Extensions in graph normal form

MICHAŁ WALICKI*, *Department of Informatics, University of Bergen, Norway.*

Abstract

Graph normal form, introduced earlier for propositional logic, is shown to be a normal form also for first-order logic. It allows to view syntax of theories as digraphs, while their semantics as kernels of these digraphs. Graphs are particularly well suited for studying circularity, and we provide some general means for verifying that circular or apparently circular extensions are conservative. Traditional syntactic means of ensuring conservativity, like definitional extensions or positive occurrences guaranteeing exsitence of fixed points, emerge as special cases.

1 Graph normal form

Graph normal form (GNF) for propositional logic was introduced in [1] and applied in [5] to analysis of paradoxes. We begin by showing in this section that it is normal form also for first-order logic (FOL). Section 2 shows how to represent theories in GNF as graphs and their classical semantics as graph kernels. Logical circularity emerges as graph cycles, of which only odd ones are vicious, leading possibly to inconsistency. Using graph-theoretic terms, Section 3 formulates conditions ensuring conservativity of extensions, which generalize those utilizing usual syntactic restrictions. Definitional extensions provide a special case, while the conditions are applicable in many situations involving extensions with circular or apparently circular definitions, like fixed-point definitions.

$$* * *$$

Given an FOL language $\mathcal{L}$ and a set D, by $\mathbf{T}_D$ we denote the free term algebra over D, and by $\mathbf{A}_D$ the atomic formulas formed by application of any predicate symbol $P \in \mathcal{L}$ to elements of D. The usual atoms are thus $\mathbf{A}_{\mathbf{T}_X}$, for some set of variables X.

DEFINITION 1.1
A formula of an FOL language $\mathcal{L}$ is in GNF if it is an equivalence where

- the left side is an atom, $LS \in \mathbf{A}_{\mathbf{T}_X}$,
- the right side, RS, is a (universally quantified) conjunction of negated atoms,
- all free variables of RS occur in LS, $\mathcal{V}(RS) \subseteq \mathcal{V}(LS)$.

A theory, i.e. a set of formulas, Γ is in GNF if each $F \in \Gamma$ is in GNF.

Atomic formulas are special cases of GNF, consisting of LS with empty RS. Predicate symbols not occurring in LS of any formula of a GNF theory Γ are referred to as 'undefined' by Γ. (Occurring elsewhere, they are as defined as others, but this figure of speech will find some justification.)

Given any formula ϕ in prenex normal form (PNF), its $GNF(\phi)$ is obtained by constructing first a restricted form of Morleyization, $GNF^-(\phi)$, introducing fresh predicates by a series of definitional extensions. In the resulting formulas, we write the universal quantifier $\forall x$ as $\bigwedge_x$.

---

*E-mail: michal@uib.no

EXAMPLE 1.2
In a PNF formula, say $\phi = \forall x \exists y Pxy$,[1] we first replace $\exists$ by $\neg\forall\neg$, obtaining $\forall x \neg \forall y \neg Pxy$, and then introduce fresh variable z and fresh predicates A, S with equivalences:

$Az \leftrightarrow \bigwedge_x \neg Sx$,
$Sx \leftrightarrow \bigwedge_y \neg Pxy$.

For $\exists y \forall x Pxy$, we first obtain $\neg\forall y \neg \forall x Pxy$, introduce a fresh variable z and then

$Nz \leftrightarrow \neg Rz$
$Rz \leftrightarrow \bigwedge_y \neg Qy$
$Qy \leftrightarrow \bigwedge_x \neg \overline{P}xy$
$\overline{P}xy \leftrightarrow \neg Pxy$.

Substituting back the introduced predicates shows the equivalence to the original formula:

$Az \leftrightarrow \bigwedge_x \neg Sx \leftrightarrow \bigwedge_x \neg \bigwedge_y \neg Pxy$,
$Nz \leftrightarrow \neg Rz \leftrightarrow \neg \bigwedge_y \neg Qy \leftrightarrow \neg \bigwedge_y \neg \bigwedge_x \neg \overline{P}xy \leftrightarrow \neg \bigwedge_y \neg \bigwedge_x \neg\neg Pxy \leftrightarrow \neg \bigwedge_y \neg \bigwedge_x Pxy$.

The remaining propositional matrix, if nonatomic, is rewritten using only $\neg$ and $\wedge$ and processed in an analogous manner, introducing new predicates. The following two definitions give the general construction of $GNF^-(\phi)$ for an arbitrary formula ϕ. The reader satisfied with the example can skip them on the first reading and continue now after Definition 1.4.

Function $\langle\_\rangle_-$ in Definition 1.3 transforms quantifier prefix of a PNF formula into GNF, sending the quantifier-free matrix for further processing by function $pGNF^-$ from Definition 1.4. The numerical argument i in $\langle\_\rangle_i$, counting the number of generated symbols, ensures that all introduced predicate symbols are distinct. All $\exists x$ in the initial formula are assumed replaced by $\neg\forall x \neg$. A block of quantifiers $\forall x_1 ... \forall x_n$ not separated by $\neg$ is abbreviated below as a single quantifier $\forall x$, with x abbreviating the sequence $x_1 ... x_n$. In the generated formulas, $\bigwedge_{x_1} ... \bigwedge_{x_n}$ replace then $\bigwedge_x$.

The two main cases in Definition 1.3 correspond to ϕ starting with $\neg\forall$ (1 and 3) or $\forall$ (2 and 4), each having two subcases, when ϕ is open (o) or closed (c). Cases (o) show also how $GNF^-(\phi)$ can be constructed for open ϕ. (Alternatively, one could construct it for ϕ's universal closure.) One starts typically with case 1 or 2, then performs step 2o until reaching the last quantifier, to which equation 2o, 3o or 4o is applied. Point 5 sends the remaining quantifier-free matrix ρ to $pGNF^-\langle\rho\rangle_j$ given in Definition 1.4. In case of only one quantifier, there is only one step, which may be any of 1 through 4.

