case where all constants are present. Furthermore, settling this will be at least as hard as settling the similar dichotomy for CSP. However, we would like to specifically echo the conjecture of Chen in [8] (where it appears written in two conjectures).

Conjecture. Let $\mathcal{B}$ be finite and expanded with all constants; then $\mathrm{QCSP}_c(\mathcal{B})$ is in NP iff id-Pol($\mathcal{B}$) has the PGP.

Acknowledgment

The authors would like to thank the four anonymous reviewers for their patience, stamina and very useful suggestions, which have been a great help to prepare the final version of this paper.

References

[1] B. Martin, "QCSP on partially reflexive forests," in *Principles and Practice of Constraint Programming - 17th International Conference, CP 2011*, 2011.

[2] P. Dapic, P. Markovic, and B. Martin, "QCSP on semicomplete digraphs," in *Automata, Languages, and Programming - 41st International Colloquium, ICALP 2014*, 2014, pp. 847–858. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-43948-7_70

[3] H. Chen, "The complexity of quantified constraint satisfaction: Collapsibility, sink algebras, and the three-element case," *SIAM J. Comput.*, vol. 37, no. 5, pp. 1674–1701, 2008.

[4] F. R. Madelaine and B. Martin, "On the complexity of the model checking problem," *CoRR*, vol. abs/1210.6893, 2012. [Online]. Available: http://arxiv.org/abs/1210.6893

[5] T. Feder and M. Vardi, "The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory," *SIAM Journal on Computing*, vol. 28, pp. 57–104, 1999.

[6] A. Bulatov, A. Krokhin, and P. G. Jeavons, "Classifying the complexity of constraints using finite algebras," *SIAM Journal on Computing*, vol. 34, pp. 720–742, 2005.

[7] H. Chen, "A rendezvous of logic, complexity, and algebra," *SIGACT News*, 2006.

[8] ——, "Meditations on quantified constraint satisfaction," in *Logic and Program Semantics - Essays Dedicated to Dexter Kozen on the Occasion of His 60th Birthday*, 2012, pp. 35–49.

[9] F. R. Madelaine and B. Martin, "Containment, equivalence and coreness from CSP to QCSP and beyond," in *Principles and Practice of Constraint Programming - 18th International Conference, CP 2012, Québec City, QC, Canada, October 8-12, 2012. Proceedings*, ser. Lecture Notes in Computer Science, M. Milano, Ed., vol. 7514. Springer, 2012, pp. 480–495. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-33558-7_36

[10] H. Chen, "Quantified constraint satisfaction and the polynomially generated powers property," *Algebra universalis*, vol. 65, no. 3, pp. 213–241, 2011, an extended abstract appeared in ICALP B 2008. [Online]. Available: http://dx.doi.org/10.1007/s00012-011-0125-4

[11] J. Wiegold, "Growth sequences of finite semigroups," *Journal of the Australian Mathematical Society (Series A)*, vol. 43, pp. 16–20, 8 1987, communicated by H. Lausch. [Online]. Available: http://journals.cambridge.org/article_S1446788700028925

[12] H. Chen, F. Madelaine, and B. Martin, "Quantified constraints and containment problems," in *23rd Annual IEEE Symposium on Logic in Computer Science*, 2008, pp. 317–328.

[13] F. Börner, A. A. Bulatov, H. Chen, P. Jeavons, and A. A. Krokhin, "The complexity of constraint satisfaction games and qcsp," *Inf. Comput.*, vol. 207, no. 9, pp. 923–944, 2009.

[14] B. Martin and F. Madelaine, "Towards a trichotomy for quantified H-coloring," in *2nd Conf. on Computatibility in Europe, LNCS 3988*, 2006, pp. 342–352.

[15] F. R. Madelaine and B. Martin, "A tetrachotomy for positive first-order logic without equality," in *LICS*, 2011, pp. 311–320.

[16] T. Feder, P. Hell, P. Jonsson, A. A. Krokhin, and G. Nordh, "Retractions to pseudoforests," *SIAM J. Discrete Math.*, vol. 24, no. 1, pp. 101–112, 2010.

[17] F. R. Madelaine and B. Martin, "The complexity of positive first-order logic without equality," in *LICS*. IEEE Computer Society, 2009, pp. 429–438.

[18] H. Chen, F. R. Madelaine, and B. Martin, "Quantified constraints and containment problems," *CoRR*, vol. abs/1310.1016, 2013.

Material omitted from § 3.

Partially reflexive paths (*c.f.*3.1)

Cases with the PGP

In the proof of Lemma 4, we refer to the fact that loop-connected p.r. paths have a majority polymorphism. In the reference, it is not fully explicit how one builds such a majority operation, and we highlight it here for the sake of completeness.

Let $\mathcal{P}$ be a loop-connected path labelled in ascending natural numerical order. Let L be the irreflexive component left of the central loops and R be the irreflexive component right of the loops. If there are no loops let the whole path be in L.

Recall first that the operation median over the elements of $\mathcal{P}$ returns the argument that is neither minimal, nor maximal when the arguments are pairwise distinct, and behave as a majority operation otherwise.

Define $\text{Feder}(x, y, z) := \text{median}(x, y, z)$, if all x, y, z have the same parity, and $\text{Feder}(x, y, z)$:=max of the repeated parity, otherwise (this operation was communicated to one of the author by email by Tomás Feder, hence its name).

We define $f(x, y, z) := \text{Feder}(x, y, z)$, if $x, y, z \subset L$ or $x, y, z \subset R$, and $f(x, y, z) := \text{median}(x, y, z)$, otherwise. This operation f is a majority polymorphism and a polymorphism of $\mathcal{P}$.

Lemma 5. Let $\mathcal{P}_{0^a1^b\alpha}$, with $b > 0$, be a quasi-loop-connected path on vertices $[n]$. For each $y \in [n]$ there is a binary idempotent polymorphism f_y of $\mathcal{P}_{0^a1^b\alpha}$ so that $f_y(1, x) = x$ (for all x) and $f_y(n, 1) = y$.

Proof. Let y be given. Suppose $\mathcal{P}_{0^a1^b\alpha}$ is of odd length and has centre at position q (the argument for even length is very similar with central vertices q, q'). Choose p minimal ($1 \leq p \leq q$) so it is a looped vertex. Let r be so that $r - q = q - p$, i.e. p and r are first and last in the block 1^b, and we have $1 \leq p \leq q \leq r \leq n$). An idempotent binary polymorphism on domain $[n]$ may be visualised as a matrix X with leading diagonal $1, \ldots, n$. We consider the top-left and bottom-left parts of the matrix X_{tl} and X_{bl}, respectively, to include as their farthest right column the central column of the matrix X at position q. X_{tl} and X_{bl} will also overlap on the bottom row of the former which is the top row of the latter. Let us consider what constraints a polymorphism must satisfy. Across the whole matrix, diagonal neighbours must be adjacent elements. In X_{tl}, in fact, only the diagonals are needed to be considered to satisfy polymorphism. But in X_{bl} (indeed the whole bottom half) there may be some horizontal lines that must satisfy the adjacency condition and in the right half there might be some vertical lines that need to satisfy this adjacency condition too. To see an example of this we direct the reader to $\mathcal{P}_{00001110110}$ in Figure 3. We will rebuild X_{tl} and X_{bl} to satisfy all horizontals, even though we do not need them all, and the right half of the matrix will satisfy all potential vertical lines.

When viewed as a matrix, the entire right half (from and including the middle column q) will obey $f(u, v) = v$. We now turn our attention to X_{tl}. The farthest right column of X_{tl} is already set to q, and we will set the entire bottom row to q. We now remove these already-set positions and then set the farthest right column and bottom row of the remainder of X_{tl} to $q-1$. We iterate this until we reach and have done this for p. Note that this is consistent with idempotency. We have now filled in X other than a $(p-1)\times(p-1)$ matrix in the top-left which we call X'_{tl} and a $(q-1)\times(q-1)$ matrix in the bottom-left which we still call X_{bl}. This is depicted in Figure 1 and satisfies all local conditions for polymorphism. The matrix X'_{tl} must additionally satisfy leading diagonal idempotency and must also satisfy the boundary condition of p against its right-most column and bottom row. The matrix X_{bl} must satisfy position $(n, 1)$ being y and the boundary condition of q against its right-most column and top row.

