SIAM J. COMPUT.
Vol. 37, No. 1, pp. 132–163

CONSTRAINT SATISFACTION, LOGIC AND FORBIDDEN PATTERNS*

FLORENT MADELAINE† AND IAIN A. STEWART†

This paper is dedicated to the memory of our friend and colleague Clemens Lautemann

Abstract. In the 1990s, Feder and Vardi attempted to find a large subclass of NP which exhibits a dichotomy, that is, where every problem in the subclass is either solvable in polynomial-time or NP-complete. Their studies resulted in a candidate class of problems, namely, those definable in the logic MMSNP. While it remains open as to whether MMSNP exhibits a dichotomy, for various reasons it remains a strong candidate. Feder and Vardi added to the significance of MMSNP by proving that, although MMSNP strictly contains CSP, the class of constraint satisfaction problems, MMSNP and CSP are computationally equivalent. We introduce here a new class of combinatorial problems, the class of forbidden patterns problems FPP, and characterize MMSNP as the finite unions of problems from FPP. We use our characterization to detail exactly those problems that are in MMSNP but not in CSP. Furthermore, given a problem in MMSNP, we are able to decide whether the problem is in CSP or not (this whole process is effective). If the problem is in CSP, then we can construct a template for this problem; otherwise, for any given candidate for the role of template, we can build a counterexample (again, this process is effective).

Key words. finite model theory, constraint satisfaction, existential monadic second-order logic

AMS subject classifications. 68Q19, 68Q15, 0v3C13

DOI. 10.1137/050634840

1. Introduction.

Descriptive complexity theory seeks to classify problems, i.e., classes of finite structures, as to whether they can be defined using formulae of some specific logic, in relation to their computational complexity. One of the seminal results in descriptive complexity is Fagin's theorem [10], which states that a problem can be defined in existential second-order logic if and only if it is in the complexity class NP (throughout we equate a logic with the class of problems definable by the sentences of that logic). In a relatively recent paper and based upon Fagin's characterization of NP, Feder and Vardi [15] attempted to find a large (syntactically defined) subclass of NP which exhibits a dichotomy, that is, where every problem in the subclass is either solvable in polynomial-time or NP-complete (recall Ladner's theorem [22, 26], which states that if P $\neq$ NP, then there is an infinite number of distinct polynomial-time equivalence classes in NP). What emerged from Feder and Vardi's consideration was a (candidate) class of problems called MMSNP, defined by imposing syntactic restrictions upon the existential fragment of second-order logic. Their focus on a fragment of existential second-order logic was so that they might apply tools and techniques of finite model theory to possibly obtain a dichotomy result.

The logic MMSNP is defined by insisting that formulae of the fragment SNP of existential second-order logic must in addition be monotone, be monadic, and not involve inequalities (full definitions follow later). Feder and Vardi considered the imposition of combinations of these three restrictions (monadic, monotone, and without

*Received by the editors June 30, 2005; accepted for publication (in revised form) December 27, 2006; published electronically April 10, 2007.
http://www.siam.org/journals/sicomp/37-1/63484.html

†Department of Computer Science, University of Durham, Science Labs, South Road, Durham DH1 3LE, UK (f.r.madelaine@durham.ac.uk, i.a.stewart@durham.ac.uk). Much of this work was undertaken while the first author was at the Université de la Réunion.

inequalities) and showed that under any combination, excepting the imposition of all three restrictions, the resulting logic does not have a dichotomy (assuming $P \neq NP$). They were unable to make any similar claim about the logic obtained by imposing all three restrictions. However, they proved that MMSNP properly contains CSP, the class of combinatorial problems known as *constraint satisfaction problems* and, further, that the two classes are closely related in a computational sense.

THEOREM 1 (Feder and Vardi [15]). *Every problem in CSP is definable by a sentence of MMSNP, and every problem definable by a sentence of MMSNP is computationally equivalent to a problem in CSP.*

(By "computationally equivalent" above we mean that the MMSNP problem can be reduced to the CSP problem by a randomized polynomial-time Turing reduction, and the CSP problem can be reduced to the MMSNP problem by a polynomial-time Karp reduction.)[1]

The class CSP of constraint satisfaction problems is of great importance in computer science and artificial intelligence and has strong ties with database theory, graph theory, and universal algebra (see, for instance, [7, 30, 18, 20, 21]). For example, it is well-known that constraint satisfaction problems can be modeled in terms of the existence of homomorphisms between structures [21], in that every constraint satisfaction problem can be realized as the class of structures for which there exists a homomorphism to some fixed template structure. The close relationship between CSP and MMSNP prompted Feder and Vardi [15] to make explicit their conjecture that every problem in CSP is either NP-complete or solvable in polynomial-time. There are numerous results supporting this conjecture. For example, Schaefer [30] proved that if the template structure corresponding to some constraint satisfaction problem has size 2, then the conjecture holds, with Bulatov [3] recently extending Schaefer's result to templates of size 3. Also, Hell and Nešetřil [18] proved that the conjecture holds for all constraint satisfaction problems involving undirected graphs. Various other related dichotomy results have recently been determined; see, for example, [4, 5, 6, 8, 9, 11, 12, 13, 14, 15, 27, 28].

It is with the "border" between CSP and MMSNP that we are concerned in this paper. Feder and Vardi exhibited specific problems in MMSNP that are not in CSP, with their proofs relying essentially on counting arguments (they did not examine in any detail the inclusion relationship between CSP and MMSNP as classes of problems). We gave more examples of such problems in [25] although our proofs were of a different nature; they involved the explicit construction of particular families of graphs. We attempt in this paper to generalize the constructions in [25] so that we might develop a method by which we can ascertain whether *any* problem definable in MMSNP is in CSP or not. To this end, we give a new combinatorial characterization of MMSNP as the class of finite unions of *forbidden patterns problems* (from the class FPP). We use our new combinatorial characterization to answer the following questions in the affirmative: "*Can we characterize exactly those problems that are in MMSNP but not in CSP*?"; "*given a problem in MMSNP, is it decidable whether it is in CSP or not*?"; and "*if a problem in MMSNP can be shown to be in CSP then can we construct a template witnessing its inclusion in CSP*?"

As we shall see, forbidden patterns problems are given by representations that involve a finite set of colored structures, and we introduce the key notion of a recoloring between representations. The notions of a representation and a recoloring somehow generalize the notion of a structure and a homomorphism. The concept of a recoloring,

[1] Gábor Kun has recently derandomized this computational equivalence.

together with two notions that were implicitly present in the proof of Theorem 1 (the notion of a template of a representation and of a Feder–Vardi transformation), allow us to derive for *any* forbidden patterns problem a normal representation. Given any normal representation, we are then able to decide (according to simple criteria) whether the corresponding problem is in CSP or not. If it is in CSP, then we show how to construct its template; if it is not, then we show how to construct a counterexample to any potential template. Finally, we extend these results about problems in FPP to answer the questions (about MMSNP) above.

We end this section with a brief word about MMSNP and our research direction. The logic MMSNP has recently been shown to be related to constraint satisfaction problems where the template is infinite. In particular, Bodirsky and Dalmau [2] have shown that any problem in MMSNP that is nontrivial and closed under disjoint unions can be realized as a constraint satisfaction problem with an ω-categorial template. As regards our interest in the differences between MMSNP and CSP, there are numerous decidability investigations into the relative expressibilities of different logics in the literature, and we highlight a selection of these investigations here. In [1], Benedikt and Segoufin extend the well-known result that on strings, it is decidable whether a monadic second-order problem (that is, a regular language) is definable in first-order logic, to trees. In [16], Gaifman et al. show that the problem of deciding whether a given Datalog program is equivalent to one without recursion (and therefore to a formula of existential positive first-order logic) is undecidable. Finally, one very recent (and pertinent) result is that the problem of deciding whether a constraint satisfaction problem is first-order definable is decidable; indeed, it is NP-complete [23]. It turns out that first-order definable constraint satisfaction problems are forbidden patterns problems with a single color (logically, they correspond to the first-order fragment of MMSNP). The dual question (that asks, given such a forbidden patterns problem, whether it is a constraint satisfaction problem or not) is directly related to a popular notion in structural combinatorics, namely, that of a duality pair. Duality pairs have been characterized by Tardif and Nešetřil [31].

This paper is organized as follows. In the next section, we formally define CSP and FPP. In section 3, we recall the definition of Feder and Vardi's logic MMSNP and show how it relates to the class of problems FPP. In section 4, we introduce normal representations and related notions. In section 5, we prove our main result, i.e., an exact characterization of problems in FPP as to whether they are in CSP or not, provided that they can be given by connected representations. Next, in section 6, we extend this result to the disconnected case (this requires us to generalize normal representations to what we call normal sets) and then extend our results from FPP to MMSNP. Finally, in section 7, we conclude with some closing remarks.

2. Preliminaries. In this section, we give precise definitions of many of the concepts involved in this paper. We define many well-known notions in a slightly nonstandard way as many of these notions are extended very soon to analogous ones for colored structures.

Structures. A *signature* is a finite set of relation symbols (with each relation symbol having some finite arity). Let σ denote some fixed signature. A *σ-structure* $\mathcal{A}$ consists of a nonempty set A, the *domain*, together with an interpretation $R^{\mathcal{A}} \subseteq A^m$, for every m-ary relation symbol R in σ. Throughout this paper, we only ever consider finite σ-structures. Hence, in the following we simply write "a structure" instead of "a finite σ-structure." We denote structures by $\mathcal{A}, \mathcal{B}, \mathcal{C}$, etc., and their respective domains by A, B, C, etc., or alternatively by $|\mathcal{A}|, |\mathcal{B}|, |\mathcal{C}|$, etc.

Let $\mathcal{A}$ be a structure. We denote tuples of elements by $\mathbf{s}, \mathbf{t}$, etc., and we write "let $\mathbf{t}$ in A" as an abbreviation for "let $\mathbf{t}$ be a tuple of elements of A." Let R be a relation symbol in σ. We feel free to specify only when it is relevant the precise length of a tuple, and when we write "$R^{\mathcal{A}}(\mathbf{t})$" this automatically implies that the tuple of elements $\mathbf{t}$ has the same length as the arity of the relation symbol R. We write "a tuple $R^{\mathcal{A}}(\mathbf{t})$" as an abbreviation for "a tuple of elements $\mathbf{t}$ in A such that $R^{\mathcal{A}}(\mathbf{t})$ holds." We always use R to refer to a relation symbol of σ unless otherwise stated.

Let $\mathcal{A}$ and $\mathcal{B}$ be two structures. A *homomorphism from $\mathcal{A}$ to $\mathcal{B}$* is a mapping $h : A \to B$ such that for any relation symbol R in σ and for any tuple $R^{\mathcal{A}}(\mathbf{t})$, we have that $R^{\mathcal{B}}(h(\mathbf{t}))$, where $h(\mathbf{t})$ denotes the tuple obtained from $\mathbf{t}$ by a componentwise application of h. To denote that h is a homomorphism from $\mathcal{A}$ to $\mathcal{B}$, we write $\mathcal{A}\xrightarrow{h}\mathcal{B}$. If, furthermore, h is onto (respectively, one-to-one), then h is an *epimorphism* (respectively, a *monomorphism*), and we write $\mathcal{A}\overset{h}{\twoheadrightarrow}\mathcal{B}$ (respectively, $\mathcal{A}\overset{h}{\hookrightarrow}\mathcal{B}$). If both $\mathcal{A}\overset{h}{\twoheadrightarrow}\mathcal{B}$ and $\mathcal{A}\overset{h}{\hookrightarrow}\mathcal{B}$, then we write $\mathcal{A}\overset{h}{\hookrightarrow\!\!\!\!\twoheadrightarrow}\mathcal{B}$. If $\mathcal{A}\overset{h}{\hookrightarrow\!\!\!\!\twoheadrightarrow}\mathcal{B}$ and $\mathcal{A}\overset{h^{-1}}{\longrightarrow}\mathcal{B}$, then h is an *isomorphism*, and we write $\mathcal{A} \approx \mathcal{B}$. If there exists a homomorphism (respectively, a monomorphism) of $\mathcal{A}$ to $\mathcal{B}$, then we write $\mathcal{A}\to\mathcal{B}$ (respectively, $\mathcal{A}\hookrightarrow\mathcal{B}$). When something does not hold, we use the same notation but place a / through the symbol. For example, we write $\mathcal{A} \not\twoheadrightarrow \mathcal{B}$ if it is not the case that $\mathcal{A}\to\mathcal{B}$.

If $\mathcal{A}\overset{h}{\hookrightarrow}\mathcal{B}$, then $\mathcal{A}$ is a *substructure* of $\mathcal{B}$, and if, furthermore, for any tuple $R^{\mathcal{B}}(h(\mathbf{t}))$, we have that $R^{\mathcal{A}}(\mathbf{t})$ holds, then $\mathcal{A}$ is an *induced substructure* of $\mathcal{B}$. If $\mathcal{A}\overset{h}{\twoheadrightarrow}\mathcal{B}$ and every tuple $R^{\mathcal{B}}(\mathbf{t}')$ is in the image of h (more formally, there exists a tuple $\mathbf{t}$ in $\mathcal{A}$ such that $h(\mathbf{t}) = \mathbf{t}'$ and $R^{\mathcal{A}}(\mathbf{t})$ holds), then $\mathcal{B}$ is an *homomorphic image* of $\mathcal{A}$. If $\mathcal{A}\xrightarrow{h}\mathcal{B}$, then the homomorphic image of $\mathcal{A}$ under h, which we denote by $h(\mathcal{A})$, is the substructure of $\mathcal{B}$ that consists only of those tuples $R^{\mathcal{B}}(\mathbf{t}')$ that are in the image of h.

A *retract* of a structure $\mathcal{B}$ is a structure $\mathcal{A}$ for which there are two homomorphisms $\mathcal{A}\overset{i}{\hookrightarrow}\mathcal{B}$ and $\mathcal{B}\overset{s}{\twoheadrightarrow}\mathcal{A}$ such that $s \circ i = \mathrm{id}_{\mathcal{A}}$ (where $\mathrm{id}_{\mathcal{A}}$ denotes the identity homomorphism on $\mathcal{A}$, so, in particular, if $\mathcal{A}$ is a retract of $\mathcal{B}$, then $\mathcal{A}$ is isomorphic to an induced substructure of $\mathcal{B}$). Moreover, $\mathcal{A}$ is a *proper retract* whenever $\mathcal{A} \not\approx \mathcal{B}$. If $\mathcal{B}$ does not have any proper retracts, then $\mathcal{B}$ is *automorphic* (we use the terminology of [17]). An automorphic retract of $\mathcal{B}$ is called a *core*. It is well known that a core is unique up to isomorphism (see [17] or [19]).

Let $\mathcal{A}$ be a structure, let s and t be in A, and let $n \geq 1$. A *path of length n in $\mathcal{A}$ joining s and t* consists of n tuples $R_1^{\mathcal{A}}(\mathbf{t}_1), R_2^{\mathcal{A}}(\mathbf{t}_2), \ldots, R_n^{\mathcal{A}}(\mathbf{t}_n)$ such that each R_i is a relation symbol in σ of arity at least two (these relation symbols need not be distinct nor need the tuples), s occurs in $\mathbf{t}_1$, t occurs in $\mathbf{t}_n$, and for every $1 \leq i < n$, the tuples of elements $\mathbf{t}_i$ and $\mathbf{t}_{i+1}$ have a common element. If a path joins two distinct elements s and t, then they are *connected*. A structure $\mathcal{A}$ is *connected* if and only if any two distinct elements are connected.

Let $\mathcal{B}$ and $\mathcal{C}$ be two substructures of $\mathcal{A}$ and let $x \in A$. If

- $B \cap C = \{x\}$;
- $B \cup C = A$;
- for every relation symbol R of σ that has arity at least two and for every tuple $R^{\mathcal{A}}(\mathbf{t})$, either $R^{\mathcal{B}}(\mathbf{t})$ or $R^{\mathcal{C}}(\mathbf{t})$ holds but not both;
- for every monadic symbol M and for every element y in B (respectively, C), $M(y)$ holds in $\mathcal{B}$ (respectively, $\mathcal{C}$) if and only if $M(y)$ holds in $\mathcal{A}$; and
- each substructure $\mathcal{B}$ and $\mathcal{C}$ has at least one tuple $R(\mathbf{t})$ (where R has arity at least two),

then we say that $\mathcal{A}$ admits a *decomposition with components $\mathcal{B}$ and $\mathcal{C}$ in the articu-*

lation point x, and we write $\mathcal{A} = \mathcal{B}\overset{x}{32}\ \mathcal{C}$. If $\mathcal{A}$ is connected and does not admit any decomposition, then $\mathcal{A}$ is *biconnected*.

Let $\mathcal{A}$ be a structure. A tuple $R^{\mathcal{A}}(\mathbf{t})$ is said to be *antireflexive* if and only if no element in $\mathbf{t}$ occurs more than once. A *cycle of size* 1 in $\mathcal{A}$ consists of one tuple $R^{\mathcal{A}}(\mathbf{t})$ that is not antireflexive. An element that occurs more than once in a cycle of size 1 is called an *articulation point* of the cycle. A *cycle of size* 2 in $\mathcal{A}$ consists of two antireflexive tuples $R_1^{\mathcal{A}}(\mathbf{t}_1)$ and $R_2^{\mathcal{A}}(\mathbf{t}_2)$, for which we have that if $R_1 = R_2$, then $\mathbf{t}_1$ and $\mathbf{t}_2$ differ and which have at least two distinct common elements, each of which is called an *articulation point* of the cycle. Let $n > 2$. A *cycle of size* n in $\mathcal{A}$ consists of n tuples $R_1^{\mathcal{A}}(\mathbf{t}_1), R_2^{\mathcal{A}}(\mathbf{t}_2), \ldots, R_n^{\mathcal{A}}(\mathbf{t}_n)$ such that:

- for every $1 \leq i \leq n$, the tuple $R_i^{\mathcal{A}}(\mathbf{t}_i)$ is antireflexive;
- for every $1 \leq i < j \leq n$, if $j = i + 1$ or ($i = 1$ and $j = n$), the tuples $\mathbf{t}_i$ and $\mathbf{t}_j$ have one, and only one, common element $a_{i,j}$; otherwise, they have none; and
- the elements $a_{i,j}$, each of which is called an *articulation point* of the cycle, are pairwise distinct.

Colored structures. Let $\mathcal{T}$ be a structure. A $\mathcal{T}$*-colored structure* is a pair $(\mathcal{A}, a)$, where $\mathcal{A}$ is a structure and $\mathcal{A}\xrightarrow{a}\mathcal{T}$. We call: $\mathcal{T}$ the *target* of $(\mathcal{A}, a)$; a the *coloring*; and $\mathcal{A}$ the *underlying structure*. Let $(\mathcal{A}, a)$ and $(\mathcal{B}, b)$ be two $\mathcal{T}$-colored structures. A $\mathcal{T}$*-colored homomorphism* of $(\mathcal{A}, a)$ to $(\mathcal{B}, b)$ is a homomorphism $\mathcal{A}\xrightarrow{h}\mathcal{B}$ such that $a = b \circ h$. All notions defined above extend to $\mathcal{T}$-colored structures, so that colorings are respected by morphisms. For example, a retract of a $\mathcal{T}$-colored structure $(\mathcal{B}, b)$ is a $\mathcal{T}$-colored structure $(\mathcal{A}, a)$ for which there are two homomorphisms $\mathcal{A}\overset{i}{\hookrightarrow}\mathcal{B}$ and $\mathcal{B}\overset{s}{\twoheadrightarrow}\mathcal{A}$ such that $s \circ i = \mathrm{id}_{\mathcal{A}}$, $b \circ i = a$ and $a \circ s = b$. We use the same terminology but add the prefix "$\mathcal{T}$-colored," e.g., as in "$\mathcal{T}$-colored retract," and we use the same notation, e.g., $(\mathcal{A}, a)\xrightarrow{h}(\mathcal{B}, b)$ for a $\mathcal{T}$-colored homomorphism from $(\mathcal{A}, a)$ to $(\mathcal{B}, b)$. However, for simplicity, we may drop the prefix $\mathcal{T}$-colored when it does not cause confusion. At times, we deal with different targets, and so to avoid confusion, we sometimes write the target as a superscript, e.g., as in $(\mathcal{A}, a^{\mathcal{T}})$. We often refer to the elements of $|\mathcal{T}| = T$ as colors. We shall use the following lemmas later on, but we include them here so that readers can familiarize themselves with colored structures.