DEFINITION 1.3
For a formula $\phi(X)$ in PNF, with $X = \mathcal{V}(\phi)$, we define $GNF^-(\phi) = \langle\phi(X)\rangle_1$, with the recursive function $\langle\_\rangle_-$ given by (z is a fresh free variable):

$$
\begin{array}{llll}
1o. & \langle\neg\forall x \neg \psi(X,x)\rangle_i & = & \{B_i(X) \leftrightarrow \neg B_{i+1}(X)\} & \cup & \langle\forall x \neg \psi(X,x)\rangle_{i+1} \\
1c. & \langle\neg\forall x \neg \psi(x)\rangle_i & = & \{B_i(z) \leftrightarrow \neg B_{i+1}(z)\} & \cup & \langle\forall \neg \psi(x)\rangle_{i+1} \\
2o. & \langle\forall x \neg \psi(X,x)\rangle_i & = & \{B_i(X) \leftrightarrow \bigwedge_x \neg B_{i+1}(X,x)\} & \cup & \langle\psi(X,x)\rangle_{i+1} \\
2c. & \langle\forall x \neg \psi(x)\rangle_i & = & \{B_i(z) \leftrightarrow \bigwedge_x \neg B_{i+1}(x)\} & \cup & \langle\psi(x)\rangle_{i+1}
\end{array}
$$

---

[1] We write Bx for $B(x)$, with x denoting a list of variables matching the arity of B, i.e. $x \in X^{ar(B)}$. Btx denotes an application of the predicate B to term(s) t, possibly with some variables x, where x may also comprise other arguments of B. $ar(Bt)$ denotes then the arity of the derived predicate resulting from such a substitution. Abbreviation $Bx \leftrightarrow \bigwedge_y Cxy$ assumes $x \cap y = \emptyset$ but admits all cases of $\mathcal{V}(Cxy) \setminus y \subseteq x$, with $\mathcal{V}(\phi)$ denoting free variables of ϕ. Substitution of t for x yields then Bt and Cty, where $\mathcal{V}(Cty) = (\mathcal{V}(Cxy) \setminus x) \cup \mathcal{V}(t)$. Such syntactic details would only clutter the main ideas, distinguishing cases which appear uniform from the level relevant to our considerations.

when quantifier-free ρ does not start with $\neg$:

$3o.$ $\quad \langle\neg\forall x\rho(X,x)\rangle_i \;=\; \{B_i(X) \leftrightarrow \neg B_{i+1}(X)\} \qquad\qquad \cup \quad \langle\forall x\rho(X,x)\rangle_{i+1}$

$3c.$ $\quad\;\; \langle\neg\forall x\rho(x)\rangle_i \;=\; \{B_i(z) \leftrightarrow \neg B_{i+1}(z)\} \qquad\qquad\;\; \cup \quad \langle\forall x\rho(x)\rangle_{i+1}$

$4o.$ $\quad\;\; \langle\forall x\rho(X,x)\rangle_i \;=\; \{B_i(X) \leftrightarrow \bigwedge_x \neg B_{i+1}(X,x),$
$$B_{i+1}(X,x) \leftrightarrow \neg B_{i+2}(X,x)\} \quad \cup \quad \langle\rho(X,x)\rangle_{i+2}$$

$4c.$ $\quad\;\;\; \langle\forall x\rho(x)\rangle_i \;=\; \{B_i(z) \leftrightarrow \bigwedge_x \neg B_{i+1}(x),$
$$B_{i+1}(x) \leftrightarrow \neg B_{i+2}(x)\} \qquad\;\; \cup \quad \langle\rho(x)\rangle_{i+2}$$

when ρ is quantifier-free, continue with Definition 1.4:

$5.$ $\qquad\qquad\quad \langle\rho\rangle_j \;=\; pGNF^-\langle\rho\rangle_j.$

For instance:

$i)$ $\qquad \langle\neg\forall x\neg\rho(x)\rangle_1 \;\overset{1c}{=}\; \{B_1(z) \leftrightarrow \neg B_2(z)\} \qquad\qquad \cup \quad \langle\forall x\neg\rho(x)\rangle_2$

$\qquad\qquad \langle\forall x\neg\rho(x)\rangle_2 \;\overset{2c}{=}\; \{B_2(z) \leftrightarrow \bigwedge_x \neg B_3(x)\} \qquad\;\; \cup \quad \langle\rho(x)\rangle_3...$

$ii)$ $\;\; \langle\forall x\neg\forall y\neg\rho(x,y)\rangle_1 \;\overset{2c}{=}\; \{B_1(z) \leftrightarrow \bigwedge_x \neg B_2(x)\} \qquad\;\; \cup \quad \langle\forall y\neg\rho(x,y)\rangle_2$

$\qquad\;\; \langle\forall y\neg\rho(x,y)\rangle_2 \;\overset{2o}{=}\; \{B_2(x) \leftrightarrow \bigwedge_y \neg B_3(x,y)\} \quad \cup \quad \langle\rho(x,y)\rangle_3...$

$iii)$ $\;\; \langle\neg\forall x\neg\forall y\rho(x,y)\rangle_1 \;\overset{1c}{=}\; \{B_1(z) \leftrightarrow \neg B_2(z)\} \qquad\qquad \cup \quad \langle\forall x\neg\forall y\rho(x,y)\rangle_2$

$\qquad\;\; \langle\forall x\neg\forall y\rho(x,y)\rangle_2 \;\overset{2c}{=}\; \{B_2(z) \leftrightarrow \bigwedge_x \neg B_3(x)\} \qquad\;\; \cup \quad \langle\forall y\rho(x,y)\rangle_3$

$\qquad\qquad\;\; \langle\forall y\rho(x,y)\rangle_3 \;\overset{4o}{=}\; \{B_3(x) \leftrightarrow \bigwedge_y \neg B_4(x,y),$
$$B_4(x,y) \leftrightarrow \neg B_5(x,y)\} \quad \cup \quad \langle\rho(x,y)\rangle_5...$$

DEFINITION 1.4

Given a quantifier-free $\rho(x)$ and index j, to form $pGNF_2^-\langle\rho\rangle_j$:

1. write $\rho(x)$ in CNF as: $(L_{1_1} \vee ... \vee L_{1_{z_1}}) \wedge ... \wedge (L_{n_1} \vee ... \vee L_{n_{z_n}})$ – each L_{i_k} is a literal;
2. using DeMorgan, transform it to: $\neg(L'_{1_1} \wedge ... \wedge L'_{1_{z_1}}) \wedge ... \wedge \neg(L'_{n_1} \wedge ... \wedge L'_{n_{z_n}})$, where L' removes $\neg$ from negative literals L and adds $\neg$ to the positive ones;
3. add the formula $B_j(x) \leftrightarrow \neg B_{j+1}(x_1) \wedge ... \wedge \neg B_{j+n}(x_n)$, where $x = x_1 \cup ... \cup x_n$ and each B_i, for $1 \leq i \leq n$, has the arity x_i of the i-th conjunct $L_{i_1} \vee ... \vee L_{i_{z_i}}$;
4. for each $1 \leq i \leq n$, add $B_{j+i}(x_i) \leftrightarrow \overline{L}_{i_1} \wedge ... \wedge \overline{L}_{i_{z_i}}$, where for $1 \leq j \leq zi$ and the predicate symbol R in L'_{i_j}, if $L'_{i_j} = \neg R(x_{i_j})$ then $\overline{L}_{i_j} = L'_{i_j} = \neg R(x_{i_j})$, while if $L'_{i_j} = R(x_{i_j})$, then $\overline{L}_{i_j} = \neg\overline{R}(x_{i_j})$ for a fresh symbol $\overline{R}$ with $ar(\overline{R}) = ar(R)$, and additional $\overline{R}(x_{i_j}) \leftrightarrow \neg R(x_{i_j})$; free variables $x_i = x_{i_1} \cup ... \cup x_{i_{z_i}}$.

$pGNF^-\langle\rho\rangle_j$ contains formulas from points 3 and 4 and $pGNF^-\langle\rho\rangle_j \models \rho(x) \leftrightarrow B_j x$. For instance:

$pGNF^-\langle(Cxy \rightarrow Dx) \wedge Ey\rangle_0$:

1. CNF $= (\neg Cxy \vee Dx) \wedge (Ey)$
2. $\neg(Cxy \wedge \neg Dx) \wedge \neg(\neg Ey)$
3. $B_0xy \leftrightarrow \neg B_1xy \wedge \neg B_2y$
4. $B_1xy \leftrightarrow \neg \overline{C}xy \wedge \neg Dx,$
$\overline{C}xy \leftrightarrow \neg Cxy$ and $B_2y \leftrightarrow \neg Ey$.

$pGNF^-\langle\neg Px\rangle_0$:

1. CNF $= \neg Px$
2. $\neg(Px)$, with $L'x = Px$
3. $B_0x \leftrightarrow \neg B_1x$
4. $B_1x \leftrightarrow \neg\overline{P}x$ $\overline{P}x \leftrightarrow \neg Px$

For an $\mathcal{L}$-formula ϕ, $GNF^-(\phi)$ denotes a GNF theory in an extended language $\mathcal{L}^-$, resulting from the procedure defined and exemplified above. The choice of the predicate symbols is inessential as long as they are all different, so we speak also about the $GNF^-(\phi)$. The predicate B_1 in $GNF^-(\phi)$ (Az in Example 1.2), not occurring in any RS and satisfying $GNF^-(\phi) \models B_1z \leftrightarrow \phi$, is called the

top predicate of $GNF^-(\phi)$. For a theory Γ, its $GNF^-(\Gamma)$ is obtained as the union of $GNF^-(\phi)$, for each $\phi \in \Gamma$, where distinct $GNF^-(\phi)$ introduce distinct predicate symbols.

$GNF^-(\Gamma)$ is a definitional extension of Γ, in the following sense. An *explicit definition*, in an FOL language $\mathcal{L}$, is $Bx \leftrightarrow \phi$, where predicate $B \notin \mathcal{L}$ and ϕ is an $\mathcal{L}$-formula with free variables $\mathcal{V}(\phi) \subseteq x$. A *definitional extension of* (the language of) Γ is a sequence $\langle D_i \rangle_{i < \lambda}$, for an ordinal λ, where (i) D_0 is an explicit definition in the language of Γ, (ii) for $j = i + 1 \leq \lambda$, D_j is an explicit definition in the language of all $D_k, k < j$ and (iii) $D_l = \bigcup_{i < l} D_i$ for limits $l \leq \lambda$. A set of formulas is a definitional extension of Γ if it can be well-ordered into one, and is a definitional extension *simpliciter*, if it is a definitional extension of some language.

FACT 1.5
For every FOL theory Γ, $GNF^-(\Gamma)$ is a definitional extension of Γ.

PROOF. We view the process from Definitions 1.3 and 1.4 bottom-up, so that for $j > i$, B_j is introduced before B_i. Step i adds the explicit definition $\bigwedge_x B_i(x) \leftrightarrow \bigwedge_y \neg B_{i+1}(x, y)$ of fresh predicate B_i in the language of the previous stage $i + 1$ (containing, in addition to the original symbols, the predicate symbols B_j for $j > i$). Thus, $GNF^-(A)$ is a definitional extension of A, for every $A \in \Gamma$. This yields the claim for finite Γ. If Γ is infinite, then we well-order Γ using axiom of choice. Since distinct $GNF^-(A_i)$ introduce distinct predicates, the resulting well-founded chain of explicit definitions is a definitional extension of Γ. $\qquad\square$

Now, theory $GNF^-(\phi)$ is satisfiable even if ϕ is not. To obtain equisatisfiability of ϕ and its GNF, we add one more formula.

DEFINITION 1.6
$GNF(\phi)$ is $GNF^-(\phi)$ with a fresh predicate symbol A', where $ar(A') = ar(A)$ for the top predicate A of $GNF^-(\phi)$, and the GNF formula

$$A'x \leftrightarrow (\neg Ax \wedge \neg A'x). \tag{1.7}$$

For a theory Γ, $GNF(\Gamma) = \bigcup_{\phi \in \Gamma} GNF(\phi)$, where distinct $GNF(\phi)$ introduce distinct symbols.

EXAMPLE 1.8
For $\phi = \forall x \exists y Pxy$ from Example 1.2, its $GNF^-(\phi)$

$\quad Az \quad \leftrightarrow \quad \bigwedge_x \neg Sx$
$\quad Sx \quad \leftrightarrow \quad \bigwedge_y \neg Pxy$

becomes $GNF(\phi)$ when extended with: $A'z \leftrightarrow (\neg Az \wedge \neg A'z)$.