(Construction of X'_{tl}.) We explain how to fill in position $(1, i)$ and $(i, 1)$ for $i \in [p-1]$ because each diagonal proceeding towards the centre of the matrix will contain an increasing arithmetic sequence with step 1. Set $(1, i)$ to be i and $(i, 1)$ to be i (when i is odd) and $i+1$ (when i is even). A simple calculation now yields the precise specification: if $\lambda < \mu$, set (λ, μ) to μ; if $\lambda > \mu$, set (λ, μ) to λ (if $\lambda - \mu + 1$ is odd) and to $\lambda + 1$ (if $\lambda - \mu + 1$ is even). It is easy to see that this satisfies polymorphism. Indeed, it satisfies polymorphism on the horizontals where it is not necessary (but will become necessary for X_{bl}).

(Construction of X_{bl}.) The upward diagonal from y at position $(n, 1)$ to q is filled $y, y \pm 1 \ldots, q, \ldots, q$. That is, if $y \leq q$ we increase by one until we reach q and then repeat q, and if $y \geq q$ we decrease by one until we reach q and then repeat q.

All rows and columns in X_{bl} that contain a vertex $z \in \{p, \ldots, q, \ldots, r\}$ on the upward diagonal from $(n, 1)$ are now filled in with z. At this point we are left with some $s \times s$ submatrix X'_{bl} of X_{bl} not filled in. X'_{bl} might be empty if $y \in \{p, \ldots, q, \ldots, r\}$, but if X'_{bl} is not empty then we have the boundary condition of either p or r against its right-most column and top row. We now fill this in in precisely the dual fashion to our filling in of X'_{tl}. We will give the argument when the boundary condition is r (the other case of boundary p being very similar). We explain how to fill in position (n, i) and (i, n) for $i \in \{n, \ldots, n-s+1)$ because each diagonal proceeding towards the centre of the matrix will contain a decreasing (increasing if boundary is instead p) arithmetic sequence with step 1. Set (n, i) to be $y - i + 1$ and (i, n) to be $i - n + y$ (when i is odd) and $i - 1 - n + y$ (when i is even). It is not hard to see that this satisfies polymorphism, even on its horizontals.

Two examples, for the graph $\mathcal{P}_{0^4 1^3 \alpha}$ with $|\alpha| = 4$, are given in Figure 2. The left-hand example is for ($n = 11$ where $p = 5, q = 6, r = 7$ and) $y = 10$; and the right-hand example is for ($n = 11$ where $p = 5, q = 6, r = 7$ and) $y = 3$. $\square$

$$
\begin{array}{ccccccccc}
1 & & p & \cdots & q & \cdots & r & \cdots & n \\
 & \ddots & \vdots & & \vdots & & \vdots & & \vdots \\
p & \cdots & p & \cdots & q & \cdots & r & \cdots & n \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
q & \cdots & q & \cdots & q & \cdots & r & \cdots & n \\
 & & & & \vdots & & \vdots & & \vdots \\
 & & & & q & \cdots & r & \cdots & n \\
 & & & & \vdots & & \vdots & & \vdots \\
y & & & & q & \cdots & r & \cdots & n
\end{array}
$$

Figure 1: First part of the construction for the proof of Lemma 5.

$$
\begin{array}{ccccccccccc}
1 & 2 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
3 & 2 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
3 & 4 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
5 & 4 & 5 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
\cline{1-4}
5 & 5 & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
6 & 6 & 6 & 6 & 6 & 6 & 7 & 8 & 9 & 10 & 11 \\
7 & 7 & 7 & 7 & 7 & 6 & 7 & 8 & 9 & 10 & 11 \\
7 & 7 & 7 & 7 & 7 & 6 & 7 & 8 & 9 & 10 & 11 \\
\cline{1-3}
8 & 7 & \multicolumn{1}{c|}{8} & 7 & 7 & 6 & 7 & 8 & 9 & 10 & 11 \\
8 & 9 & \multicolumn{1}{c|}{8} & 7 & 7 & 6 & 7 & 8 & 9 & 10 & 11 \\
10 & 9 & \multicolumn{1}{c|}{8} & 7 & 7 & 6 & 7 & 8 & 9 & 10 & 11
\end{array}
$$

$$
\begin{array}{ccccccccccc}
1 & 2 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
3 & 2 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
3 & 4 & 3 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
5 & 4 & 5 & \multicolumn{1}{c|}{4} & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
\cline{1-4}
5 & 5 & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
6 & 6 & 6 & 6 & 6 & 6 & 7 & 8 & 9 & 10 & 11 \\
5 & 5 & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
5 & 5 & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
5 & 5 & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
\cline{1-2}
5 & \multicolumn{1}{c|}{4} & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
3 & \multicolumn{1}{c|}{4} & 5 & 5 & 5 & 6 & 7 & 8 & 9 & 10 & 11
\end{array}
$$

Figure 2: Two polymorphism of the graph $\mathcal{P}_{0^4 1^3 \alpha}$, with α any string of 0s and 1s of length 4. The lines indicate the boundaries of X'_{tl} and X'_{bl}.

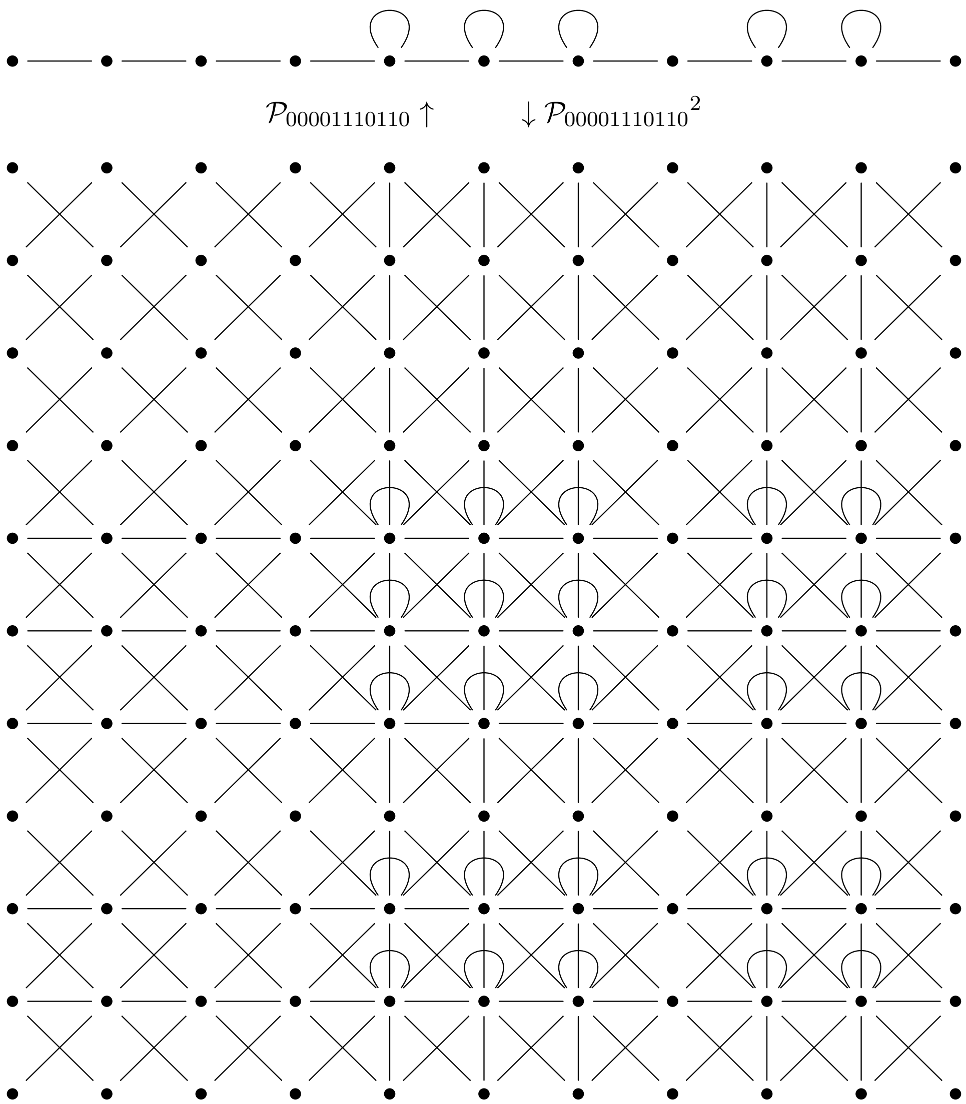

Figure 3: Path $\mathcal{P}_{0^41^301^20}$ and its square

Lemma 50. *Let $\mathcal{P}_{0^a\alpha}$ be a quasi-loop-connected path on vertices $[n]$ (that is not of the form $\mathcal{P}_{0^a1^b\alpha}$ with $|\alpha| = a$). For each $y \in [n]$ there is a binary idempotent polymorphism f_y of $\mathcal{P}_{0^a\alpha}$ so that $f_y(1, x) = x$ (for all x) and either $f_y(n, 1) = y$ or $f_y(n, 2) = y$.*