Lemma 2. *Let $(\mathcal{A}, a^{\mathcal{T}})$ be a $\mathcal{T}$-colored structure, let $\mathcal{T}'$ be a structure such that $\mathcal{T}'\overset{e}{\hookrightarrow}\mathcal{T}$, and let $(\mathcal{A}, a^{\mathcal{T}'})$ be a $\mathcal{T}'$-colored structure, where $a^{\mathcal{T}} = e \circ a^{\mathcal{T}'}$. If $(\mathcal{A}, a^{\mathcal{T}})$ is automorphic, then $(\mathcal{A}, a^{\mathcal{T}'})$ is automorphic.*

Proof. Suppose that $(\mathcal{A}, a^{\mathcal{T}})$ is automorphic, and suppose that $(\mathcal{B}, b^{\mathcal{T}'})$ is a proper retract of $(\mathcal{A}, a^{\mathcal{T}'})$. That is, we have that $(\mathcal{B}, b^{\mathcal{T}'})\overset{i}{\hookrightarrow}(\mathcal{A}, a^{\mathcal{T}'})$ and $(\mathcal{A}, a^{\mathcal{T}'})\overset{s}{\twoheadrightarrow}(\mathcal{B}, b^{\mathcal{T}'})$, where $s \circ i = \mathrm{id}_{\mathcal{B}}$, $a^{\mathcal{T}'} \circ i = b^{\mathcal{T}'}$, and $b^{\mathcal{T}'} \circ s = a^{\mathcal{T}'}$ (cf. the left commutative diagram of Figure 1) so that $(\mathcal{A}, a^{\mathcal{T}'}) \not\approx (\mathcal{B}, b^{\mathcal{T}'})$. We can compose the two $\mathcal{T}'$-colorings

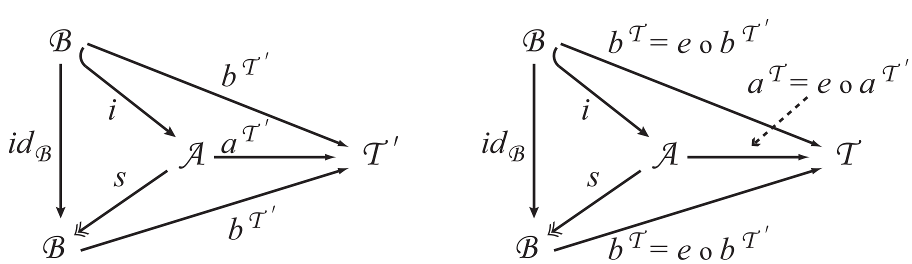

Fig. 1. *Proof of Lemma* 2.

with e to yield $\mathcal{T}$-colored structures, i.e., let $a^{\mathcal{T}} := e \circ a^{\mathcal{T}'}$ and $b^{\mathcal{T}} := e \circ b^{\mathcal{T}'}$ (cf. the right commutative diagram of Figure 1). Thus, $(\mathcal{B}, b^{\mathcal{T}})$ is a retract of $(\mathcal{A}, a^{\mathcal{T}})$, and so is isomorphic to $(\mathcal{A}, a^{\mathcal{T}})$. Thus, i is an isomorphism, with inverse s. Consequently, $(\mathcal{A}, a^{\mathcal{T}'}) \approx (\mathcal{B}, b^{\mathcal{T}'})$. But $(\mathcal{B}, b^{\mathcal{T}'})$ is a proper retract of $(\mathcal{A}, a^{\mathcal{T}'})$ which yields a contradiction. The result follows. □

The proofs of the next two lemmas are almost identical to analogous proofs in [19], for example, but are included here to allow readers to familiarize themselves with colored structures.

LEMMA 3. *The $\mathcal{T}$-colored structure $(\mathcal{A}, a^{\mathcal{T}})$ is automorphic if and only if whenever $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}})$, we have that $f(\mathcal{A}, a^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$.*

Proof. Assume that $(\mathcal{A}, a^{\mathcal{T}})$ is automorphic and also that $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}})$. From all such homomorphisms, choose g such that $g(\mathcal{A}, a^{\mathcal{T}})$ has a minimal number of elements and from those structures also a minimal number of tuples. Define h to be g restricted to $g(\mathcal{A}, a^{\mathcal{T}})$.

Note that $g(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{h} g(\mathcal{A}, a^{\mathcal{T}})$, and so h is one-to-one and onto, as otherwise $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{h \circ g} (\mathcal{A}, a^{\mathcal{T}})$ contradicts the minimality of g. So, h is an isomorphism. Thus, $(\mathcal{A}, a^{\mathcal{T}}) \overset{h^{-1} \circ g}{\twoheadrightarrow} g(\mathcal{A}, a^{\mathcal{T}})$ and $g(\mathcal{A}, a^{\mathcal{T}}) \overset{i}{\hookrightarrow} (\mathcal{A}, a^{\mathcal{T}})$, where i is the identity on $g(\mathcal{A}, a^{\mathcal{T}})$. For any $x \in |g(\mathcal{A}, a^{\mathcal{T}})|$, $h^{-1} \circ g \circ i(x) = h^{-1} \circ g(x) = h^{-1} \circ h(x) = x$. Hence, $g(\mathcal{A}, a^{\mathcal{T}})$ is a retract of $(\mathcal{A}, a^{\mathcal{T}})$, and so $g(\mathcal{A}, a^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$. Consequently, $f(\mathcal{A}, a^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$ by minimality of g.

Conversely, assume that whenever $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}})$, we have $f(\mathcal{A}, a^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$. Suppose that $(\mathcal{B}, b^{\mathcal{T}}) \overset{i}{\hookrightarrow} (\mathcal{A}, a^{\mathcal{T}})$ and $(\mathcal{A}, a^{\mathcal{T}}) \overset{s}{\twoheadrightarrow} (\mathcal{B}, b^{\mathcal{T}})$, with $s \circ i = id_{\mathcal{B}}$. Define $f := i \circ s$. Thus, $f(\mathcal{A}, a^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$, with i an epimorphism and s a monomorphism. Consequently, $(\mathcal{B}, b^{\mathcal{T}}) \approx (\mathcal{A}, a^{\mathcal{T}})$, and $(\mathcal{A}, a^{\mathcal{T}})$ is automorphic. □

LEMMA 4. *Every $\mathcal{T}$-colored structure has a $\mathcal{T}$-colored core that is unique up to $\mathcal{T}$-colored isomorphism.*

Proof. Trivially, every $\mathcal{T}$-colored structure has a $\mathcal{T}$-colored core. Suppose that $(\mathcal{A}_1, a_1)$ and $(\mathcal{A}_2, a_2)$ are cores of $(\mathcal{B}, b)$ such that $(\mathcal{A}_1, a_1) \not\approx (\mathcal{A}_2, a_2)$. In particular:

- $(\mathcal{A}_1, a_1) \overset{i_1}{\hookrightarrow} (\mathcal{B}, b)$ and $(\mathcal{B}, b) \overset{s_1}{\twoheadrightarrow} (\mathcal{A}_1, a_1)$ such that $s_1 \circ i_1 = \mathrm{id}_{\mathcal{A}_1}$, $b \circ i_1 = a_1$, and $s_1 \circ a_1 = b$; and
- $(\mathcal{A}_2, a_2) \overset{i_2}{\hookrightarrow} (\mathcal{B}, b)$ and $(\mathcal{B}, b) \overset{s_2}{\twoheadrightarrow} (\mathcal{A}_2, a_2)$ such that $s_2 \circ i_2 = \mathrm{id}_{\mathcal{A}_2}$, $b \circ i_2 = a_2$, and $s_2 \circ a_2 = b$.

Then $f_1 := s_2 \circ i_1 : (\mathcal{A}_1, a_1^{\mathcal{T}}) \to (\mathcal{A}_2, a_2^{\mathcal{T}})$ is a homomorphism as is $f_2 := s_1 \circ i_2 : (\mathcal{A}_2, a_2^{\mathcal{T}}) \to (\mathcal{A}_1, a_1^{\mathcal{T}})$. Hence, by Lemma 3, $f_2 \circ f_1(\mathcal{A}_1, a_1^{\mathcal{T}}) \approx (\mathcal{A}_1, a_1^{\mathcal{T}})$ and $f_1 \circ f_2(\mathcal{A}_2, a_2^{\mathcal{T}}) \approx (\mathcal{A}_2, a_2^{\mathcal{T}})$. Consequently, $(\mathcal{A}_1, a_1^{\mathcal{T}})$ and $(\mathcal{A}_2, a_2^{\mathcal{T}})$ are isomorphic, and the result follows. □

Patterns and representations. A structure $(\mathcal{A}, a^{\mathcal{T}})$ is a *$\mathcal{T}$-pattern* whenever for every $y \in A$, there exists a relation symbol R in σ and a tuple $\mathbf{t}$ in A in which y occurs such that $R^{\mathcal{A}}(\mathbf{t})$ holds (that is, every element occurs in some tuple in some relation of $\mathcal{A}$; i.e., $\mathcal{A}$ has no isolated elements). A $\mathcal{T}$-pattern $(\mathcal{A}, a^{\mathcal{T}})$ is *conform* if and only if $\mathcal{A}$ consists solely of an antireflexive tuple $R^{\mathcal{A}}(\mathbf{t})$: That is, there exists a relation symbol R in σ such that $R^{\mathcal{A}} = \{\mathbf{t}\}$, where every element of A occurs in $\mathbf{t}$ exactly once, and for every other relation symbol R' in σ, we have $R'^{\mathcal{A}} = \emptyset$. We denote conform patterns explicitly as in $(R(\mathbf{t}), a^{\mathcal{T}})$.

A *representation* is a pair $(\mathscr{F}, \mathcal{T})$, where $\mathcal{T}$ is a structure, called the *target*, and $\mathscr{F}$ is a finite set of $\mathcal{T}$-patterns, called the *forbidden patterns*. If every forbidden pattern in $\mathscr{F}$ is connected, then we say that $(\mathscr{F}, \mathcal{T})$ is *connected*. Let $(\mathscr{F}, \mathcal{T})$ be a representation. A $\mathcal{T}$-colored structure $(\mathcal{A}, a^{\mathcal{T}})$ is *valid* (respectively, *weakly valid*) with respect to $(\mathscr{F}, \mathcal{T})$ if and only if there is no forbidden pattern $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$ such that

$(\mathcal{B}, b^{\mathcal{T}}) \to (\mathcal{A}, a^{\mathcal{T}})$ (respectively, $(\mathcal{B}, b^{\mathcal{T}}) \hookrightarrow (\mathcal{A}, a^{\mathcal{T}})$). A structure $\mathcal{A}$ is *valid* (respectively, *weakly valid*) with respect to $(\mathscr{F}, \mathcal{T})$ if and only if there exists a homomorphism $\mathcal{A} \overset{a^{\mathcal{T}}}{\to} \mathcal{T}$ such that $(\mathcal{A}, a^{\mathcal{T}})$ is valid (respectively, weakly valid) with respect to $(\mathscr{F}, \mathcal{T})$.

Constraint satisfaction problems. It is well-known that constraint satisfaction problems can be modeled in terms of the existence of homomorphisms between structures [21]. Recall that the *nonuniform constraint satisfaction problem with template* $\mathcal{T}$, denoted by CSP($\mathcal{T}$), is the problem defined as follows:

- instances: structures $\mathcal{A}$ (over the same signature as $\mathcal{T}$);
- yes instances: those instances $\mathcal{A}$ for which $\mathcal{A} \to \mathcal{T}$.

We denote by CSP the class of nonuniform constraint satisfaction problems. Note that in [21], the adjective "nonuniform" was coined to distinguish such problems from *uniform* constraint satisfaction problems where the template $\mathcal{T}$ is not fixed but may range over a class of structures (all structures in general) and is part of the input. Since we do not deal with uniform problems in this paper, from now on we drop the phrase nonuniform.

Forbidden patterns problems. The *forbidden patterns problem* given by the representation $(\mathscr{F}, \mathcal{T})$, and denoted by FPP($\mathscr{F}, \mathcal{T}$), is the problem defined as follows:

- instances: structures $\mathcal{A}$ (over the same signature as $\mathcal{T}$);
- yes instances: those instances $\mathcal{A}$ that are valid w.r.t. $(\mathscr{F}, \mathcal{T})$.

We denote by FPP the class of forbidden patterns problems. If two representations define the same forbidden patterns problem, then we say that the representations are *equivalent*.

Remark 5. A problem in CSP is clearly monotone, i.e., closed under substructures. Furthermore, it is closed under inverse homomorphisms. To see this, let $\mathcal{B}$ and $\mathcal{T}$ be two structures. If $\mathcal{B} \in$ CSP($\mathcal{T}$), then $\mathcal{A} \in$ CSP($\mathcal{T}$) for any $\mathcal{A}$ such that $\mathcal{A} \to \mathcal{B}$. It is not difficult to check that if $\mathcal{B} \in$ FPP($\mathscr{F}, \mathcal{T}$), then $\mathcal{A} \in$ FPP($\mathscr{F}, \mathcal{T}$) for any $\mathcal{A}$ such that $\mathcal{A} \to \mathcal{B}$. Moreover, note that the containment problem, i.e., given two structures $\mathcal{T}$ and $\mathcal{T}'$, decide whether CSP($\mathcal{T}$) $\subseteq$ CSP($\mathcal{T}'$), is nothing other than the uniform constraint satisfaction problem (as CSP($\mathcal{T}$) $\subseteq$ CSP($\mathcal{T}'$) if and only if $\mathcal{T} \to \mathcal{T}'$).

THEOREM 6. CSP $\subsetneq$ FPP.

Proof. The inclusion is clear, as a problem from CSP with template $\mathcal{T}$ can be given equivalently as the forbidden patterns problem with representation $(\emptyset, \mathcal{T})$. It follows from counterexamples given in [15, 25] that this inclusion is strict. □

This provokes the following question, which is intrinsic to this paper: *When is a forbidden patterns problem not a constraint satisfaction problem*?

3. Feder and Vardi's logic. The logic SNP is the fragment of existential second-order logic, ESO, consisting of formulae Φ of the form $\exists \mathbf{S} \forall \mathbf{t} \varphi$, where $\mathbf{S}$ is a tuple of relation symbols (not in σ), $\mathbf{t}$ is a tuple of (first-order) variables, and φ is quantifier-free. Furthermore: Φ is in *monadic* SNP whenever $\mathbf{S}$ is a tuple of monadic relation symbols; Φ is in *monotone* SNP whenever every occurrence in φ of a symbol R from σ appears in the scope of an odd number of $\neg$ symbols; and Φ is in SNP *without inequalities* whenever the symbol $=$ does not appear in φ (either positively or negatively). If one thinks about the intuitive properties of the existence of a homomorphism from one structure to another, one might find it plausible to consider imposing some of the above restrictions on ESO. For instance, the existence (cf. the existential second-order quantifiers) of a homomorphism from an arbitrary source graph to a fixed target graph is equivalent to finding a partition of the domain of the source graph into sets (cf. the monadic restriction), one for each element of the target graph, so that every edge of the source graph (cf. the universal prefix of first-order quantifiers) maps to an edge of the target graph (cf. the monotone restriction above, reflecting that we are interested only in positive information, that is, about mappings of edges, not about mappings of "nonedges"). The "without inequalities" aspect of MMSNP comes about as homomorphisms do not distinguish between different elements.

Feder and Vardi considered the imposition of combinations of these three restrictions (monadic, monotone, and without inequalities) and showed that under any combination excepting the imposition of all three restrictions, the resulting logic does not have a dichotomy (assuming P $\neq$ NP). However, they were unable to make any similar claim about the logic obtained by imposing all three restrictions, and they observed that this logic subsumes CSP. This motivated the following definition.

DEFINITION 7. *Monotone monadic SNP without inequality (MMSNP) is the fragment of ESO consisting of those formulae* Φ *of the following form*:

$$\exists \mathbf{M} \forall \mathbf{t} \bigwedge_i \neg\big(\alpha_i(\sigma, \mathbf{t}) \wedge \beta_i(\mathbf{M}, \mathbf{t})\big),$$

where $\mathbf{M}$ *is a tuple of monadic relation symbols (not in* σ*),* $\mathbf{t}$ *is a tuple of (first-order) variables, and for every negated conjunct* $\neg(\alpha_i \wedge \beta_i)$:

- α_i *consists of a conjunction of positive atoms involving relation symbols from* σ *and variables from* $\mathbf{t}$; *and*
- β_i *consists of a conjunction of atoms or negated atoms involving relation symbols from* $\mathbf{M}$ *and variables from* $\mathbf{t}$.

(Notice that the equality symbol does not occur *in* Φ.*)*

Feder and Vardi showed that CSP is subsumed by MMSNP and, furthermore, that MMSNP is computationally equivalent to CSP. (Theorem 8 is a more detailed reformulation of Theorem 1 and is included for completeness.)

THEOREM 8 (Feder and Vardi [15]). *Every problem in CSP is definable by a sentence of MMSNP, but there are problems in MMSNP that are not in CSP. However, for every problem* $\Omega \in$ *MMSNP, there exists a problem* $\Omega' \in$ *CSP such that* Ω *reduces to* Ω' *via a polynomial-time Karp reduction, and* Ω' *reduces to* Ω *via a randomized polynomial-time Turing reduction.*[2]

(A more detailed proof of Theorem 8 than that in [15] can be found in [24].)

In the remainder of this section, we show that the logic MMSNP essentially corresponds to the class FPP of forbidden patterns problems. Let us begin by looking at some illustrative examples.

Example 9. Consider the signature $\sigma_2 = \langle E \rangle$, where E is a binary relation symbol. Define Φ_1 as

$$\exists C\, \forall x\, \forall y\, \forall z\, \Big(\neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x) \wedge C(x) \wedge C(y) \wedge C(z)\big) \\ \wedge \neg\big(E(x,y) \wedge E(y,z) \wedge E(z,x) \wedge \neg C(x) \wedge \neg C(y) \wedge \neg C(z)\big)\Big).$$

We can easily ascertain that Φ_1 defines the forbidden patterns problem with representation $(\mathscr{F}, \mathcal{T})$, where $|\mathcal{T}| := \{0,1\}$, $E^{\mathcal{T}} := |\mathcal{T}|^2$, and $\mathscr{F}$ contains two forbidden patterns, one for each negated conjunct, both having as the underlying structure a directed triangle (domain $\{x,y,z\}$ and relation $E = \{(x,y),(y,z),(z,x)\}$): In the first forbidden pattern all vertices of this directed triangle are colored 0, whereas in the second forbidden pattern the vertices are all colored 1 (the colorings are given by C and correspond to $x,y,z \mapsto 0$ and $x,y,z \mapsto 1$, respectively, and the colors are the

[2] As mentioned earlier, Gábor Kun has recently derandomized this reduction.

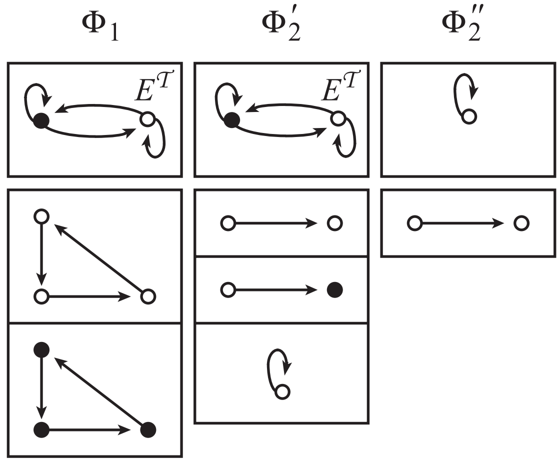

FIG. 2. *Primitive sentence and representations.*

names of the elements of the template). For simplicity, from now on we usually give representations in a pictorial fashion. For example, the representation we have just defined is depicted on the left in Figure 2; the top cell depicts the template, and the other cells depict the forbidden patterns. Note that the template is not a colored structure; however, to depict the homomorphisms from the forbidden patterns to the template, we have colored the elements of the template accordingly.

It is not so clear to which forbidden patterns problem the following sentence corresponds:

$$\Phi_2 := \exists C\ \forall x\ \forall y\ \big(\neg\big(E(x,y) \wedge C(x)\big) \wedge \neg\big(E(x,x) \wedge C(x) \wedge C(y)\big)\big).$$

However, it can be transformed into equivalent sentences as follows. First, we list all possibilities for the monadic predicate, to ensure that we have "fully colored" structures:

$$\exists C\ \forall x\ \forall y\ \big(\neg\big(E(x,y) \wedge C(x) \wedge C(y)\big) \wedge \neg\big(E(x,y) \wedge C(x) \wedge \neg C(y)\big) \wedge \neg\big(E(x,x) \wedge C(x) \wedge \neg C(y)\big)\big).$$

The last negated conjunct is comprised of two "independent" parts, namely, $(E(x,x) \wedge C(x))$ and $C(y)$, and does not correspond to a pattern (y does not appear in any atomic σ-relation). We can rewrite the above formula as the disjunction of two formulae Φ_2' and Φ_2'', where

$$\Phi_2' = \exists C\ \forall x\ \forall y\ \big(\neg\big(E(x,y) \wedge C(x) \wedge C(y)\big) \wedge \neg\big(E(x,y) \wedge C(x) \wedge \neg C(y)\big) \wedge \neg\big(E(x,x) \wedge C(x)\big)\big)$$

and

$$\Phi_2'' = \exists C\ \forall x\ \forall y\ \big(\neg\big(E(x,y) \wedge C(x) \wedge C(y)\big) \wedge \neg\big(E(x,y) \wedge C(x) \wedge \neg C(y)\big) \wedge \neg\big(\neg C(y)\big)\big)$$

(we leave the fact that Φ_2 can be so decomposed as a simple exercise). Now from each formula we can extract a suitable representation: This is easy in the case of Φ_2'; and, in the case of Φ_2'', note that the last negated conjunct essentially forces us to use a single color, so we can ignore all negated conjuncts which mention $\neg C(z)$ for some variable z. Finally, this leads to the representations depicted in the middle and on the right in Figure 2, respectively.