Although $GNF(\Gamma)$ is no longer a definitional extension of Γ, due to formula (1.7), it shares its essential feature: every model of Γ has a unique expansion to a model of $GNF(\Gamma)$, because for any structure $M : M \models A'z \leftrightarrow (\neg Az \wedge \neg A'z)$ if and only if $M \models Az$. On the other hand, reduct of every model of $GNF(\Gamma)$, forgetting A', is a model of Γ. This gives the following fact, where $X \leftrightarrows Y$ denotes a bijection. (We show existence of injections $X \to Y$ and $Y \to X$, also when X, Y are classes, assuming their theory with Schröder–Bernstein theorem, e.g. NGB.)

FACT 1.9
$Mod(\Gamma) \leftrightarrows Mod(GNF(\Gamma))$, for each FOL theory Γ.

PROOF. If $M \models \Gamma$ then, as a consequence of Fact 1.5, M has a unique expansion M^- to the language $\mathcal{L}^-$, so that $M^- \models \Gamma \cup GNF^-(\Gamma)$. Since $M \models Ax$ for each $Ax \in \Gamma$, also $M^- \models Ax$. Interpreting

$A' = \emptyset$ in M^-, yields an expansion M' satisfying (1.7), i.e. $M' \models GNF(\Gamma)$. No other interpretation of A' over M^- satisfies (1.7), so the expansion of M to $M' \in Mod(GNF(\Gamma))$ is unique.

Conversely, if $M' \models GNF(\Gamma)$ then, in particular, $M' \models Ax$, for each $Ax \in \Gamma$, since M' satisfies each equivalence (1.7). Hence $M'|_{\mathcal{L}} \in Mod(\Gamma)$, where $M'|_{\mathcal{L}}$ denotes the $\mathcal{L}$-reduct of M'. □

GNF is thus a normal form for FOL, and we study only theories in GNF. Theory $GNF(\Gamma)$, given by Definitions 1.3–1.6, is only one possible GNF for Γ respecting the equivalence from the fact above, and we will occasionally use other, simpler forms.

2 Graphs as syntax

The syntax of a GNF theory can be represented by a graph, while the semantics amounts to specific properties of this syntax graph. By a 'graph' we mean directed graph, namely, a pair $G = (\mathbf{V}_G, \mathbf{E}_G)$ with vertices, $\mathbf{V}_G$ and edges, $\mathbf{E}_G \subseteq \mathbf{V}_G \times \mathbf{V}_G$. For $x \in \mathbf{V}_G$, by $\mathbf{E}_G(x)$ we denote the set of vertices with an edge from x, that is $\mathbf{E}_G(x) = \{y \in \mathbf{V}_G \mid \langle x, y \rangle \in \mathbf{E}_G\}$ and, dually, $\mathbf{E}_G^-(y) = \{x \in \mathbf{V}_G \mid \langle x, y \rangle \in \mathbf{E}_G\}$. $\mathbf{E}_G^*, \mathbf{E}_G^{-*}$ denote the reflexive transitive closure of $\mathbf{E}_G, \mathbf{E}_G^-$. This notation is extended to sets, e.g. $\mathbf{E}_G(X) = \bigcup_{v \in X} \mathbf{E}_G(v)$, etc.

One proviso is needed. Definition 1.1 admits in a GNF theory several equivalences with the same predicate in LSs, say, $Pa \leftrightarrow ...$ and $Pb \leftrightarrow ...$, with a, b being constants or even unifiable terms. Such general cases are treated in Section 2.2. Until then, definitions, results and examples are formulated for FOL without equality or function symbols, FOL$^-$, to convey the main ideas undisturbed by more detailed technicalities. Thus, $\mathbf{T}_X = X$ and the only terms in LS of each GNF formula are variables, $LS \in \mathbf{A}_X$. We assume also that each predicate symbol P occurs at most once in LS of some formula of a theory, to which we refer as P's equivalence.

We let B_i range over predicate symbols from the language $\mathcal{L}$ of a GNF theory Γ. Any formula from Γ (with the restrictions just mentioned) is represented schematically by the following pattern (where B may occur in RS as one of $B_1, ..., B_n$):

$$Bx \leftrightarrow \bigwedge_y \left(\neg B_1 xy \wedge ... \wedge \neg B_n xy \right). \tag{2.1}$$

DEFINITION 2.2
For a GNF theory Γ in FOL$^-$ and set D, the graph $G = \mathcal{G}_D(\Gamma)$ is given by:

1. $\mathbf{V}_G = \mathbf{A}_D \cup \overline{\mathbf{A}}_D$, with $\overline{\mathbf{A}}_D$ given in point 3.
2. For each axiom (2.1) of Γ, each vertex $Bd \in \mathbf{A}_D$, instantiating x by $d \in D^{ar(B)}$ in its LS, has the outgoing edges to $\mathbf{E}_G(Bd) = \{B_i dc \mid 1 \leq i \leq n, c \in D^{ar(B_i d)}\}$.
3. For each undefined predicate symbol P, we add a fresh symbol $\overline{P}$ and the 2-cycle $Pd \leftrightarrows \overline{P}d$ for each $d \in D^{ar(P)}$. Vertices $\overline{P}d$ form $\overline{\mathbf{A}}_D$.

A nonempty D is a domain of an FOL-structure interpreting the language $\mathcal{L}$ of the theory. We exclude the empty graph, $\mathcal{G}_\emptyset(\Gamma) = (\emptyset, \emptyset)$, from considerations, and denote by $Gr(\Gamma)$ the class of graphs $\mathcal{G}_D(\Gamma)$, for all nonempty sets D.

For each instance Bd, $d \in D^{ar(B)}$, of an atomic axiom Bx (with no RS), point 2 gives the empty set of neighbours, including Bd in $sinks(G) = \{v \in \mathbf{V}_G \mid \mathbf{E}_G(v) = \emptyset\}$. As will be seen in Section 2.1, sinks are included in every kernel of a graph, which represents their valuation as true. The copies $\overline{P}d$, added along with the 2-cycles for each undefined Pd in point 3, ensure that such atoms are not sinks and can obtain arbitrary values, restricted only by the rest of the theory.