Proof. Suppose first that $y \in \{1, 2\}$. Assume n is even (the argument for the odd case is very similar). Our proof has similarities to that of Lemma 5. We will rebuild X_{tl} roughly as before, but now we rebuild X_{bl} as a mirror image of X_{tl}. We will set the columns $n/2 + 1, \ldots, n$ of our matrix to be full columns of $n/2 + 1, \ldots, n$, respectively. Take the remainder of the matrix, on columns $1, \ldots, n/2$ and split it into two across

$$\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 4 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 4 & 5 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 6 & 5 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
5 & 6 & 5 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 4 & 5 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 4 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10
\end{array}$$

$$\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 4 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 4 & 5 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 6 & 5 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
7 & 6 & 7 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
\hline
7 & 6 & 7 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 6 & 5 & 6 & 5 & 6 & 7 & 8 & 9 & 10 \\
5 & 4 & 5 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
3 & 4 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10
\end{array}$$

Figure 4: Examples for the proof of Lemma 50. The line indicates the axis of symmetry for the mirror image.

its central horizontal. Call the top left X_{tl} and the bottom left X_{bl}. We will fill in X_{tl} according to the construction of X'_{tl} from Lemma 5. We will now fill X_{bl} to be a mirror image of X_{tl} in the central horizontal. An example of this is depicted in Figure 4.

Now for general $y \in \{z, z+1\}$ with z odd, we shift the horizontal split between X_{tl} and X_{bl} downwards (making X_{tl} larger). The split will be just after row $n/2+(z-1)/2$, i.e. at $z = n-1$ the matrix X_{bl} is empty. We now build X_{tl} according to the construction of X'_{tl} from Lemma 5 and now fill X_{bl} to be a mirror image of the bottom part of X_{tl} in the horizontal just after row $n/2+(z-1)/2$. An example of this is depicted in Figure 4. □

Lemma 7. Let $\mathcal{P}_{0^a\alpha}$, for $|\alpha| \in \{a, a-1\}$, be a quasi-loop-connected path on vertices $[n]$ (that is not of the form $\mathcal{P}_{0^a1^b\alpha}$ with $|\alpha| = a$). Let $\mathbb{A}$ be the algebra specified by id-Pol($\mathcal{P}_{0^a\alpha}$). For each m, $\mathbb{A}^m$ is generated from the $2n+2$ m-tuples $(1,1,\ldots,1)$, $(2,2,\ldots,2), (n,1,\ldots,1), (1,n,\ldots,1), \ldots, (1,1,\ldots,n)$,$(n,2,\ldots,2), (2,n,\ldots,2), \ldots, (2,2,\ldots,n)$.

Proof. The proof is as in the Lemma 6 but relies upon Lemma 50 in place of Lemma 5. □

Cases with the EGP

For a digraph $\mathcal{H}$, the distance, $d_{\mathcal{H}}$, between two vertices is the number of edges in a shortest path connecting them. By $\mathcal{H}^n$ we mean the tensor product of $\mathcal{H}$ with itself n times.

We note that polymorphisms do not increase distances in graphs, i.e. if f is an n-ary polymorphism of $\mathcal{H}$ and $u, v \in H^n$ then $d_{\mathcal{H}^n}(u, v) \leq d_{\mathcal{H}}(f(u), f(v))$.

Lemma 8 is proved by induction on the arity of the polymorphisms. We deal first with the base case.

Lemma 51. *Let α be any sequence of zeros and ones. All idempotent binary polymorphisms of $\mathcal{P}_{10\alpha01}$ are projections.*

Proof. We label the vertices of $\mathcal{P} = \mathcal{P}_{10\alpha01}$ left to right over $0, 1, \ldots, t = |\mathcal{P}| - 1$ and start by showing that any binary polymorphism f of $\mathcal{P}$ must satisfy the following

$$f(i, j) \leq \max\{i, j\} \quad \text{and} \quad f(i, j) \geq \min\{i, j\},$$

with $i, j = 0, \ldots, t$ and considering the natural linear ordering of the labelling of vertices of $\mathcal{P}$. Assume, for a contradiction, that there exist i, j such that $f(i, j) = k$ with $k > i, j$. Without loss of generality we assume that $i < j$. There exists a path of length at most j from $f(i, j)$ to $f(0, 0) = 0$, via the vertices $f(i-1, j-1), \ldots, f(0, 1), f(0, 0)$, but clearly $d_{\mathcal{H}}(0, k) = k$, so we get a contradiction. Dually, we can show that we also cannot have $k < i, j$.

We now show that $f_{|\{x,x+k\}}$ is, without loss of generality, the first projection, by induction on $k \geq 1$.

There is an edge, in $\mathcal{P}$, from $f(0, 1)$ and from $f(1, 0)$ to $f(0, 0) = 0$, so $f(0, 1), f(1, 0) \in \{1, 0\}$. There is also an edge from $f(0, 1)$ to $f(1, 0)$, so they cannot both be equal to 1. In a similar way we can check that $f(t, t-1), f(t-1, t) \in \{t-1, t\}$ and they cannot both be equal to $t - 1$.

We have $d_{\mathcal{H}}(f(0, 1), f(t-1, t), d_{\mathcal{H}}(f(1, 0), f(t, t-1) \leq t - 1$, since $f(t, t-1)$ and $f(t-1, t)$ cannot both be equal to $t-1$, this immediately implies that $f(1, 0)$ and $f(0, 1)$ cannot both be equal to 0. Hence it follows that $f_{|\{1,2\}}$ must be a projection. Assume, without loss of generality, that it is the first projection.

To be able to get the correct distances from $f(0, 1)$ to $f(t-1, t)$ we must have that f restricted to any two consecutive vertices must be the first projection, i.e. $f(x, x+1) = x$ and $f(x+1, x) = x+1$ for all $x = 0, \ldots, t-1$.

Now, assume that $f_{|\{x,x+l\}}$ is the first projection, for all $l < m$ and all $x = 0, \ldots, t-l$. We show that $f_{|\{x,x+m\}}$ is also the first projection, by induction on x. For the base case $x = 0$, we know that there is an edge from $f(0, m)$ to $f(0, m-1)$ and an edge from $f(m, 0)$ to $f(m-1, 0)$. By the inductive hypothesis, $f(0, m-1) = 0$ and $f(m-1, 0) = m-1$, so we must have $f(0, m) \in \{1, 0\}$ and $f(m, 0) \in \{m-2, m-1, m\}$.

Also, there is an edge from $f(0,m)$ to $f(1,m-1)$, by the inductive hypothesis $f(1,m-1)=1$, so we must have $f(0,m)=0$. We now just need to consider the case $f(m,0)$.

Case 1: Suppose that $f(m,0)=m-2$; $d_{\mathcal{H}}(f(m,0),f(t,t-(m+1))\leq t-m$, and by the inductive hypothesis $f(t,t-(m+1))=t$. So $d_{\mathcal{H}}(f(m,0),t)\leq t-m$, but $d_{\mathcal{H}}(m-2,t)=t-(m-2)$, so we get a contradiction.