The above examples motivate the following definition and proposition.

DEFINITION 10. *A sentence Φ of MMSNP, where Φ is as in Definition 7, is* primitive *if and only if for every negated conjunct $\neg(\alpha \wedge \beta)$:*

- *for every first-order variable x that occurs in $\neg(\alpha \wedge \beta)$ and for every monadic symbol C in $\mathbf{M}$, exactly one of $C(x)$ and $\neg C(x)$ occurs in β; and*
- *unless x is the only first-order variable that occurs in $\neg(\alpha \wedge \beta)$, an atom of the form $R(\mathbf{t})$, where x occurs in $\mathbf{t}$ and R is a relation symbol from σ, must occur in α.*

PROPOSITION 11. *Every sentence of MMSNP is logically equivalent to a finite disjunction of primitive sentences.*

Proof. Let Φ be a sentence of MMSNP that is not primitive. Assume that Φ does not satisfy the first property of Definition 10. Let $\neg\big(\alpha(\sigma, \mathbf{t}) \wedge \beta(\mathbf{M}, \mathbf{t})\big)$ be a negated conjunct in Φ where there exists a (first-order) variable x that occurs in this negated conjunct and a monadic symbol C in $\mathbf{M}$ such that neither $C(x)$ nor $\neg C(x)$ occurs in β. Replace $\neg(\alpha \wedge \beta)$ in Φ by the conjunction of two negated conjuncts:

$$\neg\big(\alpha \wedge \beta \wedge C(x)\big) \wedge \neg\big(\alpha \wedge \beta \wedge \neg C(x)\big).$$

This new formula belongs to MMSNP and is logically equivalent to Φ. We iterate this process until the sentence satisfies the first property of Definition 10. Let Φ' denote this new sentence.

It may be the case that the second property does not hold for Φ' because of a negated conjunct of the form $\neg\big(\alpha(\sigma, \mathbf{t}) \wedge \beta_0(\mathbf{M}, \mathbf{t}) \wedge \beta_1(\mathbf{M}, x)\big)$, where x does not occur in $\mathbf{t}$, where $\alpha(\sigma, \mathbf{t})$ may be empty, and where β_1 is the conjunction of all atoms and negated atoms of β involving symbols from $\mathbf{M}$ and the variable x (β_0 is a conjunction of the remaining atoms and negated atoms of β). Let $\Phi'' = \Phi'_1 \vee \Phi'_2$, where

- Φ'_1 is obtained from Φ' by replacing $\neg(\alpha \wedge (\beta_0 \wedge \beta_1))$ in Φ' by $\neg(\alpha \wedge \beta_0)$; and
- Φ'_2 is obtained from Φ' by replacing $\neg(\alpha \wedge (\beta_0 \wedge \beta_1))$ in Φ' by $\neg\beta_1$.

First, note that Φ'_1 and Φ'_2 are both in MMSNP. Second, it is not hard to check that Φ'' is logically equivalent to Φ'. We iterate this transformation until each sentence in the disjunction satisfies the second property of Definition 10. ∎

We are now ready to state exactly what the correspondence is between MMSNP and FPP.

THEOREM 12. *The class of problems captured by the primitive fragment of the logic MMSNP is exactly the class FPP of forbidden patterns problems.*

Proof. Let $\Phi = \exists \mathbf{M} \forall \mathbf{t} \varphi$ be a primitive sentence of MMSNP. We shall build a representation $(\mathscr{F}, \mathcal{T})$ from Φ. A conjunction $\chi(\mathbf{M}, x)$ of atoms and negated atoms involving only relation symbols from $\mathbf{M}$ and the sole first-order variable x, where for each relation symbol C in $\mathbf{M}$, exactly one of $C(x)$ or $\neg C(x)$ occurs, is referred to as an $\mathbf{M}$-color. So, associated with every negated conjunct $\neg(\alpha \wedge \beta)$ in Φ (more precisely, with β in every such negated conjunct) and every variable occurring in this negated conjunct is a unique $\mathbf{M}$-color; in fact, β can be written as the conjunction of these $\mathbf{M}$-colors. Construct the structure $\mathcal{T}$ from Φ as follows:

- Its domain T consists of all $\mathbf{M}$-colors $\chi(\mathbf{M}, x)$ that are not explicitly forbidden in Φ by some negated conjunct $\neg(\alpha \wedge \beta)$ of φ having the form $\neg\chi(\mathbf{M}, x)$, i.e., so that α is empty and β is the $\mathbf{M}$-color χ; and
- for every relation symbol R of arity m in σ, set $R^{\mathcal{T}} := T^m$.

Start with $\mathscr{F} := \emptyset$, and for every negated conjunct $\neg(\alpha \wedge \beta)$ in φ, add to $\mathscr{F}$ the structure $(\mathcal{A}_\alpha, a^{\mathcal{T}}_\beta)$, where

- $\mathcal{A}_\alpha$ is the structure defined as follows:
  - the domain consists of all first-order variables that occur in the negated conjunct $\neg(\alpha \wedge \beta)$; and
  - for every relation symbol R in σ, there is a tuple $R^{\mathcal{A}_\alpha}(\mathbf{t})$ if and only if the atom $R(\mathbf{t})$ appears in α;
- for every $x \in |\mathcal{A}_\alpha|$, set $a^{\mathcal{T}}_\beta(x) := \chi$, where χ is the $\mathbf{M}$-color of x in β.

(The fact that Φ is primitive makes these definitions well-defined.)

Let $\mathcal{B}$ be a structure such that $\mathcal{B} \models \Phi$. So, there exists an assignment $\Pi : \mathbf{M} \to 2^B$ (where 2^B denotes the power set of B) such that $\mathcal{B} \models \forall \mathbf{t} \varphi(\Pi(\mathbf{M}), \mathbf{t})$ (here, $\varphi(\Pi(\mathbf{M}), \mathbf{t})$ denotes the formula φ where every monadic predicate is instantiated as the subset of B given by the assignment Π). Since $\Phi = \exists \mathbf{M} \forall \mathbf{t} \varphi$ is primitive, the formula φ is of the form:

$$\neg\chi_1(\mathbf{M}, x) \wedge \neg\chi_2(\mathbf{M}, x) \wedge \cdots \wedge \neg\chi_k(\mathbf{M}, x) \wedge \psi(\sigma, \mathbf{M}, \mathbf{t}),$$

where $k \geq 0$, and for every $1 \leq i \leq k$, χ_i is an $\mathbf{M}$-color (with all such $\mathbf{M}$-colors distinct) and ψ is a conjunction of negated conjuncts that are not $\mathbf{M}$-colors.

The assignment Π induces a map $\pi^{\mathcal{T}}$ from B to the set T that sends an element $u \in B$ to χ, where χ is the unique $\mathbf{M}$-color for which $\chi(\Pi(\mathbf{M}), u)$ holds (note that $\chi \neq \chi_i$ for $i = 1, 2, \ldots, k$, as $\neg\chi_i(\Pi(\mathbf{M}), u)$ holds for all $u \in B$).

Let $\neg(\alpha \wedge \beta)$ be a negated conjunct of φ, where α is nonempty, and suppose that $(\mathcal{A}_\alpha, a^{\mathcal{T}}_\beta) \xrightarrow{h} (\mathcal{B}, \pi^{\mathcal{T}})$.

Let $R(x_1, x_2, \ldots, x_a)$ be an atom appearing in α. So, $R^{\mathcal{A}_\alpha}(x_1, x_2, \ldots, x_a)$ holds and consequently $R^{\mathcal{B}}(h(x_1), h(x_2), \ldots, h(x_a))$ holds. Thus, if $\mathbf{t}'$ is the tuple of variables appearing in $\neg(\alpha \wedge \beta)$, then $\alpha^{\mathcal{B}}(h(\mathbf{t}'))$ holds. Also, $\pi^{\mathcal{T}} \circ h = a^{\mathcal{T}}_\beta$ and so $\pi^{\mathcal{T}}(h(\mathbf{t}')) = a^{\mathcal{T}}_\beta(\mathbf{t}')$. That is, $\beta^{\mathcal{B}}(\Pi(\mathbf{M}), h(\mathbf{t}'))$ holds. Thus $(\alpha \wedge \beta)^{\mathcal{B}}(\Pi(\mathbf{M}), h(\mathbf{t}'))$ holds, which contradicts the fact that $\mathcal{B} \models \Phi$, witnessed by $\Pi(\mathbf{M})$. Hence, $\mathcal{B} \in \mathrm{FPP}(\mathscr{F}, \mathcal{T})$.

Conversely, suppose that $\mathcal{B} \in \mathrm{FPP}(\mathscr{F}, \mathcal{T})$, witnessed by the homomorphism $\mathcal{B} \xrightarrow{\pi^{\mathcal{T}}} \mathcal{T}$. Clearly, $\pi^{\mathcal{T}}$ gives rise to an assignment $\Pi : \mathbf{M} \to 2^B$, where $u \in \Pi(C)$ for some $C \in \mathbf{M}$ and $u \in B$, if and only if $C(y)$ appears in $\chi(y)$, where $\pi^{\mathcal{T}}(u) = \chi$. Assume that $\mathcal{B} \models \alpha(h(\mathbf{t}')) \wedge \beta(\Pi(\mathbf{M}), h(\mathbf{t}'))$ for some map $h : |\mathcal{A}_\alpha| \to |\mathcal{B}|$, where $\neg(\alpha \wedge \beta)$ is a negated conjunct of φ, and $\mathbf{t}'$ is the tuple of variables appearing in $\alpha \wedge \beta$.

If $R^{\mathcal{A}_\alpha}(x_1, x_2, \ldots, x_a)$ holds, then $R^{\mathcal{B}}(h(x_1), h(x_2), \ldots, h(x_a))$ holds. If β is of the form $\bigwedge_{i=1}^{d} \chi^i(x_i)$, where $\mathbf{t}' = (x_1, x_2, \ldots, x_d)$ and each χ^i is an $\mathbf{M}$-color, then $\chi^i(\Pi(\mathbf{M}), h(x_i))$ holds for each $i = 1, 2, \ldots, d$. However, by definition $a^{\mathcal{T}}_\beta(x_i) = \chi^i$, and so $\pi^{\mathcal{T}}(h(x_i)) = a^{\mathcal{T}}_\beta(x_i)$ for each $i = 1, 2, \ldots, d$. Hence, $(\mathcal{A}_\alpha, a^{\mathcal{T}}_\beta) \xrightarrow{h} (\mathcal{B}, \pi^{\mathcal{T}})$, which yields a contradiction. Thus, $\mathcal{B} \models \Phi$, witnessed by the assignment $\Pi(\mathbf{M})$, and the implication follows.

Conversely, given a representation $(\mathscr{F}, \mathcal{T})$, we shall build a corresponding primitive sentence of MMSNP. Let $\mathbf{M} = \{C_1, C_2, \ldots, C_k\}$ be a set of monadic predicates that are not in σ such that $k = \lceil \log_2 |\mathcal{T}| \rceil$. To each element x_i of $|\mathcal{T}|$, we associate some arbitrary $\mathbf{M}$-color χ_{x_i}. Let $\chi_{|\mathcal{T}|+1}, \ldots, \chi_{2^k}$ denote the remaining $\mathbf{M}$-colors (if $|\mathcal{T}| < 2^k$). Let $\Phi = \exists \mathbf{M} \forall \mathbf{t} \phi$, where $\forall \mathbf{t} \phi$ is the universal closure of the conjunction of the following negated conjuncts:

- If $|\mathcal{T}| < 2^k$, then for every i such that $|\mathcal{T}| < i \leq 2^k$, we add the negated conjunct $\neg\chi_i(y)$.

- For each tuple $R(i_1, i_2, \ldots, i_r)$ that does not hold in $\mathcal{T}$, we add the negated conjunct $\neg(R(y_1, y_2, \ldots, y_r) \wedge \chi_{i_1}(y_1) \wedge \chi_{i_2}(y_2) \wedge \cdots \wedge \chi_{i_r}(y_r))$, where the variables $y_1, y_2, \ldots, y_r$ are pairwise distinct.
- For each forbidden pattern $(\mathcal{A}, a^{\mathcal{T}})$ in $\mathscr{F}$, we add the negated conjunct $\neg(\alpha \wedge \beta)$, where α is the conjunction of the tuples of $\mathcal{F}$, and β is the conjunction $\bigwedge_{x \in |\mathcal{A}|} \chi_{a^{\mathcal{T}}(x)}(x)$.

The first type of negated conjunct ensures that we may use only the $\mathcal{M}$-colors that correspond to elements of $\mathcal{T}$. The second type of negated conjunct describes that there is a homomorphism to $\mathcal{T}$. Finally, the last type of negated conjunct enforces that this homomorphism is not compatible with any of the forbidden patterns. Consequently, a structure $\mathcal{B}$ is a yes instance of the forbidden patterns problem with representation $(\mathscr{F}, \mathcal{T})$ if and only if $\mathcal{B} \models \Phi$. The formal proof of this equivalence is similar to that of the first implication. This concludes the proof. ∎

By Proposition 11, every forbidden patterns problem is described by a primitive sentence of MMSNP. Since the disjunction of two sentences of MMSNP is logically equivalent to a sentence of MMSNP, we get the following corollary from the above theorem.

COROLLARY 13. *The class of problems captured by the logic MMSNP corresponds exactly to the class of finite unions of problems in FPP.*

4. A normal form for problems in FPP. In this section, we introduce normal representations and show how any representation can be effectively rewritten into an equivalent normal representation. The transformation is achieved through a combination of different operations so as to enforce various properties. We shall make clear later, in section 5, why we need these properties.

However, before we proceed, let us try and give some idea here of the direction of travel by stating the properties we wish to enforce and our intended goal. We shall state the properties again at the appropriate point in the text, as we do with the definition and result stated below. Let $(\mathscr{F}, \mathcal{T})$ be a representation. The properties we wish to enforce upon $(\mathscr{F}, \mathcal{T})$ are as follows.

($\mathfrak{p}1$) Any structure is valid if and only if it is weakly valid.
($\mathfrak{p}2$) Every pattern of $\mathscr{F}$ is automorphic.
($\mathfrak{p}3$) It is not the case that $(\mathcal{B}_1, b_1^{\mathcal{T}}) \hookrightarrow (\mathcal{B}_2, b_2^{\mathcal{T}})$ for any distinct patterns $(\mathcal{B}_1, b_1^{\mathcal{T}})$ and $(\mathcal{B}_2, b_2^{\mathcal{T}})$ of $\mathscr{F}$.
($\mathfrak{p}4$) No pattern of $\mathscr{F}$ is conform.
($\mathfrak{p}5$) Every forbidden pattern is biconnected.
($\mathfrak{p}6$) The representation $(\mathscr{F}, \mathcal{T})$ is automorphic.

We say that a connected representation for which properties $\mathfrak{p}1$ to $\mathfrak{p}6$ hold is a *normal representation*. In the process of reducing our representation to a normal representation, we will show that this can be done by an effective procedure.

4.1. Our first batch of reductions. Let $(\mathscr{F}, \mathcal{T})$ be a representation. We now define a number of operations on representations so that we might enforce certain properties. However, before we start, we wish our representation to have the following property:

($\mathfrak{p}1$) Any structure is valid if and only if it is weakly valid.

Let $\mathbf{H}\mathscr{F}$ be the set of homomorphic images of the patterns from $\mathscr{F}$, up to isomorphism. Recall that a forbidden pattern is a colored structure; hence, an homomorphic image of a forbidden pattern $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$ is a colored structure $(\mathcal{C}, c^{\mathcal{T}})$ such that there exists an epimorphism $\mathcal{B} \overset{h}{\twoheadrightarrow} \mathcal{C}$ with the properties that:

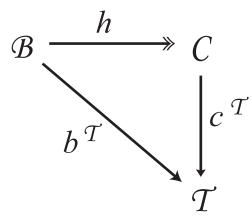

FIG. 3. *A commuting diagram.*

- for each symbol $R \in \sigma$ and for each tuple $R^{\mathcal{C}}(\tilde{\mathbf{t}})$, there exists a tuple $R^{\mathcal{B}}(\mathbf{t})$ such that $h(\mathbf{t}) = \tilde{\mathbf{t}}$ and
- the diagram in Figure 3 commutes.

LEMMA 14. *The representation* $(\mathbf{H}\mathscr{F}, \mathcal{T})$ *is equivalent to* $(\mathscr{F}, \mathcal{T})$.

Proof. Let $\mathcal{A}$ be valid w.r.t. $(\mathscr{F}, \mathcal{T})$, witnessed by $\mathcal{A} \xrightarrow{a^{\mathcal{T}}} \mathcal{T}$. Assume for contradiction that $(\mathcal{A}, a^{\mathcal{T}})$ is not valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$, and let $(\mathcal{C}, c^{\mathcal{T}}) \in \mathbf{H}\mathscr{F}$ (defined from $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$, using h as above) be such that $(\mathcal{C}, c^{\mathcal{T}}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}})$. By composition, it follows that $(\mathcal{B}, b^{\mathcal{T}}) \xrightarrow{f \circ h} (\mathcal{A}, a^{\mathcal{T}})$. This yields a contradiction, and so $(\mathcal{A}, a^{\mathcal{T}})$ is valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$.

Conversely, if $\mathcal{A}$ is valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$, then $\mathcal{A}$ is valid w.r.t. $(\mathscr{F}, \mathcal{T})$ since $\mathscr{F} \subseteq \mathbf{H}\mathscr{F}$. □

LEMMA 15. *The representation* $(\mathbf{H}\mathscr{F}, \mathcal{T})$ *satisfies* $\mathfrak{p}1$.

Proof. Let $(\mathcal{A}, a^{\mathcal{T}})$ be weakly valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$. Assume for contradiction that $(\mathcal{A}, a^{\mathcal{T}})$ is not valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$, and let $(\mathcal{C}, c^{\mathcal{T}}) \in \mathbf{H}\mathscr{F}$ (defined from $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$, using h as above) be such that $(\mathcal{C}, c^{\mathcal{T}}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}})$. By construction, $f(\mathcal{C}, c^{\mathcal{T}})$ belongs to $\mathbf{H}\mathscr{F}$, and $f(\mathcal{C}, c^{\mathcal{T}}) \hookrightarrow (\mathcal{A}, a^{\mathcal{T}})$. This yields a contradiction.

Conversely, if $(\mathcal{A}, a^{\mathcal{T}})$ is valid w.r.t. $(\mathbf{H}\mathscr{F}, \mathcal{T})$, then it is trivially weakly valid. The result follows. □

Our next property to enforce is the following:

($\mathfrak{p}2$) Every pattern of $\mathscr{F}$ is automorphic.

DEFINITION 16. *Let* $(\mathscr{F}, \mathcal{T})$ *be a representation, and let* $(\mathscr{F}', \mathcal{T})$ *be the representation obtained by replacing a pattern of* $\mathscr{F}$ *with its core. We call this a* core reduction *on* $(\mathscr{F}, \mathcal{T})$.

Note that Definition 16 is well-defined by Lemma 4.

LEMMA 17. *Let the representation* $(\mathscr{F}', \mathcal{T})$ *be obtained from the representation* $(\mathscr{F}, \mathcal{T})$ *by a core reduction.*

- $(\mathscr{F}', \mathcal{T})$ *is equivalent to* $(\mathscr{F}, \mathcal{T})$.
- *If* $(\mathscr{F}, \mathcal{T})$ *satisfies property* $\mathfrak{p}1$, *then so does* $(\mathscr{F}', \mathcal{T})$.

Proof. If $(\mathcal{C}, c^{\mathcal{T}})$ is the core of $(\mathcal{B}, b^{\mathcal{T}})$, then $(\mathcal{B}, b^{\mathcal{T}}) \to (\mathcal{A}, a^{\mathcal{T}})$ if and only if $(\mathcal{C}, c^{\mathcal{T}}) \to (\mathcal{A}, a^{\mathcal{T}})$ for any structure $(\mathcal{A}, a^{\mathcal{T}})$. Hence, $(\mathscr{F}', \mathcal{T})$ is equivalent to $(\mathscr{F}, \mathcal{T})$.