EXAMPLE 2.3

For $\phi = \forall x \exists y Pxy$ with $\Gamma = GNF(\phi)$ from Example 1.8, repeated to the left, and for the set $D = \{c, d\}$, Definition 2.2 yields the graph $\mathcal{G}_D(\Gamma)$ to the right:
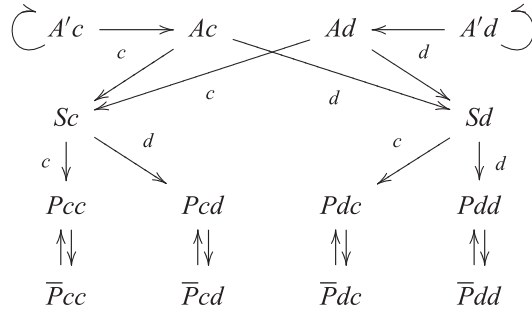
The subgraph induced[2] by $\{Ac, Ad, Sc, Sd\}$ and all Pxy vertices corresponds to $GNF^-(\phi)$, with vertices $\overline{P}xy$ and their 2-cycles to Pxy added according to Definition 2.2.3. The top vertices $A'c, A'd$ with their edges arise from formula (1.7) in Definition 1.6.

Generally, in $\mathcal{G}_D(\Gamma)$, each vertex Ad, for $d \in D$, has edges to $|D|$ copies of a subgraph with a source Sx and edges to Pxy, for all $x, y \in D$. Each pair $Pxy, \overline{P}xy$ forms a 2-cycle.

2.1 Kernels as models

Given a GNF theory Γ, we denote its usual models by $Mod(\Gamma)$, while its models over a given set D by $Mod_D(\Gamma)$. Graphs from Definition 2.2 provide an equivalent representation of these models.

Graph $\mathcal{G}_D(\Gamma)$ mixes syntax, using the predicate symbols from the language of Γ, with semantics, applying these symbols to the elements of the interpretation domain D. Typically, by 'atoms' we refer to such mixed expressions. One uses such a notational abbreviation when, for a formula $\phi(x)$, one writes $\phi(d)$ for some $d \in D$. It may denote (i) the formula $\phi(x)$ with a new constant—naming the object d—substituted for x, or else, (ii) the interpretation of the formula $\phi(x)$ under a valuation of variables assigning the object d to x. Vertices of our graphs come closest to (i)—formulas with names of the objects from D substituted for variables. They obtain truth values, becoming (ii), relatively to the solutions of the graph, which we now define.

A *kernel* (or a *solution* [4]) of a graph G is a subset $K \subseteq \mathbf{V}_G$ which

(a) is independent, i.e. $\mathbf{E}_G^-(K) \subseteq \mathbf{V}_G \setminus K$, and

(b) absorbs its complement, i.e. $\mathbf{E}_G^-(K) \supseteq \mathbf{V}_G \setminus K$,

in short, such that $\mathbf{E}_G^-(K) = \mathbf{V}_G \setminus K$. Equivalently, a kernel is an assignment $\alpha \in \mathbf{2}^{\mathbf{V}_G}$ such that $\forall x \in \mathbf{V}_G : \alpha(x) = \mathbf{1} \Leftrightarrow (\forall y \in \mathbf{E}_G(x) : \alpha(y) = \mathbf{0})$.

A graph G is *solvable* when $Ker(G) \neq \emptyset$, where $Ker(G)$ denotes the set of all kernels of G.

Vertices of $\mathcal{G}_D(\Gamma)$ contain all D instances of all atomic formulas. An assignment $v \in D^{\mathcal{V}(\phi)}$ to free variables $\mathcal{V}(\phi)$ of a formula ϕ, along with a kernel K, determine a valuation of all atoms over $\mathcal{V}(\phi)$, with atoms in K assigned $\mathbf{1}$, as given in line 1 below. Satisfaction of arbitrary formulas is defined from this basis in the usual way, with the needed adjustments in the last two lines. (For a structure M (or kernel of a graph) over a set D, a formula ϕx and $d \in D$, we write $M \models \phi d$ if M satisfies ϕx

---

[2]Given $X \subseteq \mathbf{V}_G$, the subgraph of G *induced* by X is $(X, \mathbf{E}_G \cap (X \times X))$.

with x assigned d.)

$$
\begin{array}{llll}
1. & (G,K) \models_v Px & \Leftrightarrow & P(v(x)) \in K, \text{ for atomic } Px \\
2. & (G,K) \models_v \neg\phi & \Leftrightarrow & (G,K) \not\models_v \phi \\
3. & (G,K) \models_v \phi_1 \wedge \phi_2 & \Leftrightarrow & (G,K) \models_v \phi_1 \text{ and } K \models_v \phi_2 \\
4. & (G,K) \models_v \forall x\phi x & \Leftrightarrow & \forall d \in D : (G,K) \models_v \phi d \\
and & (G,K) \models \phi & \Leftrightarrow & \forall v \in D^{\mathcal{V}(\phi)} : (G,K) \models_v \phi \\
& G \models \phi & \Leftrightarrow & \forall K \in Ker(G) : (G,K) \models \phi \\
& \Gamma \models \phi & \Leftrightarrow & \forall G \in Gr(\Gamma) : G \models \phi
\end{array}
\tag{2.4}
$$

Repeating the standard definition of satisfaction, this gives $\Gamma \models \phi$ coinciding with the standard notion $Mod(\Gamma) \models \phi$, provided that kernels of graphs in $Gr(\Gamma)$ correspond to $Mod(\Gamma)$. We establish this now, showing first that kernels of $\mathcal{G}_D(\Gamma)$ correspond to $Mod_D(\Gamma)$. (Given a graph G, $K \models \phi$ abbreviates $(G,K) \models \phi$.)

FACT 2.5
For a GNF theory Γ in FOL$^-$ and any nonempty set D, there are injections

$kr : Mod_D(\Gamma) \to Ker(\mathcal{G}_D(\Gamma))$ and $md : Ker(\mathcal{G}_D(\Gamma)) \to Mod_D(\Gamma)$,

such that $M \models \phi \Leftrightarrow kr(M) \models \phi$ and $K \models \phi \Leftrightarrow md(K) \models \phi$ for $M \in Mod_D(\Gamma), K \in Ker(\mathcal{G}_D(\Gamma))$ and formula ϕ of the language of Γ.