Case 2: Suppose that $f(m,0)=m-1$. Since there is an edge from $f(m,0)$ to $f(m-1,0)$, and $f(m-1,0)=m-1$ by the inductive hypothesis, it follows that $m-1$ must be a loop. Now, there is an edge from $f(m,0)$ to $f(m+1,1)$ and from this to $f(m,2)$. Since $f(m,2)=m$, by the inductive hypothesis. We have that $f(m+1,1)\in\{m-1,m\}$. If $f(m+1,1)=m-1$ we get a similar contradiction as in Case 1, so we must have $f(m+1,1)=m$, which also implies that m must be a loop. We now move on to $f(m+2,2)$ and using the same reasoning we get that $m+2$ must be a loop. Carrying on in this way we will eventually reach a contradiction since the vertex $t-1$ does not have a loop; unless $m=t$, in which case $f(t,0)$ is a loop and we immediately have $f(t,0)=t$.

Hence we must have $f(m,0)=m$. This proves the base case.

Assume now that $f_{|\{x,x+m\}}$ is the first projection for all $x<b$. We show that $f_{|\{b,b+m\}}$ is also the first projection. There are edges from $f(b,b+m)$ and $f(b+m,b)$ to $f(b-1,b-1+m)$ and $f(b-1+m,b-1)$ respectively. By the inductive hypothesis, $f(b-1,b-1+m)=b-1$ and $f(b-1+m,b-1)=b-1+m$. So we have $f(b,b+m)\in\{b-2,b-1,b\}$ and $f(b+m,b)\in\{b-2+m,b-1+m,b+m\}$.

There is an edge from $f(b,b+m)$ to $f(b+1,b+m-1)$, and, by the inductive hypothesis, $f(b+1,b+m-1)=b+1$. So we must have $f(b,b+m)\in\{b,b+1,b+2\}$, it immediately follows that $f(b,b+m)=b$. Like above, in Cases 1 and 2, we can show that we also must have $f(b+m,b)=b+m$.

This proves the lemma. □

Lemma 8. Let α be any sequence of zeros and ones. All idempotent polymorphisms of $\mathcal{P}_{10\alpha01}$ are projections.

Proof. Let $\mathcal{P}=\mathcal{P}_{10\alpha01}$ and label the vertices of $\mathcal{P}$ over $[t]$ with $t=|\mathcal{P}|$ left to right. Let $n\geq 2$ be arbitrary and let $f(x_1,\ldots,x_n)$ be any idempotent n-ary polymorphism of $\mathcal{P}$. We prove the lemma by induction on n, with base case given by Lemma 51.

Assume now that the lemma holds for any $n<k$, i.e. we have $f(x_1,\ldots,x_n)=x_1$ for any $x_1,\ldots,x_n$ vertices of $\mathcal{P}$ and any $n<k$. Let us consider the case when f is a polymorphism of arity k. We will show that $f(x_1,\ldots,x_k)$ is also the first projection.

Case 1: x_1 is not the left-most nor the right-most element of $x_1,\ldots,x_k$.

In this case we know that $d(f(x_1,\ldots,x_k),f(1,y_2,\ldots,y_k))\leq x_1-1$, where $y_i=x_i-x_1$ if $x_i>x_1$ and it is 1 otherwise. Now at least one of the y_is equals to 1, so at this stage $f(1,y_2,\ldots,y_k)$ matches a polymorphism of arity smaller than k and we can apply the inductive hypothesis, so that $f(1,y_2,\ldots,y_k)=1$. It follows that $d(f(x_1,\ldots,x_k),1)\leq$

$x_1 - 1$. In a similar way we obtain that $d(f(x_1,\dots,x_k), f(t,z_2,\dots,z_k)) \leq t - x_1 - 1$, so that $d(f(x_1,\dots,x_k),t) \leq t - x_1 - 1$. It follows that $f(x_1,\dots,x_k) = x_1$.

Case 2: $x_1 = 1$; Assume, wlog, that x_2 is the left-most element of $x_2,\dots,x_k$ and is not equal to x_1. Then $d(f(1,x_2,\dots,x_k), f(1,1,y_3,\dots,y_k)) \leq x_2 - 1$, with y_i defined as above. By the inductive hypothesis $f(1,1,y_3,\dots,y_k) = 1$, so that $d(f(1,x_2,\dots,x_k),1) \leq x_2 - 1$, hence $f(1,x_2,\dots,x_k) \leq x_2$. We show that $f(1,x_2,\dots,x_k) = 1$ by induction on x_2. If $x_2 = 2$ then $f(1,x_2,\dots,x_k) \in \{1,2\}$, and we know that there is an edge from $f(1,2,x_3\dots,x_k)$ to $f(2,1,x_3-1,\dots,x_k-1)$. Since there are no loops at 2 and, by Case 1, $f(2,1,x_3-1,\dots,x_k-1) = 2$, we must have $f(1,x_2,\dots,x_k) = 1$.

Assume now that the result holds whenever $x_2 < z$. Then $f(1,z,x_3,\dots,x_k) \leq z$ and there is an arc from this vertex to $f(1,z-1,x_3-1,\dots,x_k-1)$, since $f(1,z-1,x_3-1,\dots,x_k-1) = 1$ by the inductive hypothesis, we must have $f(1,z,x_3,\dots,x_k) \in \{1,2\}$.

Suppose, for a contradiction, that $f(1,z,x_3,\dots,x_k) = 2$. Since there is an arc from this vertex to $f(2,z-1,x_3-1,\dots,x_k-1)$ we have that $f(2,z-1,x_3-1,\dots,x_k-1) \in \{1,3\}$. Now $d(f(2,z-1,x_3-1,\dots,x_k-1), f(\lceil z/2\rceil+1, \lceil z/2\rceil, x_3 - \lceil z/2\rceil,\dots,x_k - \lceil z/2\rceil) \leq \lceil z/2\rceil - 1$. By Case 1, we know that $f(\lceil z/2\rceil+1, \lceil z/2\rceil, x_3-\lceil z/2\rceil,\dots,x_k-\lceil z/2\rceil) = \lceil z/2\rceil + 1$. So, we cannot have $f(2,z-1,x_3-1,\dots,x_k-1) = 1$. It follows that $f(2,z-1,x_3-1,\dots,x_k-1) = 3$, and since there is an arc from this vertex to $f(1,z_2,x_3,\dots,x_k)$ and, by the inductive hypothesis, $f(1,z_2,x_3,\dots,x_k) = 1$, we get a contradiction.

Case 3: x_1 is the left-most element of $x_1,\dots,x_k$, but is not equal to 1.

In this case we know that $d(f(1,x_2-x_1,\dots,x_k-x_1), f(x_1,\dots,x_k)) \leq x_1 - 1$, by Case 2 we know that $f(1,x_2-x_1,\dots,x_k-x_1) = 1$ it then follows that $f(x_1,\dots,x_k) \leq x_1$. Since we have already seen that $f(x_1,\dots,x_k) \geq x_1$, because x_1 is the left-most element, it immediately follows that $f(x_1,\dots,x_k) = x_1$. This proves the claim. □

EGP Methodology via Galois correspondence

The following is a restatement of the backward inclusion of the well-known Galois correspondence $\mathrm{Inv}(\mathrm{sPol}(\mathcal{B})) = \langle\mathcal{B}\rangle_{\mathrm{pH}}$ holding for finite structures $\mathcal{B}$ [13]. This direction can be proved by induction on the term-complexity of $\varphi \in \langle\mathcal{B}\rangle_{\mathrm{pH}}$.

Lemma 52. *Let $\mathcal{B}$ be a finite structure and suppose there is a k-ary surjective polymorphism of $\mathcal{B}$ that [pointwise] maps the tuples $(x_1^1,\dots,x_1^r)$, ..., $(x_k^1,\dots,x_k^r)$ to $(y^1,\dots,y^r)$. Let φ be an r-ary relation from $\langle\mathcal{B}\rangle_{\mathrm{pH}}$. If φ holds on each of $(x_1^1,\dots,x_1^r)$, ..., $(x_k^1,\dots,x_k^r)$ in $\mathcal{B}$, then φ holds on $(y^1,\dots,y^r)$ in $\mathcal{B}$.*

Together with the definition of a generating set, it can be used to derive Corollary 9.