Assume that $(\mathscr{F}, \mathcal{T})$ satisfies property $\mathfrak{p}1$. Suppose that $(\mathcal{A}, a^{\mathcal{T}})$ is weakly valid w.r.t. $(\mathscr{F}', \mathcal{T})$. If $(\mathcal{C}, c^{\mathcal{T}}) \to (\mathcal{A}, a^{\mathcal{T}})$ for some $(\mathcal{C}, c^{\mathcal{T}}) \in \mathscr{F}'$, then we must have that $(\mathcal{B}, b^{\mathcal{T}}) \to (\mathcal{A}, a^{\mathcal{T}})$ for some $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$. As $(\mathscr{F}, \mathcal{T})$ satisfies property $\mathfrak{p}1$, $(\mathcal{D}, d^{\mathcal{T}}) \hookrightarrow (\mathcal{A}, a^{\mathcal{T}})$ for some $(\mathcal{D}, d^{\mathcal{T}}) \in \mathscr{F}$. If $(\mathcal{D}, d^{\mathcal{T}}) \in \mathscr{F}'$, then we obtain a contradiction; otherwise, the core of $(\mathcal{D}, d^{\mathcal{T}})$ is in $\mathscr{F}'$, and we still obtain that some forbidden pattern of $\mathscr{F}'$ embeds into $(\mathcal{A}, a^{\mathcal{T}})$, yielding a contradiction. Hence, $(\mathscr{F}', \mathcal{T})$ satisfies property $\mathfrak{p}1$. □

Our next property to enforce is the following:

($\mathfrak{p}3$) It is not the case that $(\mathcal{B}_1, b_1^{\mathcal{T}}) \hookrightarrow (\mathcal{B}_2, b_2^{\mathcal{T}})$ for any distinct patterns $(\mathcal{B}_1, b_1^{\mathcal{T}})$ and $(\mathcal{B}_2, b_2^{\mathcal{T}})$ of $\mathscr{F}$.

Definition 18. *Let $(\mathscr{F},\mathcal{T})$ be a representation, and let $(\mathcal{B}_1,b_1^{\mathcal{T}})$ and $(\mathcal{B}_2,b_2^{\mathcal{T}})$ be distinct patterns of $\mathscr{F}$ such that $(\mathcal{B}_1,b_1^{\mathcal{T}})\hookrightarrow(\mathcal{B}_2,b_2^{\mathcal{T}})$. Let $(\mathscr{F}',\mathcal{T})$ be the representation obtained by removing the pattern $(\mathcal{B}_2,b_2^{\mathcal{T}})$ from $\mathscr{F}$. We call this an* embed reduction *on $(\mathscr{F},\mathcal{T})$.*

Lemma 19. *Let the representation $(\mathscr{F}',\mathcal{T})$ be obtained from the representation $(\mathscr{F},\mathcal{T})$ by an embed reduction.*

- *$(\mathscr{F}',\mathcal{T})$ is equivalent to $(\mathscr{F},\mathcal{T})$.*
- *If $(\mathscr{F},\mathcal{T})$ satisfies property $\mathfrak{p1}$, then so does $(\mathscr{F}',\mathcal{T})$.*

Proof. Trivially, $\mathrm{FPP}(\mathscr{F},\mathcal{T})\subseteq\mathrm{FPP}(\mathscr{F}',\mathcal{T})$. If $(\mathcal{B}_2,b_2^{\mathcal{T}})\to(\mathcal{A},a^{\mathcal{T}})$ (with reference to Definition 18), then $(\mathcal{B}_1,b_1^{\mathcal{T}})\to(\mathcal{A},a^{\mathcal{T}})$ for any structure $(\mathcal{A},a^{\mathcal{T}})$, and so $\mathrm{FPP}(\mathscr{F}',\mathcal{T})\subseteq\mathrm{FPP}(\mathscr{F},\mathcal{T})$.

Assume that $(\mathscr{F},\mathcal{T})$ satisfies property $\mathfrak{p1}$. Suppose that $(\mathcal{A},a^{\mathcal{T}})$ is weakly valid w.r.t. $(\mathscr{F}',\mathcal{T})$. If $(\mathcal{B},b^{\mathcal{T}})\to(\mathcal{A},a^{\mathcal{T}})$ for some pattern $(\mathcal{B},b^{\mathcal{T}})\in\mathscr{F}'$, and so some pattern in $\mathscr{F}$, then we have that $(\mathcal{A},a^{\mathcal{T}})$ is not weakly valid w.r.t. $(\mathscr{F},\mathcal{T})$. That is, $(\mathcal{C},c^{\mathcal{T}})\hookrightarrow(\mathcal{A},a^{\mathcal{T}})$ for some pattern $(\mathcal{C},c^{\mathcal{T}})\in\mathscr{F}$. If $(\mathcal{C},c^{\mathcal{T}})\in\mathscr{F}'$, then we obtain a contradiction; otherwise, $(\mathcal{C},c^{\mathcal{T}})$ is the pattern removed by the embed reduction and $(\mathcal{D},d^{\mathcal{T}})\hookrightarrow(\mathcal{C},c^{\mathcal{T}})$ for some pattern $(\mathcal{D},d^{\mathcal{T}})\in\mathscr{F}'$. Thus, we still obtain a contradiction, and $(\mathscr{F}',\mathcal{T})$ satisfies $\mathfrak{p1}$. □

Our next property to enforce is the following:

($\mathfrak{p4}$) No pattern of $\mathscr{F}$ is conform.

Definition 20. *Let $(\mathscr{F},\mathcal{T})$ be a representation, and let $(R(\mathbf{t}),\pi^{\mathcal{T}})$ be a conform pattern of $\mathscr{F}$. Let $\mathcal{T}'$ be the structure obtained from $\mathcal{T}$ by the removal of the tuple $R(\pi^{\mathcal{T}}(\mathbf{t}))$, and let e be the monomorphism $\mathcal{T}'\overset{e}{\hookrightarrow}\mathcal{T}$ defined by inclusion. Let $\mathscr{F}'$ denote the set of patterns of $\mathscr{F}$ that are also $\mathcal{T}'$-patterns; that is, the patterns $(\mathcal{B},b^{\mathcal{T}})\in\mathscr{F}$ for which $b^{\mathcal{T}}(\mathbf{u})\neq\pi^{\mathcal{T}}(\mathbf{t})$ for any tuple $R^{\mathcal{B}}(\mathbf{u})$. The representation $(\mathscr{F}',\mathcal{T}')$ has been obtained from $(\mathscr{F},\mathcal{T})$ by a* conform reduction.

Lemma 21. *Let the representation $(\mathscr{F}',\mathcal{T}')$ be obtained from the representation $(\mathscr{F},\mathcal{T})$ by a conform reduction.*

- *$(\mathscr{F}',\mathcal{T}')$ is equivalent to $(\mathscr{F},\mathcal{T})$.*
- *If $(\mathscr{F},\mathcal{T})$ satisfies property $\mathfrak{p1}$, then so does $(\mathscr{F}',\mathcal{T}')$.*

Proof. We denote a pattern $(\mathcal{B},b^{\mathcal{T}})\in\mathscr{F}$ that is also a $\mathcal{T}'$-pattern by $(\mathcal{B},b^{\mathcal{T}'})$ also, where $b^{\mathcal{T}'}$ is the homomorphism $\mathcal{B}\overset{b^{\mathcal{T}'}}{\to}\mathcal{T}'$ obtained directly from $b^{\mathcal{T}}$; that is, $b^{\mathcal{T}}=e\circ b^{\mathcal{T}'}$.

Assume that $(\mathcal{A},a^{\mathcal{T}'})$ is valid w.r.t. $(\mathscr{F}',\mathcal{T}')$ and define $a^{\mathcal{T}}:=e\circ a^{\mathcal{T}'}$ (so $\mathcal{A}\overset{a^{\mathcal{T}}}{\to}\mathcal{T}$ and $a^{\mathcal{T}}(\mathbf{u})\neq\pi^{\mathcal{T}}(\mathbf{t})$ for any tuple $R^{\mathcal{A}}(\mathbf{u})$). Suppose that some pattern $(\mathcal{B},b^{\mathcal{T}})\in\mathscr{F}$ is such that $(\mathcal{B},b^{\mathcal{T}})\to(\mathcal{A},a^{\mathcal{T}})$. Thus, $(\mathcal{B},b^{\mathcal{T}})$ is actually a $\mathcal{T}'$-pattern, and $(\mathcal{B},b^{\mathcal{T}'})\to(\mathcal{A},a^{\mathcal{T}'})$, which yields a contradiction.

Conversely, suppose that $(\mathcal{A},a^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F},\mathcal{T})$. There are two cases: either the map $a^{\mathcal{T}}$ yields a homomorphism $\mathcal{A}\to\mathcal{T}'$, or it doesn't.

Suppose that the map $a^{\mathcal{T}}$ yields a homomorphism $\mathcal{A}\overset{a^{\mathcal{T}'}}{\to}\mathcal{T}'$; thus, $a^{\mathcal{T}}=e\circ a^{\mathcal{T}'}$. If $(\mathcal{B},b^{\mathcal{T}'})\in\mathscr{F}'$ is such that $(\mathcal{B},b^{\mathcal{T}'})\to(\mathcal{A},a^{\mathcal{T}'})$, then we have that $(\mathcal{B},b^{\mathcal{T}})\to(\mathcal{A},a^{\mathcal{T}})$ (where $b^{\mathcal{T}}=e\circ b^{\mathcal{T}'}$, recall), which yields a contradiction. Thus, $(\mathcal{A},a^{\mathcal{T}'})$ is valid w.r.t. $(\mathscr{F}',\mathcal{T}')$.

Suppose that the map $a^{\mathcal{T}}$ does not yield a homomorphism from $\mathcal{A}$ to $\mathcal{T}'$. There must exist some tuple $R^{\mathcal{A}}(\hat{\mathbf{t}})$ such that $a^{\mathcal{T}}(\hat{\mathbf{t}})=\pi^{\mathcal{T}}(\mathbf{t})$. Define $h:|R(\mathbf{t})|\to|\mathcal{A}|$ as the map which takes $\mathbf{t}$ to $\hat{\mathbf{t}}$ (note that this is well-defined as $\mathbf{t}$ is antireflexive). Consequently, $(R(\mathbf{t}),\pi^{\mathcal{T}})\overset{h}{\to}(\mathcal{A},a^{\mathcal{T}})$, which yields a contradiction as $(\mathcal{A},a^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F},\mathcal{T})$. Hence, $(\mathscr{F}',\mathcal{T}')$ is equivalent to $(\mathscr{F},\mathcal{T})$.

Assume that $(\mathscr{F},\mathcal{T})$ satisfies property $\mathfrak{p1}$. Suppose that $(\mathcal{A},a^{\mathcal{T}'})$ is weakly valid w.r.t. $(\mathscr{F}',\mathcal{T}')$ and that there exists a pattern $(\mathcal{B},b^{\mathcal{T}'})\in\mathscr{F}'$ such that $(\mathcal{B},b^{\mathcal{T}'})\to(\mathcal{A},a^{\mathcal{T}'})$.

FIG. 4. *Depiction of tuples.*

Define $b^{\mathcal{T}} := e \circ b^{\mathcal{T}'}$ and $a^{\mathcal{T}} := e \circ a^{\mathcal{T}'}$. Thus, $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$ and $(\mathcal{B}, b^{\mathcal{T}}) \rightarrow (\mathcal{A}, a^{\mathcal{T}})$. Hence, $(\mathcal{A}, a^{\mathcal{T}})$ is not weakly valid w.r.t. $(\mathscr{F}, \mathcal{T})$. That is, $(\mathcal{C}, c^{\mathcal{T}}) \hookrightarrow (\mathcal{A}, a^{\mathcal{T}})$ for some pattern $(\mathcal{C}, c^{\mathcal{T}}) \in \mathscr{F}$. But as $a^{\mathcal{T}} = e \circ a^{\mathcal{T}'}$, so $(\mathcal{C}, c^{\mathcal{T}})$ is also a $\mathcal{T}'$-pattern and so is in $\mathscr{F}'$. This yields a contradiction, and the result follows. ◻

Remark 22. Applying embed reductions clearly preserves property $\mathfrak{p}2$. Note also that applying conform reductions preserves property $\mathfrak{p}2$. This follows directly from Lemma 2.

Example 23. Consider a representation $(\mathscr{F}, \mathcal{T})$ over the signature consisting of a binary relation symbol E and a ternary relation symbol R. The domain of $\mathcal{T}$ consists of two elements (or colors) $\circ$ and $\bullet$, with $E^{\mathcal{T}} = \{\circ, \bullet\}^2$ and $R^{\mathcal{T}} = \{\circ, \bullet\}^3$.

Consider the conform forbidden pattern consisting of the single tuple $R(x, y, z)$, where both x and y take the color $\circ$ and z takes the color $\bullet$. We depict this pattern by the left diagram in Figure 4. In the case where $x = y$, we depict the pattern by the right diagram in Figure 4.

The first (leftmost) column in Figure 5 depicts the four forbidden patterns in $\mathscr{F}$ (the top three are such that $R = \emptyset$, and the bottom is such that $E = \emptyset$). The second column depicts the representation $(\mathbf{H}\mathscr{F}, \mathcal{T})$, formed by adding all homomorphic images of the forbidden patterns in $\mathscr{F}$ (up to isomorphism). In the third column, we have performed core and embed reductions to obtain an equivalent representation satisfying properties $\mathfrak{p}1$, $\mathfrak{p}2$, and $\mathfrak{p}3$. In the fourth column, we have performed conform reductions to obtain an equivalent representation satisfying properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, and $\mathfrak{p}4$.

Note that, in general, starting from a representation satisfying property $\mathfrak{p}1$, if we apply core, embed, and conform reductions arbitrarily, then after a finite sequence of reductions, by the lemmas of this subsection, we will obtain an equivalent representation satisfying properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, and $\mathfrak{p}4$ (a simple induction suffices).

4.2. Feder–Vardi reductions. The reductions introduced so far do not suffice for us to obtain the normal form for which we are aiming. We need to interleave applications of these reductions with another reduction that we define now.

From now on, we make an important assumption regarding the representations we deal with: Until otherwise specified, we assume them to be connected (we shall deal with the disconnected case in section 6.1.1).

We say that a pattern is *biconnected* if its underlying structure is biconnected. Our aim is to enforce the following property (using techniques inspired from the proof of Theorem 8 in [15]):

($\mathfrak{p}5$) Every forbidden pattern is biconnected.

Before proceeding, we need some definitions relating to the grouping together of forbidden patterns. A *compact $\mathcal{T}$-structure* $\{\mathcal{A}, \overline{\alpha}\}$ is a structure $\mathcal{A}$ together with a map $\overline{\alpha} : A \rightarrow 2^T$ so that every map $a^{\mathcal{T}} : A \rightarrow T$ with the property that $a^{\mathcal{T}}(y) \in \overline{\alpha}(y)$, for all $y \in A$, yields a $\mathcal{T}$-colored structure $(\mathcal{A}, a^{\mathcal{T}})$. This notion is only a useful shorthand to denote a set of colored structures, as a compact structure can be *expanded* to obtain a set of colored structures, each with the same underlying structure; but we shall need this notion later on when we prove the termination of a particular sequence of transformations we employ towards the end of this section (this notion was not necessary in Feder and Vardi's original proof as the negated conjuncts correspond

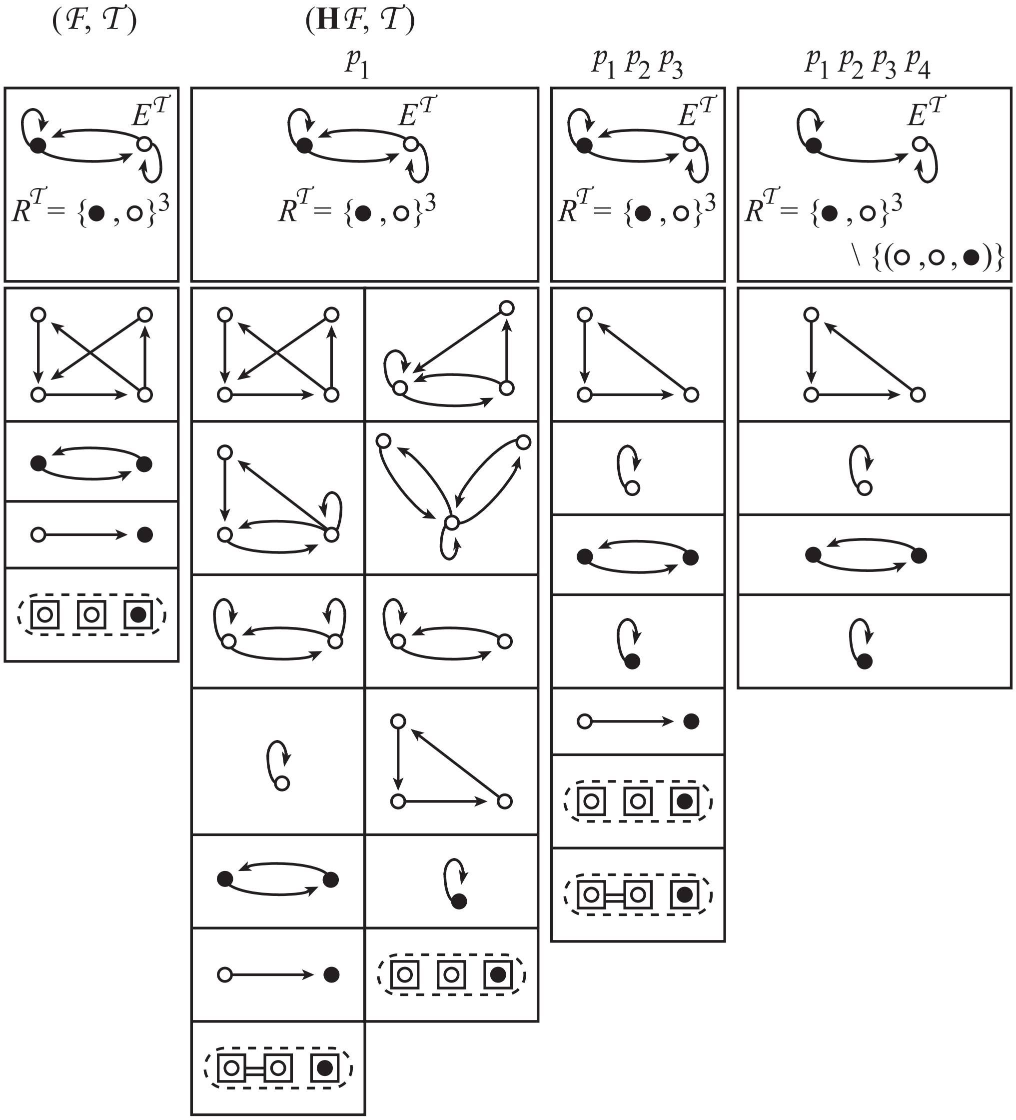

Fig. 5. *Towards a normal representation: step one.*

in general to *partially* colored structures; by choosing to work with *fully* colored structures in our combinatorial setting, this is the price we have to pay). Bearing this in mind, we can extend the definition of a representation to allow compact forbidden patterns and call it a *compact representation*, with the problem defined by a compact representation being the problem defined by the representation obtained by expanding all of the compact forbidden patterns.

Clearly, we may assume that every representation is compact, by replacing every forbidden pattern $(\mathcal{A}, a^{\mathcal{T}})$ by the compact forbidden pattern $\{\mathcal{A}, \overline{\alpha}\}$, where for every x in A, $\overline{\alpha}(x) := \{a^{\mathcal{T}}(x)\}$. We say that $(\mathcal{A}, a^{\mathcal{T}})$ is a forbidden pattern of the compact representation $(\mathscr{F}, \mathcal{T})$, and write $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$, if it is one of the forbidden patterns obtained by expanding one of the compact forbidden patterns. Notice that the notion

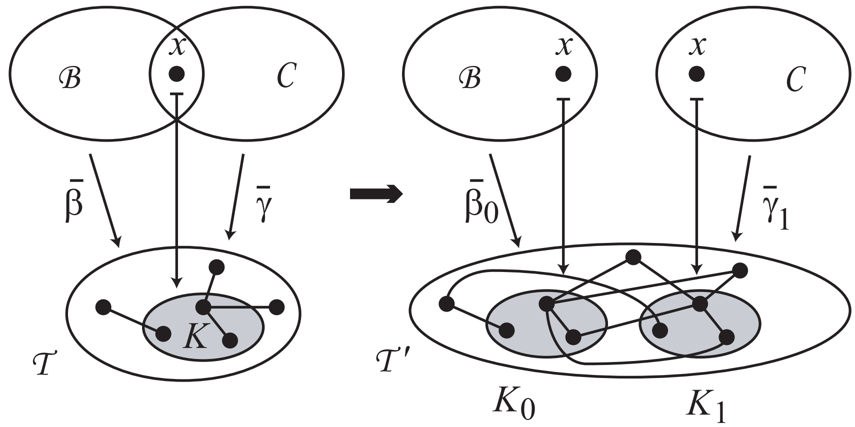

FIG. 6. *Feder–Vardi reduction.*

of a decomposition involves only the underlying structure; thus it generalizes to compact structures (of course, the definition of a decomposition in section 2 generalizes to colored structures).