PROOF. An injection $kr : Mod_D(\Gamma) \to Ker(\mathcal{G}_D(\Gamma))$ is obtained as follows. Let D^+ be a model of Γ over a set D, $G = \mathcal{G}_D(\Gamma)$ be as in Definition 2.2, and

$kr(D^+) = K = \{Bd \in \mathbf{A}_D \mid D^+ \models Bd\} \cup \{\overline{P}d \in \overline{\mathbf{A}}_D \mid D^+ \not\models Pd\}$,

where $\overline{\mathbf{A}}_D$ is as in Definition 2.2.3. If $Bd \notin K$, for Bd instantiating LS Bx of some axiom (2.1), say F, then $D^+ \not\models Bd$ and $D^+ \models \neg Bd$. Since D^+ satisfies F, for some $B_i xy$ in its RS and some $c \in D^{ar(B_i d)}$, also $D^+ \models B_i dc$, i.e. $B_i dc \in K$. Since $B_i dc \in \mathbf{E}_G(Bd)$ by Definition 2.2, so $Bd \in \mathbf{E}_G^-(K)$. For any instance Pd of an undefined Px, if $Pd \notin K$, i.e. $D^+ \not\models Pd$, then $\overline{P}d \in K$ so $Pd \in \mathbf{E}_G^-(K)$. If $\overline{P}d \notin K$, then $D^+ \models Pd$, so $Pd \in K$ and $\overline{P}d \in \mathbf{E}_G^-(K)$. Hence, $\mathbf{V}_G \setminus K \subseteq \mathbf{E}_G^-(K)$.

If $Bd \in \mathbf{E}_G^-(K)$, with Bd instantiating LS Bx of some axiom F, then $B_i dc \in K$ for some $B_i dc \in \mathbf{E}_G(Bd)$, i.e. $D^+ \models B_i dc$ and $D^+ \models \neg Bd$, since $D^+ \models F$. Hence $Bd \notin K$. For Pd instantiating undefined Px, if $Pd \in \mathbf{E}_G^-(K)$, then $\overline{P}d \in K$, i.e. $D^+ \not\models Pd$, so $Pd \notin K$. If $\overline{P}d \in \mathbf{E}_G^-(K)$, then $Pd \in K$, i.e. $D^+ \models Pd$ so $\overline{P}d \notin K$. Hence, $\mathbf{E}_G^-(K) \subseteq \mathbf{V}_G \setminus K$. The two inclusions give $\mathbf{E}_G^-(K) = \mathbf{V}_G \setminus K$, so $K \in Ker(G)$.

Since for each atom Bd, $D^+ \models Bd \Leftrightarrow K \models Bd$, this equivalence extends to arbitrary formula ϕ. Obviously, if $D^+ \neq E^+$ for two models of Γ, then $kr(D^+) \neq kr(E^+)$.

An injection $md : Ker(G) \to Mod_D(\Gamma)$ is obtained as follows. Given a $K \in Ker(G)$, we define $md(K) = M$ over D by $M \models Bd$ iff $Bd \in K$, for each $Bd \in \mathbf{A}_D$. We verify that M satisfies each $F \in \Gamma$, having the form (2.1). If $M \models Bd$ then $Bd \in K$ so that $B_i dc \notin K$ for all B_i in the RS of B's equivalence (2.1), and all $c \in D^{ar(B_i d)}$. Hence, $M \not\models B_i dc$, i.e. $M \models \neg B_i dc$ for all such B_i and c. Thus, M satisfies the implication from left to right of B's equivalence (2.1). If $M \not\models Bd$ then $Bd \notin K$ and $Bd \in \mathbf{E}_G^-(K)$, since K is a kernel of G. This means that $\mathbf{E}_G(Bd) \cap K \neq \emptyset$, i.e. $B_i dc \in K$ for some $B_i xy$ in the RS of B's equivalence (2.1) and some $c \in D^{ar(B_i d)}$. Thus, M satisfies also the right to left implication of B's (2.1), and we conclude that $M \models \Gamma$.

Since for each atom Bd (also undefined by Γ), $md(K) \models Bd \Leftrightarrow K \models Bd$, this equivalence extends to all formulas. Obviously, if two kernels K_1, K_2 of G are different, then so are $md(K_1)$ and $md(K_2)$, giving different values to at least one atom Bd. Thus, md is injective. $\square$

One sees easily that kr and md are inverses of each other.

As sinks belong to graph's every kernel, Definition 2.2.3 adds 2-cycle $Pd \leftrightarrows \overline{P}d$ at all instances of undefined Px, to admit both boolean valuations of such Pd. The subgraph $\circlearrowright A'z \rightarrow Az...$, for each axiom's top predicate Az, forces kernels to include all instances of Az.

EXAMPLE 2.6

A model over $\{c, d\}$ of the formula $\forall x \exists y Pxy$ from Example 2.3 must satisfy either Pcc or Pcd, and either Pdd or Pdc. These determine exactly the kernels of $\mathcal{G}_{\{c,d\}}(\Gamma)$ given in Example 2.2. For every kernel $K \cap \{A'c, A'd\} = \emptyset$, hence $\{Ac, Ad\} \subseteq K$, forcing $\{Sc, Sd\} \cap K = \emptyset$. $Sc \notin K$ requires $\{Pcc, Pcd\} \cap K \neq \emptyset$, while $Sd \notin K$ requires $\{Pdc, Pdd\} \cap K \neq \emptyset$.

Fact 2.5, augmented by the extension from $GNF^-(\Gamma)$ to $GNF(\Gamma)$ in Definition 1.6 and by Fact 1.9, yield the following correspondence between models of any FOL^- theory Γ and kernels of graphs from $Gr(\Gamma)$. More precisely, a graph model consists of a pair (G, K) with $G \in Gr(\Gamma)$ and $K \in Ker(G)$, and the class $Mod(\Gamma)$ corresponds to $GMod(\Gamma) = \bigcup \{\{G\} \times Ker(G) \mid G \in Gr(\Gamma)\}$.

FACT 2.7

In FOL^-:

1. For an arbitrary GNF theory Γ : $Mod(\Gamma) \leftrightarrows GMod(\Gamma)$.
2. For an arbitrary theory T : $Mod(T) \leftrightarrows GMod(GNF(T))$.
3. For an arbitrary GNF theory Γ : $Mod(\Gamma) \neq \emptyset \Leftrightarrow Ker(\mathcal{G}_\omega(\Gamma)) \neq \emptyset$.