Material omitted from § 4.

Games, adversaries and reactive composition (*c.f.*4.1)

Theorem 26. Let φ be a pH-sentence with m universal variables. Let $\mathscr{A}$ be an adversary and Ω_m a set of adversaries, both of length m.

If $\mathcal{A} \models \varphi_{\restriction \Omega_m}$ and $\mathscr{A} \trianglelefteq \Omega_m$ then $\mathscr{A} \models \varphi$.

Proof. We sketch the proof for the sake of completeness. Let $\Omega_m := \{\mathscr{B}_1, \dots, \mathscr{B}_k\}$ and f and g^i_j be as in the definition of reactive composition and witnessing that $\mathscr{A} \trianglelefteq f(\mathscr{B}_1, \dots, \mathscr{B}_k)$. Assume also that $\mathcal{A} \models \varphi_{\restriction \Omega_m}$. Given any sequence of play of the universal player according to the adversary $\mathscr{A}$, that is v_1 is played as $a_1 \in A_1$, v_2 is played as $a_2 \in A_2$, etc., we "go backwards through f" via the maps g^i_j to pinpoint *incrementally* for each $j \in [k]$ a sequence of play $v_1 = g^1_j(a_1)$, $v_2 = g^2_j(a_1, a_2)$ etc, thus yielding eventually a tuple that belongs to adversary $\mathscr{B}_j$. After each block of universal variables, we lookup the winning strategy for the existential player against each adversary $\mathscr{B}_j$ and "going forward through f", that is applying f to the choice of values for an existential variable against each adversary, we obtain a consistent choice for this variable against adversary $\mathscr{A}$ (this is because f is a polymorphism and the quantifier-free part of the sentence φ is conjunctive positive). Going back and forth we obtain eventually an assignment to the existential variables that is consistent with the universal variables being played as $a_1, a_2, \dots, a_m$. ☐

Remark 53. *In Chen's work on QCSP, constants are almost always allowed in the constraint language. This amounts with our definition to consider a relational structure $\mathcal{A}$ with all its elements named by constants. However, Chen does not necessarily explicitly add constants to the constraint language and instead moves rapidly to the algebraic setting and considers algebra. There he insists on additional technical conditions which preserves constants. For example in the above theorem, he has the additional condition that f is an* idempotent *polymorphism. Whenever we will use one of Chen's result, we will generalise it as above by considering arbitrary constraint languages and dropping technical conditions such as idempotency from the statement.*

The Π_2-case (*c.f.*4.2)

Lemma 27 (principle of union). *Let Ω_m be a set of adversaries of length m and φ a Π_2-sentence with m universal variables. Let $\mathscr{O}_{\cup\Omega} := \bigcup_{\mathscr{O} \in \Omega} \mathscr{O}$ and $\Omega_{tuples} := \{\{t\} \in \mathscr{O}_{\cup\Omega}\} = \bigcup_{\mathscr{O} \in \Omega} \{\{t\} \in \mathscr{O}\}$. We have the following equivalence.*

$$\mathcal{A} \models \varphi_{\restriction \Omega_m} \iff \mathcal{A} \models \varphi_{\restriction \mathscr{O}_{\cup\Omega}} \iff \mathcal{A} \models \varphi_{\restriction \Omega_{tuples}}$$

The forward implications

$$\mathcal{A} \models \varphi_{\restriction \Omega_m} \implies \mathcal{A} \models \varphi_{\restriction \mathscr{O}_{\cup\Omega}} \implies \mathcal{A} \models \varphi_{\restriction \Omega_{\mathrm{tuples}}}$$

of Lemma 27 hold clearly for arbitrary pH-sentences. The proof is trivial and is a direct consequence of the following obvious fact.

Fact 54. *Let Ω_m be a set of adversaries of length m and φ a Π_2-sentence with m universal variables.*

$$\mathcal{A} \models \varphi_{\restriction \Omega_m}$$
$$\Updownarrow$$
$$\forall \mathscr{O} \in \Omega_m \forall t = (a_1, \ldots, a_m) \in \mathscr{O}\ \mathcal{A} \models \varphi_{\restriction \{t\}}$$

Remark 55 (following Lemma 27)**.** *For a sentence that is not Π_2, this does not necessarily hold. For example, consider $\forall x \forall y \exists z \forall w\ E(x,z) \wedge E(y,z) \wedge E(w,z)$ on the irreflexive 4-clique $\mathcal{K}_4$. The sentence is not true, but for all individual tuples (x_0, y_0, w_0), we have $\exists z\ E(x_0,z) \wedge E(y_0,z) \wedge E(w_0,z)$.*

Proposition 28. Let φ be a Π_2-pH-sentence with m universal variables. Let $\mathscr{A}$ be an adversary and Ω_m a set of adversaries, both of length m.

If $\mathcal{A} \models \varphi_{\restriction \Omega_m}$ and Ω_m generates $\mathscr{A}$ then $\mathcal{A} \models \varphi_{\restriction \mathscr{A}}$.

Proof. The hypothesis that Ω_m generates $\mathscr{A}$ can be rephrased as follows : for each tuple t in $\mathscr{A}$, $\{t\} \trianglelefteq f_t(t_1, t_2, \ldots, t_k)$, where $t_1, t_2, \ldots, t_k$ belong to Ω_{tuples}. To see this, it remains to note that the suitable g_i^j's from the definition of composition are induced trivially as there is no choice: for every j in $[k]$ and every i in $[m]$ pick $g_i^j(a_1, a_2, \ldots, a_i) = t_{i,j}$ where $t_{i,j}$ is the ith element of t_j. So by Theorem 26, if $\mathcal{A} \models \varphi_{\restriction \Omega_{\text{tuples}}}$ then $\mathcal{A} \models \varphi_{\restriction \{t\}}$. As this holds for any tuple t in $\mathscr{A}$, via the principle of union, it follows that $\mathcal{A} \models \varphi_{\restriction \mathscr{A}}$. □

Proposition 29. Let Ω_m be a set of adversaries of length m that is not degenerate. The following are equivalent.

(i) for any Π_2-pH sentence ψ, $\mathcal{A} \models \psi_{\restriction \Omega_m}$ implies $\mathcal{A} \models \psi$.

(ii) for any Π_2-pH sentence ψ, $\mathcal{A} \models \psi_{\restriction \mathscr{O}_{\cup \Omega}}$ implies $\mathcal{A} \models \psi$.

(iii) for any Π_2-pH sentence ψ, $\mathcal{A} \models \psi_{\restriction \Omega_{\text{tuples}}}$ implies $\mathcal{A} \models \psi$.

(iv) $\mathcal{A} \models \varphi_{\mathscr{O}_{\cup \Omega}, \mathcal{A}}$

(v) $\mathcal{A} \models \varphi_{\Omega_{\text{tuples}}, \mathcal{A}}$

(vi) Ω_m generates A^m.

Proof. The first three items are equivalent by Lemma 27 (these implications have the same conclusion and equivalent premises). The fourth and fifth items are trivially equivalent since $\varphi_{\mathscr{O}_{\cup \Omega}, \mathcal{A}}$ and $\varphi_{\Omega_{\text{tuples}}, \mathcal{A}}$ are the same sentence.

We show the implication from the third item to the fifth. By construction, $\varphi_{\Omega_{\text{tuples}}, \mathcal{A}}$ is Π_2 and it suffices to show that there exists a winning strategy for $\exists$ against any adversary $\{t\}$ in Ω_{tuples}. This is true by construction. Indeed, note that there exists a winning strategy for $\exists$ in the $(\mathcal{A}, \varphi_{\Omega_{\text{tuples}}, \mathcal{A}})$-game against adversary $\{t\}$ iff there is a homomorphism from the $\sigma^{(m)}$-structure $\bigotimes_{t' \in \Omega_{\text{tuples}}} \mathfrak{A}_{\mu_{t'}}$ to the $\sigma^{(m)}$-structure $\mathfrak{A}_{\mu_t}$, where $\mu_t : [m] \to A$ is the map induced naturally by t. The projection is such a homomorphism.