DEFINITION 24 (Feder–Vardi reduction). *Let $(\mathscr{F}, \mathcal{T})$ be a compact representation with $\mathscr{F} = \mathscr{G} \cup \{\{\mathcal{B}, \overline{\beta}\} \overset{x}{32} \{\mathcal{C}, \overline{\gamma}\}\}$, and let $K = \overline{\beta}(x) = \overline{\gamma}(x)$. The new sets K_0 and K_1 are defined as $\{k_i : k \in K\}$ (that is, k_0 and k_1 are two new elements that stand as copies of k) for $i = 0, 1$, and we assume that K, K_0, and K_1 are mutually disjoint. Let $\mathcal{T}'$ be the structure obtained from $\mathcal{T}$ as follows:*

- *Replace K by K_0 and K_1 in $|\mathcal{T}|$.*
- *Set $R^{\mathcal{T}'}(\mathbf{t})$ whenever $R^{\mathcal{T}}(\tilde{\mathbf{t}})$, with $\mathbf{t}$ obtained from $\tilde{\mathbf{t}}$ by replacing every occurrence of an element $k \in K$ by either k_0 or k_1 (where two different occurrences of an element k might be replaced by k_0 and k_1; so, one tuple $R^{\mathcal{T}}(\tilde{\mathbf{t}})$ with i occurrences of elements of K gives rise to 2^i tuples $R^{\mathcal{T}'}(\mathbf{t})$).*

Let $\mathscr{F}'$ be the set of compact forbidden patterns induced from $\mathscr{F}$ as follows:

- *The compact forbidden pattern $\{\mathcal{B}, \overline{\beta}\} \overset{x}{32} \{\mathcal{C}, \overline{\gamma}\}$ is replaced by the two compact forbidden patterns induced from the decomposition so that*
  - *in the compact forbidden pattern $\{\mathcal{B}, \overline{\beta}_0\}$, $\overline{\beta}_0(x) = K_0$, and*
  - *in the compact forbidden pattern $\{\mathcal{C}, \overline{\gamma}_1\}$, $\overline{\gamma}_1(x) = K_1$.*
- *Every remaining occurrence of a color $k \in K$ in a compact forbidden pattern (including the two described above) is replaced by both k_0 and k_1; that is, every forbidden pattern obtained by expanding a compact forbidden pattern of $\mathscr{F}$ is replaced by a set of forbidden patterns, one for each possible assignment of k_0 and k_1 to occurrences of k.*

We call $(\mathscr{F}', \mathcal{T}')$ the Feder–Vardi reduction of $(\mathscr{F}, \mathcal{T})$ with respect to $\{\mathcal{B}, \overline{\beta}\} \overset{x}{32}$ $\{\mathcal{C}, \overline{\gamma}\}$.

Part of a Feder–Vardi reduction can be visualized as in Figure 6. Note that if $(\mathscr{F}, \mathcal{T})$ is a connected representation, then a Feder–Vardi reduction of $(\mathscr{F}, \mathcal{T})$ is also connected.

We reiterate that working with compact forbidden patterns is just, to some extent, a notational convenience and that a Feder–Vardi reduction has the effect of "splitting" a set of forbidden patterns in one go.

We also need to define the essential notion of a recoloring. Intuitively, a recoloring is to a (compact) representation what a homomorphism is to a structure.

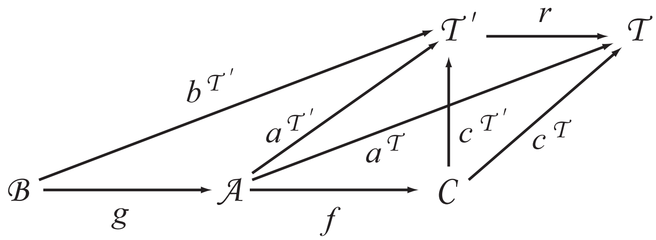

FIG. 7. *Proof of Proposition* 26.

DEFINITION 25 (Recoloring). *Let* $(\mathscr{F},\mathcal{T})$ *and* $(\mathscr{F}',\mathcal{T}')$ *be two compact representations. A* recoloring r of $(\mathscr{F}',\mathcal{T}')$ to $(\mathscr{F},\mathcal{T})$ *is a homomorphism* $\mathcal{T}'\xrightarrow{r}\mathcal{T}$ *such that any* inverse image $(\mathcal{A},a^{\mathcal{T}'})$ *of a forbidden pattern* $(\mathcal{A},a^{\mathcal{T}})$ *of* $\mathscr{F}$ *is not valid w.r.t.* $(\mathscr{F}',\mathcal{T}')$, *where by "inverse image" we mean that* $r\circ a^{\mathcal{T}'}=a^{\mathcal{T}}$. *We denote the fact that* r *is a recoloring by* $(\mathscr{F}',\mathcal{T}')\xrightarrow{r}(\mathscr{F},\mathcal{T})$ *(we use the same notation as for homomorphisms without causing any confusion). If, furthermore,* r *is onto (respectively, one-to-one), then* r *is an* epirecoloring *(respectively,* monorecoloring*). If* $(\mathscr{F}',\mathcal{T}')\overset{r}{\hookrightarrow\!\!\!\to}(\mathscr{F},\mathcal{T})$ *and* $(\mathscr{F},\mathcal{T})\xrightarrow{r^{-1}}(\mathscr{F}',\mathcal{T}')$, *then* r *is an* isorecoloring, *and we write* $(\mathscr{F},\mathcal{T})\approx(\mathscr{F}',\mathcal{T}')$.

The fact that for CSP, $\mathrm{CSP}(\mathcal{A})\subseteq\mathrm{CSP}(\mathcal{B})$ whenever $\mathcal{A}\to\mathcal{B}$, generalizes to FPP.

PROPOSITION 26. *Let* $(\mathscr{F},\mathcal{T})$ *and* $(\mathscr{F}',\mathcal{T}')$ *be two compact representations. If* $(\mathscr{F}',\mathcal{T}')\to(\mathscr{F},\mathcal{T})$, *then* $\mathrm{FPP}(\mathscr{F}',\mathcal{T}')\subseteq\mathrm{FPP}(\mathscr{F},\mathcal{T})$.

Proof. Let $(\mathscr{F}',\mathcal{T}')\xrightarrow{r}(\mathscr{F},\mathcal{T})$, and let $\mathcal{C}$ be a structure that is not valid w.r.t. $(\mathscr{F},\mathcal{T})$. If $\mathcal{C}\nrightarrow\mathcal{T}'$, then $\mathcal{C}$ is not valid w.r.t. $(\mathscr{F}',\mathcal{T}')$, and we are done. Thus, let $\mathcal{C}\xrightarrow{c^{\mathcal{T}'}}\mathcal{T}'$ and define $c^{\mathcal{T}}:=r\circ c^{\mathcal{T}'}$. By assumption, there exists a forbidden pattern $(\mathcal{A},a^{\mathcal{T}})\in\mathscr{F}$ such that $(\mathcal{A},a^{\mathcal{T}})\xrightarrow{f}(\mathcal{C},c^{\mathcal{T}})$; so define $a^{\mathcal{T}'}:=c^{\mathcal{T}'}\circ f$, with the result that $a^{\mathcal{T}}=r\circ a^{\mathcal{T}'}$ (see Figure 7). Since r is a recoloring, there exists a forbidden pattern $(\mathcal{B},b^{\mathcal{T}'})\in\mathscr{F}'$ such that $(\mathcal{B},b^{\mathcal{T}'})\xrightarrow{g}(\mathcal{A},a^{\mathcal{T}'})$. This can be summarized by the commuting diagram of Figure 7.

Hence, we can see that $(\mathcal{B},b^{\mathcal{T}'})\xrightarrow{f\circ g}(\mathcal{C},c^{\mathcal{T}'})$, which proves that $(\mathcal{C},c^{\mathcal{T}'})$ is not valid w.r.t. $(\mathscr{F}',\mathcal{T}')$, and we are done. □

PROPOSITION 27. *Let* $(\mathscr{F}',\mathcal{T}')$ *be obtained from* $(\mathscr{F},\mathcal{T})$ *via a Feder–Vardi reduction, as in Definition* 24. Then $(\mathscr{F}',\mathcal{T}')$ and $(\mathscr{F},\mathcal{T})$ are equivalent.

Proof. Let $\mathcal{T}'\xrightarrow{r}\mathcal{T}$ be the homomorphism that identifies $k_i\in K_i$ for $i=0,1$, with $k\in K$, and leaves all other elements fixed. We begin by proving that r is a recoloring.

By construction, the inverse images of any forbidden pattern of $\mathscr{G}$ belong to $\mathscr{F}'$. So, it remains to check the inverse images of the patterns expanded from the compact forbidden pattern $\{\mathcal{B},\overline{\beta}\}\overset{x}{32}\ \{\mathcal{C},\overline{\gamma}\}$. Assume without loss of generality (w.l.o.g.) that we are checking an inverse image where x takes a color from K_0. Consider the substructure of the inverse image induced by B. By construction, this substructure is one of the patterns expanded from the compact forbidden pattern $\{\mathcal{B},\overline{\beta}_0\}$ (constructed as in Definition 24), which is a compact forbidden pattern of $\mathscr{F}'$. The case when x takes a color from K_1 is similar. Hence, r is a recoloring. By Proposition 26, $\mathrm{FPP}(\mathscr{F}',\mathcal{T}')\subseteq\mathrm{FPP}(\mathscr{F},\mathcal{T})$.

Conversely, suppose that $(\mathcal{A},a^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F},\mathcal{T})$. We construct a coloring $a^{\mathcal{T}'}$ from $a^{\mathcal{T}}$ as follows:

- For any $y\in A$ such that $a^{\mathcal{T}}(y)\notin K$, set $a^{\mathcal{T}'}(y)=a^{\mathcal{T}}(y)$;
- Suppose that $a^{\mathcal{T}}(y)=k\in K$. As $(\mathcal{A},a^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F},\mathcal{T})$, there does not exist a homomorphism from any forbidden pattern expanded from the

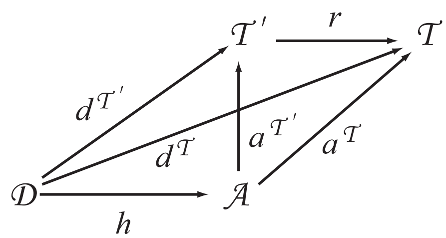

FIG. 8. *The first case.*

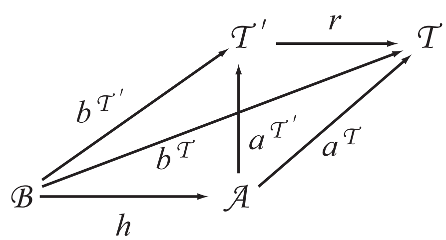

FIG. 9. *The second case.*

compact forbidden pattern $\{\mathcal{B},\overline{\beta}\}\overset{x}{32}\ \{\mathcal{C},\overline{\gamma}\}$ to $(\mathcal{A},a^{\mathcal{T}})$. That is, there does not exist $(\mathcal{B},b^{\mathcal{T}}) \in \{\mathcal{B},\overline{\beta}_0\}$ and $(\mathcal{C},c^{\mathcal{T}}) \in \{\mathcal{C},\overline{\gamma}_1\}$ such that both $(\mathcal{B},b^{\mathcal{T}})\overset{h_{\mathcal{B}}}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}})$ and $(\mathcal{C},c^{\mathcal{T}})\overset{h_{\mathcal{C}}}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}})$, with $h_{\mathcal{B}}(x) = h_{\mathcal{C}}(x) = y$. Thus

- if there exists $(\mathcal{B},b^{\mathcal{T}}) \in \{\mathcal{B},\overline{\beta}_0\}$ such that $(\mathcal{B},b^{\mathcal{T}})\overset{h_{\mathcal{B}}}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}})$, with $h_{\mathcal{B}}(x) = y$, then set $a^{\mathcal{T}'}(y) = k_1$;
- otherwise, set $a^{\mathcal{T}'}(y) = k_0$.

Suppose that $(\mathcal{D},d^{\mathcal{T}'})\overset{h}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}'})$, where $(\mathcal{D},d^{\mathcal{T}'})$ is derived from some forbidden pattern $(\mathcal{D},d^{\mathcal{T}})$ of (some compact forbidden pattern of) $\mathscr{G}$ (according to the Feder–Vardi reduction). Thus, we have the commutative diagram of Figure 8. This yields a contradiction as $(\mathcal{A},a^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F},\mathcal{T})$.

Suppose that $(\mathcal{B},b^{\mathcal{T}'})\overset{h}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}'})$, where $(\mathcal{B},b^{\mathcal{T}'})$ is derived from the compact forbidden pattern $\{\mathcal{B},\overline{\beta}\}\overset{x}{32}\ \{\mathcal{C},\overline{\gamma}\}$ (according to the Feder–Vardi reduction). Thus, we have the commutative diagram of Figure 9. In particular, $(\mathcal{B},b^{\mathcal{T}})\overset{h}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}})$, where $(\mathcal{B},b^{\mathcal{T}}) \in \{\mathcal{B},\overline{\beta}_0\}$. Set $h(x) = y$. By definition of $a^{\mathcal{T}'}$, $a^{\mathcal{T}'} \circ h(x) \in K_1$. However, by definition of $\{\mathcal{B},\overline{\beta}_0\}$, $b^{\mathcal{T}'}(x) \in K_0$. The fact that $b^{\mathcal{T}'} = a^{\mathcal{T}'} \circ h$ yields a contradiction.

Suppose that it is not the case that $(\mathcal{B},b^{\mathcal{T}'})\overset{h}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}'})$ for any $(\mathcal{B},b^{\mathcal{T}'})$ derived from the compact forbidden pattern $\{\mathcal{B},\overline{\beta}\}\overset{x}{32}\ \{\mathcal{C},\overline{\gamma}\}$ (according to the Feder–Vardi reduction) but that $(\mathcal{C},c^{\mathcal{T}'})\overset{h}{\longrightarrow}(\mathcal{A},a^{\mathcal{T}'})$ for some $(\mathcal{C},c^{\mathcal{T}'})$ derived from the compact forbidden pattern $\{\mathcal{B},\overline{\beta}\}\overset{x}{32}\ \{\mathcal{C},\overline{\gamma}\}$. A contradiction follows by reasoning analogously to the preceding case. Hence, we have that $\mathrm{FPP}(\mathscr{F},\mathcal{T}) \subseteq \mathrm{FPP}(\mathscr{F}',\mathcal{T}')$. ◻

PROPOSITION 28. *Let $(\mathscr{F}',\mathcal{T}')$ be obtained from $(\mathscr{F},\mathcal{T})$ via a Feder–Vardi reduction, as in Definition* 24. *If property* $\mathtt{p1}$ *holds for $(\mathscr{F},\mathcal{T})$, then it holds for $(\mathscr{F}',\mathcal{T}')$.*

Proof. Define $\mathcal{T}'\overset{r}{\longrightarrow}\mathcal{T}$ to be the homomorphism that identifies $k_i \in K_i$ for $i = 0,1$, with $k \in K$, and leaves all other elements fixed. Let $\mathcal{A}$ be nonvalid w.r.t. $(\mathscr{F}',\mathcal{T}')$. Since $(\mathscr{F},\mathcal{T})$ is equivalent to $(\mathscr{F}',\mathcal{T}')$, by Proposition 27, it follows that $\mathcal{A}$ is nonvalid w.r.t. $(\mathscr{F},\mathcal{T})$. We may assume that $\mathcal{A}\longrightarrow\mathcal{T}'$. Let $\mathcal{A}\overset{a^{\mathcal{T}'}}{\longrightarrow}\mathcal{T}'$ and define $a^{\mathcal{T}} := r \circ a^{\mathcal{T}'}$. As $\mathtt{p1}$ holds for $(\mathscr{F},\mathcal{T})$, there exists $(\mathcal{D},d^{\mathcal{T}}) \in \mathscr{F}$ such that $(\mathcal{D},d^{\mathcal{T}})\overset{f}{\hookrightarrow}(\mathcal{A},a^{\mathcal{T}})$. In particular,

$r \circ a^{\mathcal{T}'} \circ f = d^{\mathcal{T}}$; so, $r \circ d^{\mathcal{T}'} = d^{\mathcal{T}}$ when we define $d^{\mathcal{T}'} := a^{\mathcal{T}'} \circ f$. If $(\mathcal{D}, d^{\mathcal{T}})$ is a pattern of $\mathscr{G}$, then $(\mathcal{D}, d^{\mathcal{T}'}) \in \mathscr{F}'$. If $(\mathcal{D}, d^{\mathcal{T}})$ is a pattern of $\{\mathcal{B}, \overline{\beta}\} \overset{x}{32} \{\mathcal{C}, \overline{\gamma}\}$, then a pattern of either $\{\mathcal{B}, \overline{\beta}_0\}$ or $\{\mathcal{C}, \overline{\gamma}_1\}$ (where these are the compact forbidden $\mathcal{T}'$-patterns as constructed in Definition 24) is a (colored) substructure of $(\mathcal{D}, d^{\mathcal{T}'})$. Hence, regardless, there exists $(\mathcal{E}, e^{\mathcal{T}'}) \in \mathscr{F}'$ such that $(\mathcal{E}, e^{\mathcal{T}'}) \overset{g}{\hookrightarrow} (\mathcal{D}, d^{\mathcal{T}'})$. As $d^{\mathcal{T}'} = a^{\mathcal{T}'} \circ f$, we have that $(\mathcal{D}, d^{\mathcal{T}'}) \overset{f}{\hookrightarrow} (\mathcal{A}, a^{\mathcal{T}'})$, and so $(\mathcal{E}, e^{\mathcal{T}'}) \overset{f \circ g}{\hookrightarrow} (\mathcal{A}, a^{\mathcal{T}'})$. Consequently, if $\mathcal{A}$ is weakly valid w.r.t. $(\mathscr{F}', \mathcal{T}')$, then it is valid w.r.t. $(\mathscr{F}', \mathcal{T}')$. □

We define the *rank* of a (connected) compact structure to be the number of applications of the operator 32 in order that all resulting compact structures are biconnected. We associate with a compact representation a *rank polynomial* $P(X) = \Sigma_i a_i X^i$, where a_i is the number of compact forbidden patterns of rank i. Let $(\mathscr{F}', \mathcal{T}')$ be obtained from $(\mathscr{F}, \mathcal{T})$ via a Feder–Vardi reduction, with P the rank polynomial of $(\mathscr{F}, \mathcal{T})$ and P' that of $(\mathscr{F}', \mathcal{T}')$. It is easy to check that $P' < P$, where $<$ denotes the standard well-ordering of polynomials. Consequently, any sequence of Feder–Vardi reductions is necessarily finite. It is in order to prove this finiteness that we consider compact representations; given that we now that any sequence of Feder–Vardi reductions is necessarily finite, we can now revert to dealing with standard, as opposed to compact, representations. Of course, all of the results in this section mentioning compact representations also hold for standard representations.

4.3. Enforcing $\mathfrak{p}1$ to $\mathfrak{p}5$. We now use the reductions developed so far to obtain, from any connected representation, an equivalent representation satisfying properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, $\mathfrak{p}4$, and $\mathfrak{p}5$. We remind the reader that we are still assuming all representations to be connected, and we note that all reductions so far defined preserve the property of a representation being connected.

Definition 29. *Let $(\mathscr{F}, \mathcal{T})$ be a representation where every forbidden pattern of $\mathscr{F}$ is automorphic and nonconform. Define*

$$\rho(\mathscr{F}, \mathcal{T}) = \max\{||(\mathcal{B}, b^{\mathcal{T}})|| : (\mathcal{B}, b^{\mathcal{T}}) \text{ is a forbidden pattern of } \mathscr{F} \text{ that is not biconnected}\},$$

where $||(\mathcal{B}, b^{\mathcal{T}})||$ is the number of tuples in $\mathcal{B}$; that is, the sum of the numbers of $\{||R^{\mathcal{B}}|| : R$ is a relation symbol of the underlying signature$\}$, where $||R^{\mathcal{B}}||$ is the number of tuples in the relation $R^{\mathcal{B}}$.

Consider the following process, starting with a (connected) representation $(\mathscr{F}, \mathcal{T})$. Replace $(\mathscr{F}, \mathcal{T})$ with the representation $(\mathbf{H}\mathscr{F}, \mathcal{T})$, and so, by Lemmas 14 and 15, $(\mathbf{H}\mathscr{F}, \mathcal{T})$ is equivalent to $(\mathscr{F}, \mathcal{T})$ and satisfies $\mathfrak{p}1$. Perform a maximal sequence of core, embed, and conform reductions and denote the resulting representation by $(\mathscr{F}_1, \mathcal{T}_1)$. In particular, every forbidden pattern of $\mathscr{F}_1$ is a core and nonconform, and so $\rho(\mathscr{F}_1, \mathcal{T}_1)$ is well-defined. If $\rho(\mathscr{F}_1, \mathcal{T}_1) = 0$, then halt.