PROOF.

1. By Fact 2.5, we have injections $Mod_D(\Gamma) \leftrightarrows Ker(\mathcal{G}_D(\Gamma))$ for each set D. This gives the obvious injections $Mod(\Gamma) = \bigcup_{D \in Set} Mod_D(\Gamma) \leftrightarrows \bigcup_{D \in Set} (\{\mathcal{G}_D(\Gamma)\} \times Ker(\mathcal{G}_D(\Gamma))) = GMod(\Gamma)$.
2. By Fact 1.9, $Mod(T) \leftrightarrows Mod(GNF(T))$, so the claim follows by point 1.
3. A Γ with an uncountable model has a countable one, by Skolem–Löwenheim, so $Ker(\mathcal{G}_\omega(\Gamma)) \neq \emptyset$ by Fact 2.5. If Γ has a finite model, it also has an infinite one by a standard argument for FOL^-. Conversely, if $Ker(\mathcal{G}_\omega(\Gamma)) \neq \emptyset$, then Fact 2.5 gives a countable model of Γ.

$\square$

2.2 FOL with function symbols and equality

This section shows that graph representation from Definition 2.2 can be generalized from FOL^- to full FOL, with function symbols and equality, retaining Fact 2.7.1-2. Following sections, 2.3 and 3, can be read without absorbing the details of this section.

Definitions and facts from Section 1 remain unchanged for FOL^- with equality. The construction of $GNF^-(\Gamma)$ follows Definitions 1.3 and 1.4, with equality treated as a binary predicate.

However, introduction of terms complicates the straightforward Definition 2.2 of a theory's graph. For the first, axioms may now have a more specific form than schema (2.1), with terms instead of variables. This general form is abbreviated as

$$Btx \leftrightarrow \bigwedge_y (\neg B_1 xy \wedge ... \wedge \neg B_n xy), \tag{2.8}$$

where x in Btx may stand for variables occurring within the term t or as other arguments of B, while each $B_i xy$ can be a predicate symbol B_i applied to some terms with variables among x, y. A special case has Ba, with a constant a, in LS.

Consequently, the same predicate applied to different terms can now be defined by different formulas, while the interpretations of these terms may then coincide. For instance, given constants a, b, a unary predicate can be (partially) defined by

(1) $Pa \leftrightarrow \neg Qab$
$Pb \leftrightarrow \bigwedge_{y,z} \neg Rbyz.$

As long as a and b are interpreted as different elements of the domain, the graph may have different edges going out of vertices Pa and Pb. But if $a = b$, the mere identification of Pa with Pb in the graph will not reflect the logic, according to which $\neg Qab \leftrightarrow \bigwedge_{y,z} \neg Rbyz$. Note the difference from the single axiom

(2) $Px \leftrightarrow (x = a \wedge \neg Qab) \vee (x = b \wedge \bigwedge_{y,z} \neg Rbyz).$

Its graph is constructed as in Definition 2.2, except that also equality is needed. Unlike the two formulas in (1), it forces $Px = \mathbf{0}$ for all x distinct from a and b. To handle situations like (1) we introduce, along with terms, $\mathbf{T}_X \neq X$, also the equality predicate $eq(s, t)$, often abbreviated by eq_{st}, with the standard axioms. Vertices of graph $\mathcal{G}_D(\Gamma)$ contain all atoms $Btd \in \mathbf{A}_{\mathbf{T}_D}$, which are partitioned into two sets. *Ins* contains all atoms Btd that result from substituting some $d \in D^{ar(Bt)}$ for all variables x in the LS Btx of some axiom (2.8). For the remaining atoms, we include dual vertices $\overline{\mathbf{A}}_{\mathbf{T}_D} = \bigcup_{B \in \mathcal{L}} \overline{\mathbf{A}}_{\mathbf{T}_D, B}$, where $\overline{\mathbf{A}}_{\mathbf{T}_D, B} = \{\overline{Bs} \mid s \in \mathbf{T}_D, Bs \notin Ins\}$.

A brief explanation of the definition follows underneath.

DEFINITION 2.9
For Γ in FOL language $\mathcal{L}$ and a set D, the graph $G = \mathcal{G}_D(\Gamma)$ is given by:

1. $\mathbf{V}_G = \mathbf{A}_{\mathbf{T}_D} \cup \overline{\mathbf{A}}_{\mathbf{T}_D} \cup \{eq_{st} \mid s, t \in \mathbf{T}_D\} \cup \{\overline{eq}_{st} \mid s, t \in \mathbf{T}_D\} \cup \{\bullet_t \mid t \in \mathbf{T}_D \setminus D\} \cup Aux$, where *Aux* are auxiliary vertices used below.

2. For each pair of distinct terms, $s, t \in \mathbf{T}_D$, we form 2-cycle $eq_{st} \leftrightarrows \overline{eq}_{st}$; for each pair of distinct $a, b \in D$, we add a vertex with a loop and the edge: $\circlearrowright \bullet \to \overline{eq}(a, b) \leftrightarrows eq(a, b)$.

3. For each term $t \in \mathbf{T}_D \setminus D$ (including constants), we form first the complete digraph $C(t)$ over vertices $\mathbf{V}_{C(t)} = \{eq(t, d) \mid d \in D\}$ to which we add vertex $\bullet_t$ with a loop and an edge $\bullet_t \to eq(t, d)$, for each $eq(t, d) \in \mathbf{V}_{C(t)}$, schematically: $\circlearrowright \bullet_t \cdots\Rrightarrow C(t)$.