The penultimate item implies the last one: instantiate the universal variables of $\varphi_{\Omega_{\text{tuples}},\mathcal{A}}$ as given by the m-tuple t and pick for f_t the homomorphism from the product structure witnessing that $\exists$ has a winning strategy.

Finally, the last item implies the first one by Proposition 28. □

The unbounded case (*c.f.*4.3)

Lemma 34. Let $\mathbf{\Omega}$ be a sequence of sets of adversaries that has the m-projectivity property for some $m \geq 1$ such that $\Omega_{n.m}$ is not degenerate. The following holds.

(i) $\mathcal{A} \models \psi_{\restriction \Omega_{\mathbf{n.m}}}$, where $\psi = \varphi_{n,\Omega_{\mathbf{m}},\mathcal{A}}$

(ii) If for every Π_2-sentence ψ with $m.n$ universal variables, it holds that $\mathcal{A} \models \psi_{\restriction \Omega_{\mathbf{m.n}}}$ implies $\mathcal{A} \models \psi$, then $\mathcal{A} \models \varphi_{n,\Omega_{\mathbf{m}},\mathcal{A}}$.

Proof. The second statement is a direct consequence of the first one. The proof of the first statement generalises an argument used in the proof of Proposition 29. Consider any adversary $\mathscr{O}$ in $\Omega_{n.m}$. For convenience, we name the positions of this adversary in a similar fashion to the universal variables of the sentence, namely by a pair (i, j) in $[n] \times [m]$. By projectivity, there exists an adversary $\mathscr{O}'$ in Ω_m which dominates any adversary $\tilde{\mathscr{O}}$ in Proj$\mathscr{O}$ (obtained by projecting over an arbitrary choice of one position in each of the m blocks of size n). In the product structure underlying the formula $\varphi_{n,\Omega_m,\mathcal{A}}$, we consider the following structure:

$$\bigotimes_{\mu \in A^{[n.m]}_{\restriction \mathscr{O}'}} \mathfrak{A}_{\mathscr{O}',\mu}$$

An instantiation of the universal variables of $\varphi_{n,\Omega_m,\mathcal{A}}$ according to some tuple t from the adversary $\mathscr{O}$ corresponds naturally to a map μ_t from $[n] \times [m]$ to A. Observe that our choice of $\mathscr{O}'$ ensures that this map μ_t is consistent with $\mathscr{O}'$. An instantiation of the universal variables by μ_t induces a $\sigma^{(n.m)}$-structure $\mathfrak{A}_{\mu_t}$ and a winning strategy for $\exists$ amounts to a homomorphism from the product $\sigma^{(n.m)}$-structure underlying the sentence to this $\mathfrak{A}_{\mu_t}$. Since the component $\mathfrak{A}_{\mathscr{O}',\mu_t}$ of this product structure is isomorphic to $\mathfrak{A}_{\mu_t}$, we may take for a homomorphism the corresponding projection. This shows that $\mathcal{A} \models \psi_{\restriction \Omega_{\mathbf{n.m}}}$ where $\psi = \varphi_{n,\Omega_{\mathbf{m}},\mathcal{A}}$. □

Theorem 36. (**In abstracto**.) Let $\mathbf{\Omega}$ be a projective sequence of adversaries, none of which are degenerate. The following are equivalent.

(i) For every $m \geq 1$, For every pH-sentence ψ with m universal variables, $\mathcal{A} \models \psi_{\restriction \Omega_m}$ implies $\mathcal{A} \models \psi$.

(ii) For every $m \geq 1$, for every Π_2-pH-sentence ψ with m universal variables, $\mathcal{A} \models \psi_{\restriction \Omega_m}$ implies $\mathcal{A} \models \psi$.

(iii) For every $m \geq 1$, $\mathcal{A} \models \varphi_{n,\Omega_m,\mathcal{A}}$.

(iv) For every $m \geq 1$, $\mathcal{A} \models \varphi_{\mathscr{O} \cup \Omega,\mathcal{A}}$.

(v) For every $m \geq 1$, $A^m \trianglelefteq \Omega_m$.

(vi) For every $m \geq 1$, Ω_m generates A^m.

Proof. Propositions 29 establishes the equivalence between (ii), (iv) and (vi) for fixed values of m (numbered there as (i), (iv) and (vi), respectively).

To lift these relatively trivial equivalences to the general case, the principle of our current proof no longer preserves the parameter m. The chain of implications of Theorem 35 translates here, once the parameter is universally quantified, to the chain of implications

$$\text{(ii)} \implies \text{(iii)} \implies \text{(v)} \implies \text{(i)}$$

The fact that (i) implies (ii) is trivial[6], which concludes the proof.

□

Corollary 38. Let $\mathcal{A}$ be a structure. Let $\mathbf{\Omega}$ be a sequence of non degenerate adversaries that is effective, projective and polynomially bounded such that Ω_m generates A^m for every $m \geq 1$.

Let $\mathcal{A}'$ be the structure $\mathcal{A}$, possibly expanded with constants, at least one for each element that occurs in $\mathbf{\Omega}$. The problem QCSP($\mathcal{A}$) reduces in polynomial time to CSP($\mathcal{A}'$). In particular, if $\mathcal{A}$ has all constants, the problem QCSP$_c$($\mathcal{A}$) reduces in polynomial time to CSP$_c$($\mathcal{A}$).

Proof. To check whether a pH-sentence φ with m universal variables holds in $\mathcal{A}$, by Theorem 36, we only need to check that $\mathcal{A} \models \varphi_{\restriction \mathscr{B}}$ for every $\mathscr{B}$ in Ω_m. The reduction proceeds as in the proof of [10, Lemma 7.12], which we outline here for completeness.

Pretend first that we reduce $\mathcal{A} \models \varphi_{\restriction \mathscr{B}}$ to a collection of CSP instances, one for each tuple t of $\mathscr{B}$, obtained by instantiation of the universal variables with the corresponding constants. If x is an existential variable in φ, let x_t be the corresponding variable in the CSP instance corresponding to t. We will in fact enforce equality constraints via renaming of variables to ensure that we are constructing Skolem functions. For any two tuples t and t' in $\mathscr{B}$ that agree on their first ℓ coordinates, let Y_ℓ be the corresponding universal variables of φ. For every existential variable x such that Y_x (the universally quantified variables of φ preceding x) is contained in Y_ℓ, we identify x_t with $x_{t'}$. □

[6] We note in passing and for purely pedagogical reason that the implication (v) to (vi) is also trivial, while the natural implication (iii) to (iv) will appear as an evidence to the reader once the definition of the canonical sentences is digested.

Studies of Collapsibility (*c.f.*4.4)

p-collapsibility for $p > 0$

Application 41. A partially reflexive path $\mathcal{A}$ (no constants are present) that is quasi-loop connected has the PGP.

Proof. Indeed, a partially reflexive path $\mathcal{A}$ that is quasi-loop connected has the same QCSP as a partially reflexive path that is loop-connected $\mathcal{B}$ [9] since for some $r_a > 0$ there is a surjective homomorphism g from $\mathcal{A}^{r_a}$ to $\mathcal{B}$ and for some $r_b > 0$ there is a surjective homomorphism h from $\mathcal{B}^{r_b}$ to $\mathcal{A}$ (see main result of [12]). We also know that $\mathcal{B}$ admits a majority polymorphism m [1] and is therefore 2-collapsible from any singleton source (see Table 1) and that Theorem 39 holds for $\mathcal{B}$. Pick some arbitrary element a in $\mathcal{A}$ such that there is some b in $\mathcal{B}$ satisfying $g(a, a, \ldots, a) = b$. Use b as a source for $\mathcal{B}$.