Otherwise, perform a maximal sequence of Feder–Vardi reductions, followed by a maximal sequence of core, embed, and conform reductions. Denote the resulting representation by $(\mathscr{F}_2, \mathcal{T}_2)$. In particular, every forbidden pattern of $\mathscr{F}_2$ is a core and nonconform, and so $\rho(\mathscr{F}_2, \mathcal{T}_2)$ is well-defined. Also, the sequence of reductions performed in order to obtain $(\mathscr{F}_2, \mathcal{T}_2)$ from $(\mathscr{F}_1, \mathcal{T}_1)$ is such that:

- every forbidden pattern of $\mathscr{F}_1$ that is a biconnected ($\mathcal{T}_1$-colored) core gives rise to forbidden patterns of $\mathscr{F}_2$ that are also biconnected ($\mathcal{T}_2$-colored) cores (see Remark 22); and,
- any non-biconnected core of $\mathscr{F}_1$ is split into forbidden patterns, each of which has strictly less tuples than the original non-biconnected core of $\mathscr{F}_1$.

That is, $\rho(\mathscr{F}_2, \mathcal{T}_2) < \rho(\mathscr{F}_1, \mathcal{T}_1)$.

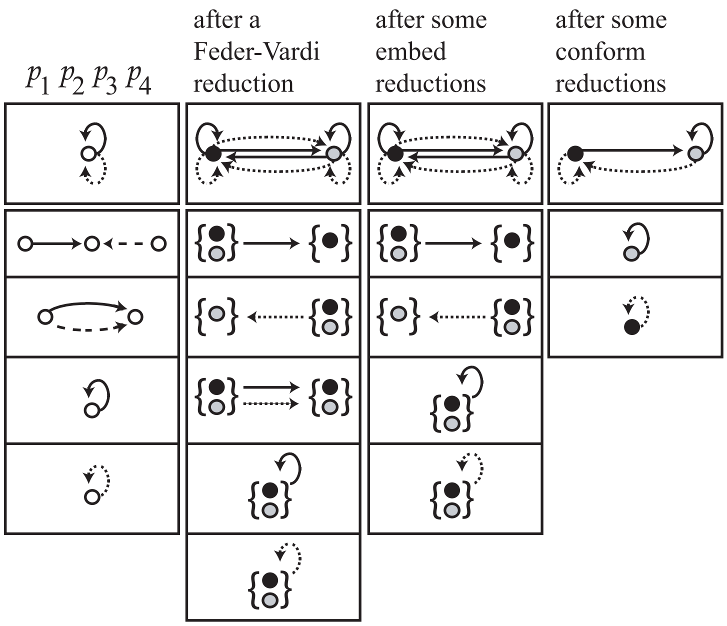

FIG. 10. *Applying our reductions.*

By iterating this process, we eventually obtain a connected representation $(\mathscr{F}', \mathcal{T}')$ that is equivalent to $(\mathscr{F}, \mathcal{T})$ and satisfies properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, $\mathfrak{p}4$, and $\mathfrak{p}5$.

Example 30. Consider a representation $(\mathscr{F}, \mathcal{T})$ over the signature consisting of two binary relation symbols E and F, where $\mathcal{T}$ and the forbidden patterns of $\mathscr{F}$ are as in the first column of Figure 10 (we represent "E-edges" by solid arrowed lines and "F-edges" by dashed arrowed lines).

As can be seen, $(\mathscr{F}, \mathcal{T})$ satisfies properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, and $\mathfrak{p}4$. However, one forbidden pattern is not biconnected, and so we perform a Feder–Vardi reduction so that the resulting compact forbidden pattern is as depicted in the second column of Figure 10. This messes up the aforementioned properties, and so we perform some embed reductions to obtain the compact representation in the third column of Figure 10 (we have left the depiction of this representation in its compact form so that the figure does not become cluttered). Finally, we perform some conform reductions to obtain the representation in the fourth column of Figure 10 which is equivalent to the original one and satisfies properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, $\mathfrak{p}4$, and $\mathfrak{p}5$.

4.4. Enforcing $\mathfrak{p}1$–$\mathfrak{p}6$. Given our notion of a recoloring of a representation, we can define a retract of a representation as follows.

DEFINITION 31. *A representation $(\mathscr{F}', \mathcal{T}')$ is a* retract *of the representation $(\mathscr{F}, \mathcal{T})$ if there exists a monorecoloring $(\mathscr{F}', \mathcal{T}') \overset{r}{\hookrightarrow} (\mathscr{F}, \mathcal{T})$ and an epirecoloring $(\mathscr{F}, \mathcal{T}) \overset{s}{\twoheadrightarrow} (\mathscr{F}', \mathcal{T}')$ such that $s \circ r = id_{\mathcal{T}'}$. We call a representation $(\mathscr{F}, \mathcal{T})$* automorphic *if whenever $(\mathscr{F}', \mathcal{T}')$ is a retract of $(\mathscr{F}, \mathcal{T})$, then $(\mathscr{F}', \mathcal{T}') \approx (\mathscr{F}, \mathcal{T})$.*

It is not difficult to see that given any representation $(\mathscr{F}, \mathcal{T})$, there is an automorphic representation $(\mathscr{F}', \mathcal{T}')$ that is a retract of $(\mathscr{F}, \mathcal{T})$ (and thus defines the same forbidden patterns problem by Proposition 26). We remark that the notion of a "core" for representations does not possess the properties that it does in the case of (colored) structures; e.g., it is not unique up to isorecoloring, but we resist the temptation to go into further details here as this has no consequence on what follows.

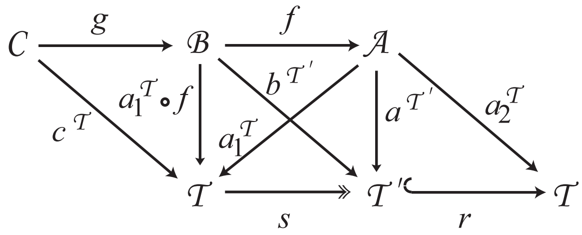

FIG. 11. *A commutative diagram.*

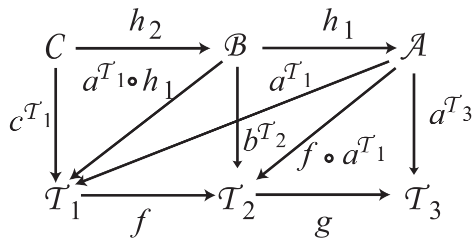

FIG. 12. *Proving transitivity.*

The next property we wish to enforce is as follows:

($\mathfrak{p}6$) The representation $(\mathscr{F}, \mathcal{T})$ is automorphic.

Suppose that $(\mathscr{F}, \mathcal{T})$ is not automorphic and that $(\mathscr{F}', \mathcal{T}') \overset{r}{\hookrightarrow} (\mathscr{F}, \mathcal{T})$, $(\mathscr{F}, \mathcal{T}) \overset{s}{\twoheadrightarrow} (\mathscr{F}', \mathcal{T}')$, and $s \circ r = id_{\mathcal{T}'}$, with $(\mathscr{F}', \mathcal{T}')$ automorphic. Define

$$\mathscr{F}'' = \{(\mathcal{A}, a^{\mathcal{T}'}) : (\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F} \text{ and } a^{\mathcal{T}} = r \circ a^{\mathcal{T}'}\}.$$

Let $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$, and let $a^{\mathcal{T}} = r \circ a^{\mathcal{T}'}$. By construction, $(\mathcal{A}, a^{\mathcal{T}'}) \in \mathscr{F}''$ and as such is not valid w.r.t. $(\mathscr{F}'', \mathcal{T}')$. Thus, $(\mathscr{F}'', \mathcal{T}') \overset{r}{\hookrightarrow} (\mathscr{F}, \mathcal{T})$. However, we also want to show that $(\mathscr{F}, \mathcal{T}) \overset{s}{\twoheadrightarrow} (\mathscr{F}'', \mathcal{T}')$.

Let $(\mathcal{A}, a^{\mathcal{T}'}) \in \mathscr{F}''$, and let $(\mathcal{A}, a_1^{\mathcal{T}})$ be such that $s \circ a_1^{\mathcal{T}} = a^{\mathcal{T}'}$; i.e., $(\mathcal{A}, a^{\mathcal{T}})$ is an inverse image of $(\mathcal{A}, a^{\mathcal{T}'})$ via s. Also, because $(\mathcal{A}, a^{\mathcal{T}'}) \in \mathscr{F}''$, by definition there exists $(\mathcal{A}, a_2^{\mathcal{T}}) \in \mathscr{F}$ such that $a_2^{\mathcal{T}} = r \circ a^{\mathcal{T}'}$. As $(\mathscr{F}', \mathcal{T}') \overset{r}{\hookrightarrow} (\mathscr{F}, \mathcal{T})$, there exists $(\mathcal{B}, b^{\mathcal{T}'}) \in \mathscr{F}'$ such that $(\mathcal{B}, b^{\mathcal{T}'}) \xrightarrow{f} (\mathcal{A}, a^{\mathcal{T}'})$. Hence, $(\mathcal{B}, a_1^{\mathcal{T}} \circ f)$ is an inverse image of $(\mathcal{B}, b^{\mathcal{T}'})$ via s, and so there exists $(\mathcal{C}, c^{\mathcal{T}}) \in \mathscr{F}$ such that $(\mathcal{C}, c^{\mathcal{T}}) \xrightarrow{g} (\mathcal{B}, a_1^{\mathcal{T}} \circ f)$ (see Figure 11). Thus, $(\mathcal{C}, c^{\mathcal{T}}) \xrightarrow{f \circ g} (\mathcal{A}, a_1^{\mathcal{T}})$ and $(\mathscr{F}, \mathcal{T}) \overset{s}{\twoheadrightarrow} (\mathscr{F}'', \mathcal{T}')$. In particular, $(\mathscr{F}'', \mathcal{T}')$ is a retract of $(\mathscr{F}, \mathcal{T})$.

We need the notion of a recoloring to be transitive.

LEMMA 32. *If $(\mathscr{F}_1, \mathcal{T}_1) \xrightarrow{f} (\mathscr{F}_2, \mathcal{T}_2)$ and $(\mathscr{F}_2, \mathcal{T}_2) \xrightarrow{g} (\mathscr{F}_3, \mathcal{T}_3)$ are recolorings, then $(\mathscr{F}_1, \mathcal{T}_1) \xrightarrow{g \circ f} (\mathscr{F}_3, \mathcal{T}_3)$ is a recoloring.*

Proof. Let $(\mathcal{A}, a^{\mathcal{T}_3}) \in \mathscr{F}_3$, and let $(\mathcal{A}, a^{\mathcal{T}_1}) \xrightarrow{g \circ f} (\mathcal{A}, a^{\mathcal{T}_3})$. As g is a recoloring of $(\mathscr{F}_2, \mathcal{T}_2)$ to $(\mathscr{F}_3, \mathcal{T}_3)$, there exists a forbidden pattern $(\mathcal{B}, b^{\mathcal{T}_2}) \in \mathscr{F}_2$ for which $(\mathcal{B}, b^{\mathcal{T}_2}) \overset{h_1}{\not\to} (\mathcal{A}, f \circ a^{\mathcal{T}_1})$. As f is a recoloring of $(\mathscr{F}_1, \mathcal{T}_1)$ to $(\mathscr{F}_2, \mathcal{T}_2)$, there exists a forbidden pattern $(\mathcal{C}, c^{\mathcal{T}_1}) \in \mathscr{F}_1$ for which $(\mathcal{C}, c^{\mathcal{T}_1}) \xrightarrow{h_2} (\mathcal{B}, a^{\mathcal{T}_1} \circ h_1)$. The situation can be depicted as in Figure 12. Consequently, $(\mathcal{C}, c^{\mathcal{T}_1}) \xrightarrow{h_1 \circ h_2} (\mathcal{A}, a^{\mathcal{T}_1})$, and $(\mathscr{F}_1, \mathcal{T}_1) \xrightarrow{g \circ f} (\mathscr{F}_3, \mathcal{T}_3)$. ∎

We have that $(\mathscr{F}'', \mathcal{T}') \overset{r}{\hookrightarrow} (\mathscr{F}, \mathcal{T})$ and $(\mathscr{F}, \mathcal{T}) \overset{s}{\twoheadrightarrow} (\mathscr{F}', \mathcal{T}')$, and consequently by Lemma 32, the identity map on $\mathcal{T}'$ is an isorecoloring from $(\mathscr{F}'', \mathcal{T}')$ to $(\mathscr{F}', \mathcal{T}')$. Thus, $(\mathscr{F}'', \mathcal{T}')$ is automorphic.

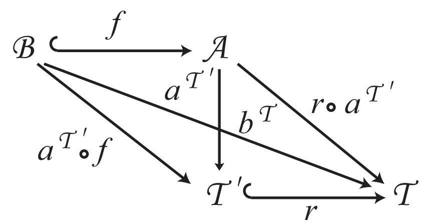

FIG. 13. *Verifying property* $\mathfrak{p}1$.

We now need to affirm the properties $\mathfrak{p}1$, $\mathfrak{p}2$, $\mathfrak{p}3$, $\mathfrak{p}4$, and $\mathfrak{p}5$ for $(\mathscr{F}'', \mathcal{T}')$; we deal with $\mathfrak{p}1$ first (note that if $(\mathscr{F}, \mathcal{T})$ is connected, then so is $(\mathscr{F}'', \mathcal{T}')$). Assume that $\mathcal{A}$ is not valid w.r.t. $(\mathscr{F}'', \mathcal{T}')$. Consequently, by Proposition 26, $\mathcal{A}$ is not valid w.r.t. $(\mathscr{F}, \mathcal{T})$. Suppose that $\mathcal{A} \stackrel{a^{\mathcal{T}'}}{\to} \mathcal{T}'$. Thus, $\mathcal{A} \stackrel{r\circ a^{\mathcal{T}'}}{\to} \mathcal{T}$. Hence, there exists $(\mathcal{B}, b^{\mathcal{T}}) \in \mathscr{F}$ such that $(\mathcal{B}, b^{\mathcal{T}}) \stackrel{f}{\hookrightarrow} (\mathcal{A}, r \circ a^{\mathcal{T}'})$; see Figure 13. So, $(\mathcal{B}, a^{\mathcal{T}'} \circ f) \in \mathscr{F}''$ and $(\mathcal{B}, a^{\mathcal{T}'} \circ f) \hookrightarrow (\mathcal{A}, a^{\mathcal{T}'})$. Thus, $\mathcal{A}$ is not weakly valid w.r.t. $(\mathscr{F}'', \mathcal{T}')$, and property $\mathfrak{p}1$ holds for $(\mathscr{F}'', \mathcal{T}')$.

Consider property $\mathfrak{p}2$. As every pattern of $(\mathscr{F}, \mathcal{T})$ is automorphic, by Lemma 2, so is every pattern of $(\mathscr{F}'', \mathcal{T}')$.

Consider property $\mathfrak{p}3$. Suppose that $(\mathcal{A}, a^{\mathcal{T}'}), (\mathcal{B}, b^{\mathcal{T}'}) \in \mathscr{F}''$ are distinct and such that $(\mathcal{B}, b^{\mathcal{T}'}) \stackrel{f}{\hookrightarrow} (\mathcal{A}, a^{\mathcal{T}'})$. Thus, we have that $(\mathcal{A}, r \circ a^{\mathcal{T}'}), (\mathcal{B}, r \circ b^{\mathcal{T}'}) \in \mathscr{F}$ and also that $(\mathcal{B}, r \circ b^{\mathcal{T}'}) \stackrel{f}{\hookrightarrow} (\mathcal{A}, r \circ a^{\mathcal{T}'})$. This yields a contradiction as $(\mathscr{F}, \mathcal{T})$ satisfies property $\mathfrak{p}3$, and so $(\mathscr{F}'', \mathcal{T}')$ satisfies property $\mathfrak{p}3$.

Trivially, $(\mathscr{F}'', \mathcal{T}')$ satisfies properties $\mathfrak{p}4$ and $\mathfrak{p}5$.

DEFINITION 33. *We say that a connected representation for which properties* $\mathfrak{p}1$–$\mathfrak{p}6$ *hold is a* normal representation.

Consequently, we have proven the following result.

THEOREM 34. *Let* $(\mathscr{F}, \mathcal{T})$ *be a connected representation. Then there is an effective procedure by which we can obtain a normal representation equivalent to* $(\mathscr{F}, \mathcal{T})$.

We end this section with a theorem crucial to what follows.

THEOREM 35. *Let* $(\mathscr{F}, \mathcal{T})$ *be a normal representation. If* $\mathscr{F} \neq \emptyset$, *then the target* $\mathcal{T}$ *is not valid w.r.t.* $(\mathscr{F}, \mathcal{T})$.

Proof. Assume for contradiction that $(\mathcal{T}, t)$ is valid w.r.t. $(\mathscr{F}, \mathcal{T})$. If t is one-to-one, then t is an isomorphism, and thus, as $\mathscr{F} \neq \emptyset$, there exists $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$ such that $t^{-1} \circ a^{\mathcal{T}}$ is a homomorphism from $(\mathcal{A}, a^{\mathcal{T}})$ to $(\mathcal{T}, t)$. This yields a contradiction, and so we may assume that t is not one-to-one.

Consider repeatedly applying the homomorphism t to obtain the homomorphism $t^k : \mathcal{T} \to \mathcal{T}$ for each $k \geq 1$. For some $k \geq 1$, it must be the case that t restricted to the image of t^k is one-to-one and thus an isomorphism. For such a k, denote the image of t^k by $\mathcal{T}'$ and the isomorphism from $\mathcal{T}'$ to $\mathcal{T}'$ induced by t by s. In particular, s^{-1} exists.

Suppose that there exists $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$ such that the image of $a^{\mathcal{T}}$ is contained in $\mathcal{T}'$. Clearly, the homomorphism $s^{-1} \circ a^{\mathcal{T}} : \mathcal{A} \to \mathcal{T}$ is well-defined and is a $\mathcal{T}$-colored homomorphism of $(\mathcal{A}, a^{\mathcal{T}})$ to $(\mathcal{T}, t)$. This contradicts our assumption that $(\mathcal{T}, t)$ is valid w.r.t. $(\mathscr{F}, \mathcal{T})$. Consequently, for every $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$, the image of $a^{\mathcal{T}}$ is not contained in $\mathcal{T}'$.

Consider the representation $(\emptyset, \mathcal{T}')$. Trivially, $(\emptyset, \mathcal{T}')$ is a retract of $(\mathscr{F}, \mathcal{T})$, and $\mathcal{T}'$ is not isomorphic to $\mathcal{T}$ (as t is not one-to-one). Thus, $(\emptyset, \mathcal{T}')$ is a proper retract of $(\mathscr{F}, \mathcal{T})$, which contradicts the fact that $(\mathscr{F}, \mathcal{T})$ is automorphic. The result follows. ∎

5. A generic construction of counterexamples. We prove in this section that any problem given by a normal representation $(\mathscr{F}, \mathcal{T})$ for which $\mathscr{F} \neq \emptyset$ is not in CSP. The proof involves a generic construction of a family of structures that provides, in a sense, a counterexample for any candidate for the role of a template; such a family of structures is called a witness family. The essence of the proof strategy employed originated in the proofs in [25] that certain graph problems are not in CSP.

Definition 36 (witness family). *Let $(\mathscr{F}, \mathcal{T})$ be a representation. A family of structures $\mathscr{W}$ is said to be a* witness family for $(\mathscr{F}, \mathcal{T})$ *if and only if $\mathscr{W} \subseteq \mathrm{FPP}(\mathscr{F}, \mathcal{T})$ and for any structure $\mathcal{B}$ (over the underlying signature), there exists $\mathcal{W} \in \mathscr{W}$ such that either $\mathcal{W} \nrightarrow \mathcal{B}$ or for some $\mathcal{W} \xrightarrow{h} \mathcal{B}$, the homomorphic image $h(\mathcal{W})$ does not belong to* $\mathrm{FPP}(\mathscr{F}, \mathcal{T})$ *(the structure $\mathcal{W}$ is said to be a* witness *for $\mathcal{B}$).*

Lemma 37. *If a representation (connected or otherwise) has a witness family, then the problem given by the representation does not belong to CSP.*

Proof. Let $\mathscr{W}$ be a witness family for some representation $(\mathscr{F}, \mathcal{T})$. Assume for contradiction that $\mathrm{FPP}(\mathscr{F}, \mathcal{T}) = \mathrm{CSP}(\mathcal{B})$ for some structure $\mathcal{B}$. By definition, there exists $\mathcal{W} \in \mathscr{W}$ such that either $\mathcal{W} \nrightarrow \mathcal{B}$ or for some $\mathcal{W} \xrightarrow{h} \mathcal{B}$, $h(\mathcal{W}) \notin \mathrm{FPP}(\mathscr{F}, \mathcal{T})$. Both cases immediately lead to a contradiction. ∎

We now state the main result of this section and a corollary.