We proceed to lift (vi) of Corollary 39 from structure $\mathcal{B}$ to $\mathcal{A}$, which we recall here for $\mathcal{B}$: for every m, for every tuple t in B^m, there is a polymorphism f_t of $\mathcal{B}$ of arity k and tuples $t_1, t_2, \ldots, t_k$ in $\Upsilon_{m,2,b}$ such that $f_t(t_1, t_2, \ldots, t_k) = t$.

Let g^k denote the surjective homomorphism from $(\mathcal{A}^{r_a})^k$ to $\mathcal{B}^k$ that applies g blockwise. Going back from t_i through g, we can find r_a tuples $t_{i,1}, t_{i,2}, \ldots, t_{i,r_a}$ all in $\Upsilon_{m,2,a}$ (adversaries based on the domain of $\mathcal{A}$) such that $g(t_{i,1}, t_{i,2}, \ldots, t_{i,r_a}) = t_i$. Thus, we can generate any $\widetilde{t}$ in $\mathcal{B}$ via $f_{\widetilde{t}} \circ (g^k)$ from tuples of $\Upsilon_{m,2,a}$.

Let $\hat{t}$ be now some tuple of $\mathcal{A}$. By surjectivity of h, let $\widetilde{t_1}, \widetilde{t_2}, \ldots, \widetilde{t_{r_b}}$ be tuples of $\mathcal{B}$ such that $h(\widetilde{t_1}, \widetilde{t_2}, \ldots, \widetilde{t_{r_b}}) = \hat{t}$. The polymorphism of $\mathcal{A}$ $(f_{\widetilde{t_1}} \circ (g^k), f_{\widetilde{t_2}} \circ (g^k), \ldots, f_{\widetilde{t_{r_b}}} \circ (g^k))$ shows that $\Upsilon_{m,2,a}$ generates $\hat{t}$. This shows that $\mathcal{A}$ is also 2-collapsible from a singleton source. □

Lemma 42. (Chen's lemma.) Let $\mathcal{A}$ be a structure with a constant x. if there is a k-ary polymorphism of $\mathcal{A}$ such that f is surjective when restricted at any position to $\{x\}$, then $\mathcal{A}$ is $k-1$-collapsible from source $\{x\}$ (i.e. $\mathcal{A}$ as a k-ary Hubie polymorphism).

Proof. We sketch the proof for pedagogical reasons. Via Corollary 39, it suffices to show that for any m, A^m is generated by $\Upsilon_{m,k-1,x}$ (instead of the notion of reactive composition).

Consider adversaries of length $m = k$ for now, that is from $\Upsilon_{k,k-1,x}$. If we apply f to these k adversaries, we generate the full adversary A^k. With a picture (adversaries are drawn as columns):

$$f\begin{pmatrix} \{x\} & A & A & \dots & A \\ A & \{x\} & A & \dots & A \\ \vdots & & \ddots & & \vdots \\ A & \dots & A & \{x\} & A \\ A & \dots & A & A & \{x\} \end{pmatrix} = \begin{pmatrix} A \\ A \\ \vdots \\ A \\ A \end{pmatrix} = A^k$$

Expanding these adversaries uniformly with singletons $\{x\}$ to the full length m, we may produce an adversary from $\Upsilon_{m,k,x}$. With a picture for *e.g.* trailing singletons:

$$f\begin{pmatrix} \{x\} & A & A & \dots & A \\ A & \{x\} & A & \dots & A \\ \vdots & & \ddots & & \vdots \\ A & \dots & A & \{x\} & A \\ A & \dots & A & A & \{x\} \\ \{x\} & \{x\} & \{x\} & \dots & \{x\} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \{x\} & \{x\} & \{x\} & \dots & \{x\} \end{pmatrix} = \begin{pmatrix} A \\ A \\ \vdots \\ A \\ A \\ \{x\} \\ \vdots \\ \{x\} \end{pmatrix}$$

Shifting the first additional row of singletons in the top block, we will obtain the family of adversaries from $\Upsilon_{m,k,x}$ with a single singleton in the first $k+1$ positions. It should be now clear that we may iterate this process to derive A^m eventually via some term f' which is a superposition of f and projections and is therefore also a polymorphism of $\mathcal{A}$. $\square$

Remark 56. *An extended analysis of our proof should convince the careful reader that we may in the same fashion prove retroactive composition (the polymorphism's action is determined for a row independently of the others). Thus, appealing to the previous section is not essential, though it does allow for a simpler argument.*

Proposition 43. Let x be a constant in $\mathcal{A}$. The following are equivalent:

(i) $\mathcal{A}$ is collapsible from $\{x\}$.

(ii) $\mathcal{A}$ has a Hubie polymorphism.

Proof. Lemma 42 shows that (ii) implies collapsibility. We prove the converse.

Assume p-collapsibility. By Fact 30, we may apply Theorem 36. For $m = p+1$, item (v) of this theorem states that there is a polymorphism f witnessing that $A^{p+1} \trianglelefteq \Upsilon_{p+1,p,x}$ (diagrammatically, we may draw a similar picture to the one we drew at the beginning of the previous proof). Clearly, f satisfies (ii). $\square$

Theorem 44. (**p-Collapsibility from a singleton source**). Let x be a constant in $\mathcal{A}$. The following are equivalent:

(i) $\mathcal{A}$ is p-collapsible from $\{x\}$.

(ii) For every $m \geq 1$, the full adversary A^m is reactively composable from $\Upsilon_{m,p,x}$.

(iii) $\mathcal{A}$ is Π_2-p-collapsible from $\{x\}$.

(iv) For every $m \geq 1$, $\Upsilon_{m,p,x}$ generates A^m.

(v) $\mathcal{A}$ models $\varphi_{n,\Upsilon_{p+1,p,x},\mathcal{A}}$ (which implies that $\mathcal{A}$ admits a particularly well behaved Hubie polymorphism with source x of arity $(p+1)n^p$).

Proof. Equivalence of the first four points appears in Corollary 39, as does the equivalence with the statement : For every $m \geq 1$, $\mathcal{A}$ models $\varphi_{n,\Upsilon_{m,p,x},\mathcal{A}}$. So they imply trivially the last point by selecting $m = p + 1$.

We show that the last point implies the penultimate one. The proof principle is similar to that of Chen's Lemma. As we have argued similarly before, the last point implies the existence of a polymorphism f. This polymorphism enjoys the following property (each column represents in fact n^p coordinates of A):

$$f\left(\begin{array}{c|c|c|c|c} \{x\} & A & A & \dots & A \\ A & \{x\} & A & \dots & A \\ \vdots & & \ddots & & \vdots \\ A & \dots & A & \{x\} & A \\ A & \dots & A & A & \{x\} \end{array}\right) = \begin{pmatrix} A \\ A \\ \vdots \\ A \\ A \end{pmatrix} = A^{p+1}$$

So arguing as in the proof of Chen's Lemma, we may conclude similarly that for all m, the full adversary A^m is composable from $\Upsilon_{m,p,x}$. □

p-collapsibility for $p > 0$ from a conservative source
We expand on Remark 46.

Theorem 57 (**p-Collapsibility from a conservative source**). *Let B be a subset of the domain of a structure $\mathcal{A}$. Assume further that $\mathcal{A}$ is B-conservative.*

The following are equivalent:

(i) $\mathcal{A}$ is p-collapsible from B.

(ii) $\mathcal{A}$ models $\varphi_{n,\Upsilon_{p+1,p,B},\mathcal{A}}$ (which implies that $\mathcal{A}$ admits a polymorphism f of arity $|B|(p+1)n^p$ that remains surjective when a position i is fixed to a suitable source element b_i in B, and that this polymorphism witnesses that A^m is reactively composable from $\Upsilon_{m,p,B}$).

Proof. Just like the case of singleton source, almost all the proof follows directly from Corollary 39 and similarly we shall only need to prove that the last point implies the penultimate one via a bootstrapping argument.

As we have argued similarly before, the last point implies the existence of a polymorphism f. Let $x_1, x_2, \dots, x_b$ enumerate the elements of the source B. This polymorphism enjoys the following property (each column represents in fact n^p coordinates of A):

$$f\left(\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
\{x_1\} & \{x_2\} & \dots & \{x_b\} & A & A & \dots & A & \dots & A & A & \dots & A \\
A & A & \dots & A & \{x_1\} & \{x_2\} & \dots & \{x_b\} & \dots & A & A & \dots & A \\
A & A & \dots & A & A & A & \dots & A & & A & A & \dots & A \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
A & A & \dots & A & A & A & \dots & A & \dots & \{x_1\} & \{x_2\} & \dots & \{x_b\}
\end{array}\right) = \begin{pmatrix} A \\ A \\ A \\ \vdots \\ A \end{pmatrix} = A^p$$

By conservativity, $f(x_1, x_2 \dots x_b, \dots, x_1, x_2 \dots x_b) \in B$ and we may assume w.l.o.g. that it is in fact equal to x_1. So adding this line at the bottom of the above we may obtain the tuple (A^p, x_1) and (similarly for the other permutations of x_1 within A^p) :

$$f\left(\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
\{x_1\} & \{x_2\} & \dots & \{x_b\} & A & A & \dots & A & \dots & A & A & \dots & A \\
A & A & \dots & A & \{x_1\} & \{x_2\} & \dots & \{x_b\} & \dots & A & A & \dots & A \\
A & A & \dots & A & A & A & \dots & A & & A & A & \dots & A \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
A & A & \dots & A & A & A & \dots & A & \dots & \{x_1\} & \{x_2\} & \dots & \{x_b\} \\
\{x_1\} & \{x_2\} & \dots & \{x_b\} & \{x_1\} & \{x_2\} & \dots & \{x_b\} & \dots & \{x_1\} & \{x_2\} & \dots & \{x_b\}
\end{array}\right) = \begin{pmatrix} A \\ A \\ A \\ \vdots \\ A \\ \{x_1\} \end{pmatrix}$$

Now, we may recopy the above picture replacing in all columns with x_1 at least one of the two occurrences of x_1 by A (we have all permutation of tuples of the form (A^p, x_1)). In particular, we may chose for the last line, the value x_2. Assuming w.l.o.g. that the image of the last line is x_2 (by B-conservativity). We obtain that :

$$f\left(\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
\{x_1\} & \{x_2\} & \dots & \{x_b\} & A & A & \dots & A & \dots & A & A & \dots & A \\
A & A & \dots & A & \{x_1\} & \{x_2\} & \dots & \{x_b\} & \dots & A & A & \dots & A \\
A & A & \dots & A & A & A & \dots & A & & A & A & \dots & A \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
A & A & \dots & A & A & A & \dots & A & \dots & \{x_1\} & \{x_2\} & \dots & \{x_b\} \\
\{x_2\} & \{x_2\} & \dots & \{x_b\} & \{x_2\} & \{x_2\} & \dots & \{x_b\} & \dots & \{x_2\} & \{x_2\} & \dots & \{x_b\}
\end{array}\right) = \begin{pmatrix} A \\ A \\ A \\ \vdots \\ A \\ \{x_2\} \end{pmatrix}$$

Iterating this trick, replacing this time the last occurrence of x_1 and x_2 (from our original picture) by x_3, we will obtain a value in B that differs from x_1 and x_2, say x_3 w.l.o.g. Eventually, we show that A^{p+2} may be generated from $\Upsilon_{o+2,p,B}$. Iterating this bootstrapping technique for higher arity, we show that for any m, the full adversary A^m may be generated from $\Upsilon_{m,p,B}$. $\square$

Corollary 58. *Given $p \geq 1$, a structure $\mathcal{A}$ that is B-conservative, we may decide whether $\mathcal{A}$ is p-collapsible from source B.*

0-collapsibility (*proofs were omitted fully from paper*)

Theorem 47. Let $\mathcal{B}$ be a finite structure. The following are equivalent.

(i) $\mathcal{B}$ is 0-collapsible from source $\{x\}$ for some x in B.

(ii) $\mathcal{B}$ admits a simple A-she.

(iii) $\mathcal{B}$ is 0-collapsible for sentences of positive equality free first-order logic from source $\{x\}$ for some x in B.

Proof. The last two points are equivalent [17, Theorem 8] (this result is stated with A-she rather than simple A-she but clearly, $\mathcal{A}$ has an A-she iff it has a simple A-she). The implication (ii) to (i) follows trivially.

We prove the implication (i) to (ii) by contraposition. Assume that $A = [n] = \{1, \dots, n\}$ and suppose that $\mathcal{A}$ has no simple A-she. We will prove that $\mathcal{A}$ does not admit universal relativisation to x for pH-sentences. We assume also *w.l.o.g.* that $x = 1$. Let Ξ be the set of simple A-shops ξ s.t. $\xi(1) = [n]$. Since each ξ is not a she of $\mathcal{A}$, we have a quantifier-free formula with $2n-1$ variables R_ξ that consists of a single positive atom (not all variables need appear explicitly in this atom) such that $\mathcal{A} \models R_\xi(1,\ldots,1,2,\ldots,n)$[7], but $\mathcal{A} \not\models R_\xi(\xi^1,\ldots,\xi^n,\xi(2),\ldots,\xi(n))$ for some $\xi^1,\ldots,\xi^n \in [n] = \xi(1)$.

This means that for each $\eta : \{2,\ldots,n\} \to [n]$ there is some $2n-1$-ary "atom" R_η such that $\mathcal{A} \models R_\eta(1,\ldots,1,1,2,\ldots,n)$[8], but $\mathcal{A} \not\models R_\eta(\xi^1,\ldots,\xi^n,\eta(2),\ldots,\eta(n))$ for some $\xi^1,\ldots,\xi^n \in [n]$. Let $\mathrm{E} = [n]^{[n-1]}$ denotes the set of ηs.

Suppose we had universal relativisation to 1. Then we know that

$$\mathcal{A} \models \bigwedge_{\eta\in\mathrm{E}} R_\eta(1,\ldots,1,1,2,\ldots,n),$$

that is,

$$\mathcal{A} \models \exists y_1,\ldots,y_n \bigwedge_{\eta\in\mathrm{E}} R_\eta(1,\ldots,1,y_1,y_2,\ldots,y_n).$$

According to relativisation this means also that

$$\mathcal{A} \models \exists y_1,\ldots,y_n \forall x_1,\ldots,x_n \bigwedge_{\eta\in\mathrm{E}} R_\eta(x_1,\ldots,x_n,y_1,y_2,\ldots,y_n).$$

But we know

$$\mathcal{A} \models \forall y_1,\ldots,y_n \exists x_1,\ldots,x_n \bigvee_{\eta\in\mathrm{E}} \neg R_\eta(x_1,\ldots,x_n,y_1,y_2,\ldots,y_n),$$

since the ηs range over all maps $[n]$ to $[n]$. Contradiction. □

Theorem 48. Let $\mathcal{B}$ be a structure. The following are equivalent.

(i) $\mathcal{B}$ is 0-collapsible from source C

(ii) $\mathcal{B}^{|C|}$ is 0-collapsible from some (any) singleton source x which is a (rainbow) $|C|$-tuple containing all elements of C.

Proof. Let $B = \{1,2,\ldots,b\}$.

- (downwards). Let x be $|B|$-tuple containing all elements of B, wlog $x = (1,2,\ldots,b)$. Let φ be a pH sentence. Assume that $\mathcal{A}^{|B|} \models \varphi_{\restriction(x,x,\ldots,x)}$. Equivalently, for any i in B, $\mathcal{A} \models \varphi_{\restriction(i,i,\ldots,i)}$. Thus, 0-collapsibility from source B implies that $\mathcal{A} \models \varphi$. Since A and its power satisfy the same pH-sentences[12, 18] we may conclude that $\mathcal{A}^{|B|} \models \varphi$.
- (upwards). Assume that for any i in B, $\mathcal{A} \models \varphi_{\restriction(i,i,\ldots,i)}$. Equivalently, $\mathcal{A}^{|B|} \models \varphi_{\restriction(x,x,\ldots,x)}$ where x is any $|B|$-tuple containing all elements of B. By assumption, $\mathcal{A}^{|B|} \models \varphi$ and we may conclude that $\mathcal{A} \models \varphi$.

□

[7] There are n ones.

[8] There are n ones.