Theorem 38. *Let $(\mathscr{F}, \mathcal{T})$ be a normal representation. If $\mathscr{F} \neq \emptyset$, then there is a witness family for $(\mathscr{F}, \mathcal{T})$.*

Corollary 39. *If $(\mathscr{F}, \mathcal{T})$ is a normal representation for which $\mathscr{F} \neq \emptyset$, then* $\mathrm{FPP}(\mathscr{F}, \mathcal{T}) \notin \mathrm{CSP}$.

The remainder of this section is devoted to a proof of the above theorem and corollary. Throughout the remainder of this section, $(\mathscr{F}, \mathcal{T})$ is a normal representation for which $\mathscr{F} \neq \emptyset$ and where the underlying signature is σ.

Opening up a structure. By Theorem 35, the structure $\mathcal{T}$ is not valid w.r.t. $(\mathscr{F}, \mathcal{T})$. Let $t^{\mathcal{T}}$ be some homomorphism $\mathcal{T} \xrightarrow{t^{\mathcal{T}}} \mathcal{T}$ (there is at least one such homomorphism: the identity). As $(\mathscr{F}, \mathcal{T})$ is normal, we may assume that some biconnected and nonconform forbidden pattern $(\mathcal{A}, a^{\mathcal{T}})$ embeds into $(\mathcal{T}, t^{\mathcal{T}})$, via some embedding f. Let $(\mathcal{D}, d^{\mathcal{T}})$ be identical to $(\mathcal{T}, t^{\mathcal{T}})$.

It is straightforward to show that any biconnected and nonconform pattern must contain a cycle; choose one of minimal size, and let C be the image of this cycle under f (and so C is a cycle). Let x be an articulation point of C, and let $\mathbf{t}$ be a tuple of C that is incident with x (thus $R^{\mathcal{D}}(\mathbf{t})$ holds for some relation symbol R). Introduce a new element x' into the domain of $\mathcal{D}$.

- Suppose that C has size 1; i.e., $\mathbf{t}$ is not antireflexive. Replace the first occurrence of x in $R^{\mathcal{D}}(\mathbf{t})$ with the new element x' (leaving all other occurrences of all elements as is).
- Suppose that C has size 2; i.e., C consists of the antireflexive tuples $R^{\mathcal{D}}(\mathbf{t})$ and $R_1^{\mathcal{D}}(\mathbf{t}_1)$, where $\mathbf{t}$ and $\mathbf{t}_1$ have at least two distinct elements in common (one of which is x) and where if $R = R_1$, then $\mathbf{t}$ and $\mathbf{t}_1$ differ. Replace the solitary occurrence of x in $R^{\mathcal{D}}(\mathbf{t})$ by x'.
- Suppose that C has size greater than 2. Replace the solitary occurrence of x in $R^{\mathcal{D}}(\mathbf{t})$ by x'.

The elements x and x' of our amended structure are called *plug points of sort* 1. We define that $d^{\mathcal{T}}(x') = d^{\mathcal{T}}(x)$ and denote the amended $\mathcal{T}$-colored structure by $(\mathcal{D}, d^{\mathcal{T}})$ also.

If there exists a forbidden pattern of $\mathscr{F}$ that embeds into $(\mathcal{D}, d^{\mathcal{T}})$, then we proceed as above by choosing an appropriate cycle and an articulation point y of this cycle and then "breaking" the cycle by introducing a new element y' and amending a specific tuple of $\mathcal{D}$ (note that if we have a cycle of size 1 or 2, then we may need more than one amendment to "break" the cycle). Again, we define $d^{\mathcal{T}}(y') = d^{\mathcal{T}}(y)$ and denote the amended $\mathcal{T}$-colored structure by $(\mathcal{D}, d^{\mathcal{T}})$ also. As above, we refer to y and y' as plug points. If y was either x or x', then y and y' are plug points of sort 1; otherwise, they are plug points of sort 2.

We proceed iteratively in this fashion until no forbidden pattern of $\mathscr{F}$ embeds into $(\mathcal{D}, d^{\mathcal{T}})$, at each stage of the iteration fixing the sort of plug points to be inherited from the corresponding articulation point or to be of a new sort (the smallest positive integer as yet unused to describe sorts) if the corresponding articulation point had not been assigned a sort. Note that this process terminates as ultimately we would obtain a cycle-free structure (into which no forbidden pattern can embed).

Denote the resulting $\mathcal{T}$-colored σ-structure by $(\mathcal{G}, g^{\mathcal{T}})$ and call it the *gadget*. Note that $(\mathcal{G}, g^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F}, \mathcal{T})$ as no forbidden pattern embeds into $(\mathcal{G}, g^{\mathcal{T}})$ (recall that $(\mathscr{F}, \mathcal{T})$ is normal). Note also that $(\mathcal{G}, g^{\mathcal{T}}) \xrightarrow{r} (\mathcal{T}, t^{\mathcal{T}})$, where r is the homomorphism which identifies plug points of the same sort and otherwise leaves elements fixed.

Preparing for plugging. Suppose that the gadget $(\mathcal{G}, g^{\mathcal{T}})$ has p_i plug points of sort i for $i = 1, 2, \ldots, k$ (and possibly other elements that have not been assigned a sort). For each $i = 1, 2, \ldots, k$, define the signature σ_i as consisting of the relation symbol P_i of arity p_i. For each $i = 1, 2, \ldots, k$ and each $m_i \geq p_i$, define the σ_i-structure $\mathcal{Q}_i^{m_i}$ to have domain $\{0, 1, \ldots, m_i - 1\}$ and relation $P_i^{\mathcal{Q}_i^{m_i}}$ defined as

$$\{(u_1, u_2, \ldots, u_{p_i}) : u_1 < u_2 < \cdots < u_{p_i}\}.$$

LEMMA 40. *Fix $b \geq 2$, fix $i \in \{1, 2, \ldots, k\}$, and suppose that $m_i \geq b(p_i - 1) + 1$. For every mapping $h : |\mathcal{Q}_i^{m_i}| \to \{0, 1, \ldots, b-1\}$, there must exist at least one tuple $P_i^{\mathcal{Q}_i^{m_i}}(u_1, u_2, \cdots, u_{p_i})$ such that $h(u_1) = h(u_2) = \ldots = h(u_{p_i})$.*

Proof. Suppose otherwise for the mapping h. So, there exist at most $p_i - 1$ distinct elements x of $|\mathcal{Q}_i^{m_i}|$ for which $h(x) = j$ for any $j \in \{0, 1, \ldots, b-1\}$. Thus, $|\mathcal{Q}_i^{m_i}| = m_i \leq b(p_i - 1)$, which yields a contradiction. ∎

Now define the signature $\overline{\sigma}$ to consist of the relation symbol P of arity $p = \Sigma_{i=1}^{k} p_i$. For any $m_1, m_2, \ldots, m_k$ for which $m_i \geq p_i$, for each $i = 1, 2, \ldots, k$, define the $\overline{\sigma}$-structure $\overline{\mathcal{Q}}$ to have a domain consisting of the disjoint union of the domains $|\mathcal{Q}_1^{m_1}|, |\mathcal{Q}_2^{m_2}|, \ldots, |\mathcal{Q}_k^{m_k}|$ and relation $P^{\overline{\mathcal{Q}}}$ defined as

$$\{(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k) : \mathbf{u}^i \in |\mathcal{Q}_i^{m_i}|^{p_i} \text{ and } u_1^i < u_2^i < \cdots < u_{p_i}^i \\ \text{for each } i = 1, 2, \ldots, k\}$$

(the notation is such that u_j^i is the jth component of the tuple $\mathbf{u}^i$). So, in a sense, $\overline{\mathcal{Q}}$ is a sort of "amalgamation" of $\mathcal{Q}_i^{m_1}, \mathcal{Q}_2^{m_2}, \ldots, \mathcal{Q}_k^{m_k}$ (note that we have suppressed the parameters "$m_1, m_2, \ldots, m_k$" in the denotation of $\overline{\mathcal{Q}}$ for ease of readability).

LEMMA 41. *Fix $b \geq 2$ and suppose that $m_i \geq b(p_i - 1) + 1$ for each $i = 1, 2, \ldots, k$. For every mapping $h : |\overline{\mathcal{Q}}| \to \{0, 1, \ldots, b-1\}$, there must exist at least one tuple $P^{\overline{\mathcal{Q}}}(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k)$ such that $h(u_1^i) = h(u_2^i) = \cdots = h(u_{p_i}^i)$ for all $i = 1, 2, \ldots, k$.*

Proof. The proof is immediate from Lemma 40. ∎

The *girth* of a structure is the length of its shortest cycle (and so if there are no cycles, then the structure has infinite girth). The following theorem is due to Feder and Vardi [15] (and generalizes a result due to Erdös; see [15]).

THEOREM 42. *Fix two positive integers r and s. For every structure $\mathcal{B}$ of size n, there exists a structure $\mathcal{B}'$ (over the same signature) of size n^a (where a depends solely on r and s) such that:*

- *the girth of $\mathcal{B}'$ is greater than r;*
- *$\mathcal{B}' \to \mathcal{B}$; and*
- *for every structure $\mathcal{C}$ of size at most s (over the same signature), $\mathcal{B} \to \mathcal{C}$ if and only if $\mathcal{B}' \to \mathcal{C}$.*

Furthermore, $\mathcal{B}'$ can be constructed from $\mathcal{B}$ in randomized polynomial time.

Remark 43. We have already mentioned that Gábor Kun has derandomized Theorem 1. To be more precise, he achieved this by giving a deterministic polynomial-time algorithm for the $\mathcal{B}'$ in the above theorem.

For each forbidden pattern $(\mathcal{A}, a^{\mathcal{T}})$ of $\mathscr{F}$, define $\gamma_{\mathcal{A}}$ to be the length of the longest cycle of $\mathcal{A}$. Define γ to be the maximum of $\{\gamma_{\mathcal{A}} : (\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}\}$.

Fix $b \geq 2$. By applying Theorem 42, there is a $\overline{\sigma}$-structure $\overline{\mathcal{Q}}'$ of girth greater than γ for which $\overline{\mathcal{Q}}' \to \overline{\mathcal{Q}}$ and for which for every structure $\mathcal{C}$ of size at most b, $\overline{\mathcal{Q}} \to \mathcal{C}$ if and only if $\overline{\mathcal{Q}}' \to \mathcal{C}$ (of course, we assume that $m_1, m_2, \ldots, m_k$ satisfy the hypothesis of Lemma 41).

LEMMA 44. *For every mapping $h : |\overline{\mathcal{Q}}'| \to \{0, 1, \ldots, b-1\}$, there must exist at least one tuple $P^{\overline{\mathcal{Q}}'}(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k)$ such that $h(u_1^i) = h(u_2^i) = \cdots = h(u_{p_i}^i)$ for all $i = 1, 2, \ldots, k$.*

Proof. The condition in the statement of the lemma (and also the statement of Lemma 41, with the same value b) is equivalent to there not being a homomorphism from $\overline{\mathcal{Q}}'$ to the $\overline{\sigma}$-structure with domain $\{0, 1, \ldots, b-1\}$ and relation

$$P = \{0, 1, \ldots, b-1\}^p \setminus \{(b_1^{p_1}, b_2^{p_2}, \ldots, b_k^{p_k}) : b_i \in \{0, 1, \ldots, b-1\} \text{ for every } i = 1, 2, \ldots, k\}$$

(where $b_i^{p_i}$ is the p_i-tuple with each component equal to b_i). The result follows by Lemma 41 and the properties of $\overline{\mathcal{Q}}'$ detailed above. □

Building the witness family. Fix some σ-structure $\mathcal{B}$ of size b. We are now in a position to build a σ-structure $\mathcal{W}_{\mathcal{B}}$ which will act as a witness for $\mathcal{B}$ (see Definition 36).

- Initialize the domain of $\mathcal{W}_{\mathcal{B}}$ to be that of $\overline{\mathcal{Q}}'$.
- For every tuple $P^{\overline{\mathcal{Q}}'}(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k)$, where each $\mathbf{u}^i \in |\overline{\mathcal{Q}}'|^{p_i}$, plug a copy of the gadget $\mathcal{G}$ by identifying the p_i sort-i plug points of $\mathcal{G}$ with the p_i "socket-points" $\mathbf{u}^i$ of $|\overline{\mathcal{Q}}'|$ for each $i = 1, 2, \ldots, k$.

All such copies of the gadget should be disjoint, except that two copies of the gadget may have plug points in common within $\mathcal{W}_{\mathcal{B}}$ and except where the gadget (possibly) contains a tuple $R^{\mathcal{G}}(\mathbf{t})$ with every element of $\mathbf{t}$ a plug point. Let us label every tuple of every relation $R^{\mathcal{W}_{\mathcal{B}}}$ with the name of the tuple of $P^{\overline{\mathcal{Q}}'}$ to which the copy of the gadget from which it comes corresponds. As just mentioned, there may be difficulties where the gadget contains a tuple $R^{\mathcal{G}}(\mathbf{t})$ with every element of $\mathbf{t}$ a plug point, as this tuple might require more than one label. In such a case, simply arbitrarily choose one label from the set of potential candidates. Finally, note that $\mathcal{W}_{\mathcal{B}} = \mathcal{W}_{\mathcal{B}'}$ whenever $|\mathcal{B}| = |\mathcal{B}'|$; i.e., the definition of $\mathcal{W}_{\mathcal{B}}$ depends solely upon b and not on the tuples of $\mathcal{B}$.

PROPOSITION 45. *The structure $\mathcal{W}_{\mathcal{B}}$ is a witness for $\mathcal{B}$.*

Proof. We begin by proving that there exists a homomorphism $\mathcal{W}_{\mathcal{B}} \stackrel{w^{\mathcal{T}}}{\to} \mathcal{T}$.

From above, $\overline{\mathcal{Q}}' \to \overline{\mathcal{Q}}$ via some homomorphism q. Recall that the domain of $\overline{\mathcal{Q}}$ is the disjoint union of $|\mathcal{Q}_1^{m_1}|, |\mathcal{Q}_2^{m_2}|, \ldots, |\mathcal{Q}_k^{m_k}|$. Hence, we can partition $|\overline{\mathcal{Q}}'|$ into disjoint subsets $S_1, S_2, \dots, S_k$, where for each $i = 1, 2, \dots, k$, $S_i = \{u \in |\overline{\mathcal{Q}}'| : q(u) \in |\mathcal{Q}_i^{m_i}|\}$. By definition of $P^{\overline{\mathcal{Q}}}$, if $P^{\overline{\mathcal{Q}}}(\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^k)$ holds, where $\mathbf{u}^i$ is a p_i-tuple of elements, then $\mathbf{u}^i \in S_i^{p_i}$ for $i = 1, 2, \dots, k$. In particular, in any copy of the gadget $\mathcal{G}$, plug points of sort i are always identified with "socket elements" from S_i for $i = 1, 2, \dots, k$. Consequently, the homomorphism $\mathcal{G} \xrightarrow{g^{\mathcal{T}}} \mathcal{T}$, under which plug points of the same sort are always mapped to the same element of $|\mathcal{T}|$, can be extended to a homomorphism $\mathcal{W}_{\mathcal{B}} \xrightarrow{w^{\mathcal{T}}} \mathcal{T}$.

Suppose that $(\mathcal{W}_{\mathcal{B}}, w^{\mathcal{T}})$ is not valid w.r.t. $(\mathscr{F}, \mathcal{T})$. So, some biconnected, nonconform forbidden pattern $(\mathcal{A}, a^{\mathcal{T}})$ embeds into $(\mathcal{W}_{\mathcal{B}}, w^{\mathcal{T}})$. As no forbidden pattern embeds into the gadget and each forbidden pattern is biconnected and nonconform, there must exist a cycle C in $\mathcal{W}_{\mathcal{B}}$ of length less than γ and involving tuples from at least two copies of the gadget within $\mathcal{W}_{\mathcal{B}}$ (we reiterate that each forbidden pattern is biconnected, and so if there were no such cycles, then we would have an articulation point) or equivalently, involving tuples labeled with at least two distinct tuples of $P^{\overline{\mathcal{Q}}'}$ (according to our labeling process as detailed prior to the statement of this proposition). However, the cycle C of $\mathcal{W}_{\mathcal{B}}$ yields a closed path of tuples in $\overline{\mathcal{Q}}'$ (by following the labels). Continuing, this closed path of tuples in $\overline{\mathcal{Q}}'$ yields a cycle in $\overline{\mathcal{Q}}'$ of length at least 2 and less than γ; this contradicts the fact that $\overline{\mathcal{Q}}'$ has girth greater than γ. Thus, $(\mathcal{W}_{\mathcal{B}}, w^{\mathcal{T}})$ is valid w.r.t. $(\mathscr{F}, \mathcal{T})$.

If $\mathcal{W}_{\mathcal{B}} \not\to \mathcal{B}$, then we are done. So, suppose that $\mathcal{W}_{\mathcal{B}} \xrightarrow{h} \mathcal{B}$. The homomorphism h induces a map $\hat{h} : |\overline{\mathcal{Q}}'| \to \{0, 1, \dots, b-1\}$, and so by Lemma 44, there exists a tuple $P^{\overline{\mathcal{Q}}'}(\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^k)$, where $\mathbf{u}^i \in |\overline{\mathcal{Q}}'|^{p_i}$ and $\hat{h}(u_1^i) = \hat{h}(u_2^i) = \cdots = \hat{h}(u_{p_i}^i)$ for $i = 1, 2, \dots, k$. Thus, by construction of $\mathcal{W}_{\mathcal{B}}$, $h(\mathcal{W}_{\mathcal{B}})$ contains a homomorphic image of the gadget $\mathcal{G}$ where all plug points of the same sort are mapped to the same element.

($\star$) Consequently, $h(\mathcal{W}_{\mathcal{B}})$ contains a homomorphic image of the structure $\mathcal{T}$, via some homomorphism $\tilde{h}$.

Suppose that $h(\mathcal{W}_{\mathcal{B}}) \xrightarrow{f} \mathcal{T}$. So, $\mathcal{T} \xrightarrow{f \circ \tilde{h}} \mathcal{T}$ and, by Theorem 35, there exists a forbidden pattern $(\mathcal{A}, a^{\mathcal{T}}) \in (\mathscr{F}, \mathcal{T})$ such that $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{\tilde{f}} (\mathcal{T}, f \circ \tilde{h})$. Hence, we have that $(\mathcal{A}, a^{\mathcal{T}}) \xrightarrow{\tilde{h} \circ \tilde{f}} (h(\mathcal{W}_{\mathcal{B}}), f)$, and $h(\mathcal{W}_{\mathcal{B}}) \notin \mathrm{FPP}(\mathscr{F}, \mathcal{T})$, as required. ∎

Thus, we have proven Theorem 38. Lemma 37 immediately yields Corollary 39.

6. MMSNP versus CSP.

We now deal with the disconnected case before turning to the more general situation involving MMSNP and CSP.

6.1. Normal sets of representations.

6.1.1. The disconnected case.

We first turn to the situation when a representation is not necessarily connected. Let $(\mathscr{F}, \mathcal{T})$ be a representation such that there exists a disconnected forbidden pattern $(\mathcal{A}, a^{\mathcal{T}}) \in \mathscr{F}$; that is, $(\mathcal{A}, a^{\mathcal{T}})$ is the disjoint union of two colored structures $(\mathcal{B}, b^{\mathcal{T}})$ and $(\mathcal{C}, c^{\mathcal{T}})$. Define $\mathscr{F}' = (\mathscr{F} \setminus \{(\mathcal{A}, a^{\mathcal{T}})\}) \cup \{(\mathcal{B}, b^{\mathcal{T}})\}$ and $\mathscr{F}'' = (\mathscr{F} \setminus \{(\mathcal{A}, a^{\mathcal{T}})\}) \cup \{(\mathcal{C}, c^{\mathcal{T}})\}$. Trivially, we have that

$$\mathrm{FPP}(\mathscr{F}, \mathcal{T}) = \mathrm{FPP}(\mathscr{F}', \mathcal{T}) \cup \mathrm{FPP}(\mathscr{F}'', \mathcal{T}).$$

By iterating this construction, we can transform $(\mathscr{F}, \mathcal{T})$ into a set of connected representations so that a structure is in $\mathrm{FPP}(\mathscr{F}, \mathcal{T})$ if and only if it is in at least one of the forbidden patterns problems corresponding to the derived connected representations.

Next, we compute the normal representation of each connected representation, just as we did in section 4. Finally, we enforce the following property on our set of normal representations:

($\mathfrak{p}7$) For any two normal representations $(\mathscr{F}', \mathcal{T}')$ and $(\mathscr{F}'', \mathcal{T}'')$, we have that $(\mathscr{F}', \mathcal{T}') \nrightarrow (\mathscr{F}'', \mathcal{T}'')$.

This property is enforced by simply removing the normal representation $(\mathscr{F}', \mathcal{T}')$ from the collection should there exist another (different) normal representation $(\mathscr{F}'', \mathcal{T}'')$ for which $(\mathscr{F}', \mathcal{T}') \rightarrow (\mathscr{F}'', \mathcal{T}'')$.

Consequently, we may assume that any representation $(\mathscr{F}, \mathcal{T})$ corresponds to a collection $\mathfrak{N}$ of normal representations (possibly containing only one such representation) for which property $\mathfrak{p}7$ holds; we call $\mathfrak{N}$ the *normal set* corresponding to $(\mathscr{F}, \mathcal{T})$. By Proposition 26, the problem FPP$(\mathscr{F}, \mathcal{T})$ is the union of the forbidden patterns problems of the representations in the normal set $\mathfrak{N}$; that is,

$$\mathrm{FPP}(\mathscr{F}, \mathcal{T}) = \bigcup \{\mathrm{FPP}(\mathscr{F}', \mathcal{T}') : (\mathscr{F}', \mathcal{T}') \in \mathfrak{N}\}.$$

6.1.2. Finite unions of forbidden patterns problems. The notion of a normal set extends naturally to finite unions of forbidden patterns problems: Given a finite set of representations, we split every disconnected representation into a set of connected representations as above, take the union of all of these sets, and simplify these sets so as to enforce $\mathfrak{p}7$. We write FPP$(\mathfrak{N})$ for $\bigcup_{(\mathfrak{F}, \mathcal{T}) \in \mathfrak{N}} \mathrm{FPP}(\mathfrak{F}, \mathcal{T})$.

PROPOSITION 46. *Let $\mathfrak{N}$ be a normal set that contains a representation $(\mathscr{F}', \mathcal{T}')$ such that $\mathscr{F}' \neq \emptyset$. Then $\mathcal{T}'$ is a no instance of* FPP$(\mathfrak{N})$.

Proof. By Theorem 35, if $\mathcal{T}'$ is valid w.r.t. $(\mathscr{F}', \mathcal{T}')$, then $\mathscr{F}' = \emptyset$. Thus, $\mathcal{T}'$ is not valid w.r.t. $(\mathscr{F}', \mathcal{T}')$.

Suppose that $\mathcal{T}'$ is valid w.r.t. $(\mathscr{F}'', \mathcal{T}'')$, where $(\mathscr{F}'', \mathcal{T}'')$ is a representation in $\mathfrak{N}$ distinct from $(\mathscr{F}', \mathcal{T}')$. That is, there exists a homomorphism $r : \mathcal{T}' \rightarrow \mathcal{T}''$ such that for every forbidden pattern $(\mathcal{A}'', a^{\mathcal{T}''}) \in \mathscr{F}''$, $(\mathcal{A}'', a^{\mathcal{T}''}) \nrightarrow (\mathcal{T}', r)$. In particular, if $(\mathcal{A}'', a^{\mathcal{T}''}) \in \mathscr{F}''$, then there does not exist a homomorphism $a^{\mathcal{T}'} : \mathcal{A}'' \rightarrow \mathcal{T}'$ for which $r \circ a^{\mathcal{T}'} = a^{\mathcal{T}''}$. Consequently, r is (trivially) a recoloring of $(\mathscr{F}', \mathcal{T}')$ to $(\mathscr{F}'', \mathcal{T}'')$. This yields a contradiction, and so $\mathcal{T}'$ is not valid w.r.t. $(\mathscr{F}'', \mathcal{T}'')$. The result follows. ∎

6.2. Finite unions.

DEFINITION 47 (strong witness family). *Let $\mathfrak{N}$ be a set of representations. A family of structures $\mathscr{W}$ is said to be a* strong witness family for $\mathfrak{N}$ *if and only if $\mathscr{W} \subseteq$* FPP$(\mathfrak{N})$ *and for any finite set of structures $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}$ (over the underlying signature), there exists $\mathcal{W} \in \mathscr{W}$ such that for every $1 \leq i \leq n$, either $\mathcal{W} \nrightarrow \mathcal{B}_i$ or for some $\mathcal{W} \xrightarrow{h} \mathcal{B}_i$, the homomorphic image $h(\mathcal{W})$ does not belong to* FPP$(\mathfrak{N})$ *(the structure $\mathcal{W}$ is said to be a* strong witness *for $\mathcal{B}$).*

LEMMA 48. *If a set of representations $\mathfrak{N}$ has a strong witness family, then the problem* FPP$(\mathfrak{N})$ *is not a finite union of constraint satisfaction problems.*

Proof. Let $\mathscr{W}$ be a strong witness family for some representation $(\mathscr{F}, \mathcal{T})$. Assume for contradiction that

$$\mathrm{FPP}(\mathfrak{N}) = \bigcup_{(\mathfrak{F}, \mathcal{T}) \in \mathfrak{N}} \mathrm{FPP}(\mathfrak{F}, \mathcal{T}) = \bigcup_{1 \leq i \leq n} \mathrm{CSP}(\mathcal{B}_i)$$

for some finite set of structures $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n\}$. By definition, there exists a strong witness $\mathcal{W} \in \mathscr{W}$. Since $\mathcal{W}$ is a yes instance of FPP$(\mathfrak{N})$, we have that $\mathcal{W} \in \mathrm{CSP}(\mathcal{B}_i)$ for some $1 \leq i \leq n$. Hence, by definition of a strong witness, there is a homomorphism $\mathcal{W} \xrightarrow{h} \mathcal{B}_i$ such that $h(\mathcal{W}) \notin$ FPP$(\mathfrak{N})$. However, $h(\mathcal{W}) \in \mathrm{CSP}(\mathcal{B}_i) = \mathrm{FPP}(\mathfrak{N})$, which is absurd. ∎

We can extend the main result of the previous section to finite unions of forbidden patterns problems and, in particular, to disconnected representations. We first deal with the case when the normal set corresponds to a finite union of constraint satisfaction problems.

THEOREM 49. *Let* $\mathfrak{N}$ *be a normal set of the form* $\{(\emptyset, \mathcal{T}_1), (\emptyset, \mathcal{T}_2), \dots, (\emptyset, \mathcal{T}_n)\}$. *Then* $\mathrm{FPP}(\mathfrak{N}) = \bigcup_{1 \le i \le n} \mathrm{CSP}(\mathcal{T}_i)$. *Moreover, if*

$$\bigcup_{1 \le i \le n} \mathrm{CSP}(\mathcal{T}_i) = \bigcup_{1 \le i \le m} \mathrm{CSP}(\mathcal{T}'_i),$$

then the following hold:

(i) *for every* $1 \le i \le m$*, there exists* $1 \le j \le n$ *such that* $\mathcal{T}'_i \rightarrow \mathcal{T}_j$*;*
(ii) *for every* $1 \le i \le n$*, there exists* $1 \le j \le m$ *such that* $\mathcal{T}_i$ *is the core of* $\mathcal{T}'_j$*;*
(iii) $m \ge n$*.*

Proof. Property (i) follows directly from the fact that $\mathcal{T}'_i \in \mathrm{CSP}(\mathcal{T}'_i)$. We now prove (ii). Using a similar argument as above, there exists $\mathcal{T}'_j$ such that $\mathcal{T}_i \rightarrow \mathcal{T}'_j$. By (i), there exists some $\mathcal{T}_k$ such that $\mathcal{T}'_j \rightarrow \mathcal{T}_k$. By composition, $\mathcal{T}_i \rightarrow \mathcal{T}_k$. Recall that, by definition of the normal set, there is no homomorphism between any $\mathcal{T}_i$ and $\mathcal{T}_k$ for any i such that $1 \le i < k \le n$. Moreover, every $\mathcal{T}_i$ is automorphic. Thus, $i = k$, and it follows that $\mathcal{T}_i$ is homomorphically equivalent to $\mathcal{T}'_j$. This proves that $\mathcal{T}_i$ is the core of $\mathcal{T}'_j$. Property (iii) follows from (ii) since $\mathcal{T}_i$ and $\mathcal{T}_k$, for any i, k such that $i \neq k$, cannot be the core of the same $\mathcal{T}'_j$; otherwise, they would be isomorphic (by uniqueness of the core). This concludes the proof. □

We can now precisely characterize when a normal set does not give rise to a finite union of constraint satisfaction problems.

THEOREM 50. *The following are equivalent:*

(i) *the normal set* $\mathfrak{N}$ *contains a representation* $(\mathscr{F}', \mathcal{T}')$*, with* $\mathscr{F}' \neq \emptyset$*;*
(ii) *the problem* $\mathrm{FPP}(\mathfrak{N})$ *is not a finite union of constraint satisfaction problems;*
(iii) *there exists a strong witness family for* $\mathfrak{N}$*.*

Proof. The implication (ii) $\implies$ (i) is the contrapositive of the (trivial statement in the) previous theorem. The implication (iii) $\implies$ (ii) holds by Lemma 48. We now prove that (i) $\implies$ (iii).

The case when $\mathfrak{N}$ is a singleton is a direct corollary of the proof of Theorem 38, as the construction of a witness family can be easily adapted to obtain a strong witness family. Indeed, as is pointed out just before the statement of Proposition 45, the construction of $\mathcal{W}_{\mathcal{B}}$ depends only on the size of $\mathcal{B}$. So, for a set of structures $\mathscr{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_n\}$, we build $\mathcal{W}_{\mathcal{B}_i}$, where $\mathcal{B}_i$ is a structure with the largest domain within this set. Now, for any structure $\mathcal{C}$ such that $|\mathcal{C}| \le |\mathcal{B}_i|$, if $\mathcal{W}_{\mathcal{B}_i} \xrightarrow{h} \mathcal{C}$, then $h(\mathcal{W}_{\mathcal{B}_i})$ contains a homomorphic image of the structure chosen as a basis for our gadget, namely, $\mathcal{T}'$ (see ($\star$) in the proof of Proposition 45), which is not a yes instance of $\mathrm{FPP}(\mathfrak{N})$ (otherwise, $\mathcal{T}'$ would also be a yes instance of $\mathrm{FPP}(\mathfrak{N})$, which would contradict Theorem 35). This means that $\mathcal{W}_{\mathcal{B}_i}$ is a strong witness for $\mathscr{B}$.

Suppose now that $\mathfrak{N}$ is not a singleton. By Proposition 46, $\mathcal{T}'$ is not valid w.r.t. $(\mathscr{F}'', \mathcal{T}'')$ for any $(\mathscr{F}'', \mathcal{T}'') \in \mathfrak{N}$. Thus, we may choose $\mathcal{T}'$ as the basis of our gadget and proceed as in the case of a singleton in order to get a strong witness family for $\mathfrak{N}$. □

6.3. The main result. We need a last definition before we can state the main result of this paper. Let Φ be a sentence of MMSNP. We call a *normal set of* Φ the normal set of the set of representations obtained from Φ as follows: First, Φ is logically equivalent to a finite set of primitive sentences, which we can build effectively as in the proof of Proposition 11; second, each such primitive sentence captures precisely a forbidden patterns problem (again, this is effective; see Theorem 12); finally, we compute the normal set of this set of representations. The main result of this paper is an exact characterization of the strict inclusion of MMSNP in CSP.

THEOREM 51. *Let* Φ *be a sentence of MMSNP. The problem defined by* Φ *is in CSP if and only if its normal set consists of a singleton* $(\emptyset, \mathcal{T})$*.*

Proof. The result follows from the definition of the normal set of Φ and from Theorems 49 and 50. ☐

7. Concluding remarks. Building upon a previous attempt by Feder and Vardi to provide a logical characterization of constraint satisfaction problems, we have introduced a new class of combinatorial problems, the forbidden patterns problems, and shown that they provide a combinatorial characterization of the logic MMSNP. Furthermore, we have provided a complete classification as to when forbidden patterns problems are in CSP, and there exists an effective procedure to decide whether a given forbidden patterns problem (or problem described by a sentence of MMSNP) is in CSP or not.

We end by describing two directions for further research. Tardif and Nešetřil [31] have characterized *duality pairs*, which correspond essentially to forbidden patterns problems with a single color (the target as only one element) that are also constraint satisfaction problems. Their elegant proof relies on a correspondence between these duality pairs and the notion of *density* (with respect to the partial order given by the existence of a homomorphism). This correspondence exists essentially because one can define the notion of the *exponential of a structure* (in graph theory, this notion plays an important role in relation with Hedetniemi's conjecture [29]). It turns out that a notion of the exponential of a representation can also be defined [24]. In a forthcoming paper, we will elaborate on this and delineate the relationship between the two approaches.

Another direction for further research relates to the containment problem and is as follows. A homomorphism problem is given by its template; hence, given two homomorphism problems CSP($\mathcal{A}$) and CSP($\mathcal{B}$) over the same signature, it is decidable whether CSP($\mathcal{A}$) $\subseteq CSP(\mathcal{B})$. As a matter of fact, the containment problem for homomorphism problems is nothing other than the uniform homomorphism problem, known to be NP-complete (as we noted in Remark 5). We would like to extend this result to the more general containment problem for forbidden patterns problems (given by their representations). Indeed, Feder and Vardi proved in [15] that the containment problem for MMSNP is decidable; hence by Theorem 13, it follows that the containment problem for forbidden patterns problems is decidable. However, to the best of our knowledge, nothing has been proved about the *complexity* of the containment problem for MMSNP.

We know that the existence of a recoloring implies the containment of the corresponding problems, and this provokes the following question: "*Does the existence of a recoloring correspond to the containment of the corresponding problems*?" However, we can answer this question negatively. Indeed, the major inconvenience of forbidden patterns problems, in comparison with homomorphism problems, is that the inclusion of two problems does not necessarily reduce to the question of the existence of a recoloring; for, in [24], an example is given where a representation is transformed into an equivalent representation, using Feder–Vardi reductions, but such that the representations are not equivalent with respect to recolorings. However, we think that the right notion of a morphism for representations should constitute a *finite sequence of recol-*

orings and Feder–Vardi reductions. More precisely, we believe that the following question can be answered affirmatively: "*Does the existence of a recoloring correspond to the containment of the corresponding problems in the case of normal (connected) representations*?" In [24], a few restricted cases for which an affirmative answer to the above question is obtained, and this leads us to propose the following conjecture (where for any representation $\mathscr{R}$, $\mathbf{normal}(\mathscr{R})$ is a normal representation equivalent to $\mathscr{R}$).

CONJECTURE 52. *Let* $\mathscr{R}_1$ *and* $\mathscr{R}_2$ *be two nontrivial connected representations.* $\mathrm{FPP}(\mathscr{R}_1) \subseteq \mathrm{FPP}(\mathscr{R}_2)$ *if and only if* $\mathbf{normal}(\mathscr{R}_1) \rightarrow \mathbf{normal}(\mathscr{R}_2)$.

Acknowledgment. The first author thanks Etienne Grandjean for his continual support.

REFERENCES

[1] M. Benedikt and L. Segoufin, *Regular tree languages definable in FO*, in Proceedings of the 22nd International Symposium on Theoretical Aspects of Computer Science (STACS'05), V. Diekert and B. Durand, eds., Lect. Notes Comput. Sci. 3404, Springer-Verlag, Berlin, 2006, pp. 327–339.

[2] M. Bodirsky and V. Dalmau, *Datalog and constraint satisfaction with infinite templates*, in Proceedings of the 23rd International Symposium on Theoretical Aspects of Computer Science (STACS'06), B. Durand and W. Thomas, eds., Lect. Notes Comput. Sci. 3884, Springer-Verlag, Berlin, 2006, pp. 646–659.

[3] A. A. Bulatov, *A dichotomy theorem for constraints on a three-element set*, in Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science (FOCS'02), 2002, pp. 649–658.

[4] A. A. Bulatov, *Tractable conservative constraint satisfaction problems*, in Proceedings of the 18th IEEE Symposium on Logic in Computer Science (LICS'03), 2003, pp. 321–330.

[5] A. A. Bulatov and V. Dalmau, *Towards a dichotomy theorem for the counting constraint satisfaction problem*, in Proceedings of the 44th IEEE Symposium on Foundations of Computer Science (FOCS'03), 2003, pp. 562–573.

[6] A. A. Bulatov, A. A. Krokhin, and P. Jeavons, *Constraint satisfaction problems and finite algebras*, in Proceedings of the 27th International Colloquium on Automata, Languages and Programming (ICALP'00), U. Montanari, J. D. P. Rolim and E. Welzl, eds., Lect. Notes Comput. Sci. 1853, Springer-Verlag, Berlin, 2000, pp. 272–282.

[7] A. K. Chandra and P. M. Merlin, *Optimal implementation of conjunctive queries in relational databases*, in Proceedings of the 9th ACM Symposium on Theory of Computing (STOC'77), 1977, pp. 77–90.

[8] J. Diaz, M. Serna, and D. M. Thilikos, *The complexity of restrictive H-coloring*, in Proceedings of the 28th International Workshop on Graph-Theoretic Concepts in Computer Science (WG'02), L. Kucera, ed., Lect. Notes Comput. Sci. 2573, Springer-Varlag, Berlin, 2002, pp. 126–137.

[9] M. Dyer and C. Greenhill, *The complexity of counting graph homomorphisms*, Random Structures Algorithms, 17 (2000), pp. 260–289.

[10] R. Fagin, *Generalized first-order spectra and polynomial-time recognizable sets*, in Complexity of Computation, R. M. Karp, ed., SIAM-AMS Proc. 7, 1974, pp. 43–73.

[11] T. Feder, *Classification of homomorphisms to oriented cycles and of k-partite satisfiability*, SIAM J. Discrete Math., 14 (2001), pp. 471–480.

[12] T. Feder and P. Hell, *List constraint satisfaction and list partition*, manuscript.

[13] T. Feder, P. Hell, and J. Huang, *List homomorphisms and circular arc graphs*, Combinatorica, 19 (1999), pp. 487–505.

[14] T. Feder, P. Hell, and J. Huang, *Bi-arc graphs and the complexity of list homomorphisms*, J. Graph Theory, 42 (1999), pp. 61–80.

[15] T. Feder and M. Y. Vardi, *The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory*, SIAM J. Comput., 28 (1998), pp. 57–104.

[16] H. Gaifman, H. Mairson, Y. Sagiv, and M. Y. Vardi, *Undecidable optimization problems for database logic programs*, J. ACM, 40 (1993), pp. 683–713.

[17] G. Hahn and C. Tardif, *Graph homomorphisms: Structure and symmetry*, in NATO Sci. Ser. C Math. Phys. Sci. 497, Kluwer, 1997, pp. 107–166.

[18] P. Hell and J. Nešetřil, *On the complexity of H-coloring*, J. Combin. Theory Ser. B, 48 (1990), pp. 92–110.

[19] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford Lecture Ser. Math. Appl., Oxford University Press, London, 2004.

[20] P. Jeavons, D. A. Cohen, and M. Gyssens, *Closure properties of constraints*, J. ACM, 44 (1997), pp. 527–548.

[21] P. G. Kolaitis and M. Y. Vardi, *Conjunctive-query containment and constraint satisfaction*, J. Comput. System Sci., 61 (2000), pp. 302–332.

[22] R. E. Ladner, *On the structure of polynomial time reducibility*, J. ACM, 22 (1975), pp. 155–171.

[23] B. Larose, C. Loten, and C. Tardif, *A characterisation of first-order constraint satisfaction problems*, in Proceedings of the 21st IEEE Symposium on Logic in Computer Science (LICS'06), IEEE, 2006, pp. 201–210.

[24] F. R. Madelaine, *Constraint satisfaction problems and related logic*, Ph.D. thesis, University of Leicester, Leicester, 2003.

[25] F. R. Madelaine and I. A. Stewart, *Some problems not definable using structure homomorphisms*, Ars Combin., 67 (2003), pp. 153–159.

[26] C. H. Papadimitriou, *Computational Complexity*, Addison-Wesley, Reading, MA, 1994.

[27] A. Puricella and I. A. Stewart, *A generic greedy algorithm, partially-ordered graphs and NP-completeness*, in Proceedings of the 27th International Workshop on Graph-Theoretic Concepts in Computer Science (WG'01), A. Brandstaedt and V. B. Le, eds., Lect. Notes Comput. Sci. 2204, Springer-Verlag, Berlin, 2001, pp. 306–316.

[28] A. Puricella and I. A. Stewart, *Greedy algorithms, H-colourings and a complexity-theoretic dichotomy*, Theoret. Comput. Sci., 290 (2003), pp. 1897–1913.

[29] N. Sauer, *Hedetniemi's conjecture – a survey*, Discrete Math., 229 (2001), pp. 261–292.

[30] T. J. Schaefer, *The complexity of satisfiability problems*, in Proceedings of the 10th ACM Symposium on Theory of Computing (STOC'78), 1978, pp. 216–226.

[31] C. Tardif and J. Nešetřil, *Duality theorems for finite structures (characterizing gaps and good characterizations)*, J. Combin. Theory Ser. B, 80 (2000), pp. 80–97.