4. We add the standard equality axioms for $eq(\_, \_)$, i.e. for all distinct $p, q, r \in \mathbf{T}_D$:
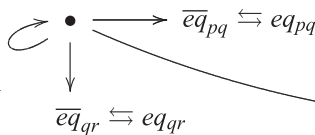(r) vertex eq_{pp} is a sink – for reflexivity,

(s) the subgraph
$$eq_{pq} \xleftarrow{} \overset{\circlearrowright}{\bullet} \xrightarrow{} \overline{eq}_{qp}$$
$$\Updownarrow \qquad\qquad\qquad \Updownarrow$$
$$\overline{eq}_{pq} \xleftarrow{} \bullet \xrightarrow{} eq_{qp}$$
– for symmetry,

(t) the subgraph
$$\overset{\circlearrowright}{\bullet} \xrightarrow{} \overline{eq}_{pq} \leftrightarrows eq_{pq}$$
$$\downarrow \qquad\qquad\qquad\qquad \searrow$$
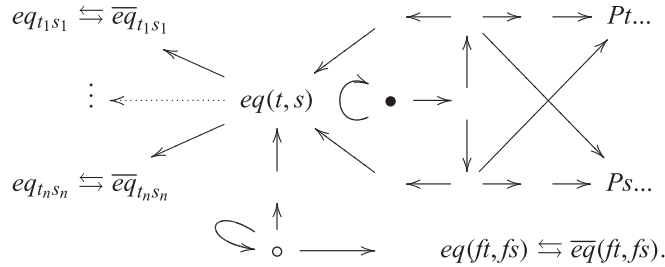$$\overline{eq}_{qr} \leftrightarrows eq_{qr} \qquad\qquad eq_{pr} \leftrightarrows \overline{eq}_{pr}$$
– for transitivity.

(c) for each function f/predicate P with arity n and pairs of terms $t_i, s_i \in \mathbf{T}_D$, $1 \le i \le n$, with

$(t_1...t_n) = t \neq s = (s_1...s_n)$, we add the congruence subgraph with the sources $\circ/\bullet$:

If t, s are single terms, edges going out of $eq(t, s)$ can be replaced by the 2-cycle to $\overline{eq}(t, s)$.

5. For atomic axioms $s = t$ or $s \neq t$ in Γ, with $s, t \in \mathbf{T}_X$, we augment each instance $sd, td \in \mathbf{T}_D$ of the 2-cycle from point 2 with a new vertex $\bullet$ with the loop and the edge:

   (a) for each $s = t \in \Gamma, d \in D^{ar(s,t)}$: $\circlearrowright \bullet \to eq(sd, td) \leftrightarrows \overline{eq}(sd, td)$;

   (b) for each $s \neq t \in \Gamma, d \in D^{ar(s,t)}$: $eq(sd, td) \leftrightarrows \overline{eq}(sd, td) \leftarrow \bullet \, \circlearrowleft$.

6. For each axiom (2.8) with Btx in LS and for each $d \in D^{ar(Bt)}$, vertex Btd obtains the outgoing edges to $\mathbf{E}_G(Btd) = \{B_i dc \mid 1 \leq i \leq n, c \in D^{ar(B_i d)}\}$.

7. For each $\overline{B}s \in \overline{\mathbf{A}}_{\mathbf{T}_D}$, vertex Bs obtains the 2-cycle $Bs \leftrightarrows \overline{B}s$.

Auxiliary vertices *Aux* are all $\bullet$ and anonymous vertices in the indicated subgraphs. For each kernel $K \in Ker(G)$, the subgraphs in respective points above ensure the following properties:

2. For distinct $a, b \in D$, $eq(a, b) \notin K$, representing inequality.

3. Unique interpretation in D of every function application. With eq representing equality, these subgraphs ensure that each application of a function to arguments from D returns a unique element of D, in particular, that each constant is interpreted as some unique $d \in D$. This follows because, in a complete graph, the kernels are exactly individual vertices, so that each kernel of $C(t)$ is exactly one $eq(t, d)$.

4. Satisfaction of the standard equality axioms by $eq(\_, \_)$. Equivalence is ensured by (r), (s) and (t), while in the subgraphs (c), vertex $\circ$ captures the congruence axiom $t = s \to ft = fs$, and vertex $\bullet$ its predicate version $t = s \to (Pt \leftrightarrow Ps)$. Vertices Pt, Ps initiate the subgraphs according to point 6.

5. Satisfaction of the atomic nonlogical (in)equality axioms.

6. Satisfaction of other nonlogical axioms (2.8).

7. If a predicate B is only partially defined (like in (1)), then each $Bd \notin Ins$ can be interpreted arbitrarily, provided $eq(d, tc) \notin K$ for each $Btc \in Ins$ (and the chosen interpretation does not collide with other restrictions).

Saying below that something follows 'by subgraphs...' refers to the points above, applied to any kernel restricted to these subgraphs. For instance, if $K \in Ker(\mathcal{G}_D(\Gamma))$ and $s = t \in \Gamma$ then, by subgraphs 5.(a), $eq(sd, td) \in K$ for each $d \in D^{ar(s,t)}$, because kernel requires exclusion of vertex $\bullet$ with the loop, which forces $eq(sd, td) \in K$.

EXAMPLE 2.10
For a predicate P, if there is no equivalence with Px in LS, then 7.(b) yields the 2-cycle $Pd \leftrightarrows \overline{P}d$, for each $d \in D^{ar(P)}$. For example (1) from the beginning of this subsection, point 6 yields the

edges

- $Pa \rightarrow Qab...$
- $Pb \rightarrow Rbd_1d_2...$ for all $d_1, d_2 \in D^2$.

By subgraphs 2.9.4.(c) with Pa, Pb and Ps, $s \in D^{ar(P)}$, if $K \in Ker(G)$ and $eq(a,s) \in K$ then $K \models Pa \leftrightarrow Ps$, while if $eq(a,b) \in K$ then also $K \models \neg Qaa \leftrightarrow \bigwedge_{y,z} \neg Rayz$. But if $\{\overline{eq}(d,a), \overline{eq}(d,b)\} \subseteq K$, then by subgraphs 7.(b), either $Pd \in K$ or $\overline{P}d \in K$.

EXAMPLE 2.11

For the axiom (P) $Px \leftrightarrow \neg Psx$, and the set with one element 0, terms $\mathbf{T}_{\{0\}}$ are (isomorphic to) natural numbers and the graph becomes a ray $P0 \rightarrow Ps0 \rightarrow Pss0 \rightarrow ...$, with subgraphs 2.9.4.(c) for each pair s^n0, s^m0, $n \neq m$. Since all these terms are interpreted identically over $\{0\}$, for any kernel $K \in Ker(\mathcal{G}_{\{0\}}(P))$, the subgraph 2.9.4.(c) with the source $\bullet$ for $P0, Ps0$ is such that $eq(0, s0) \in K$, while the edge $P0 \rightarrow Ps0$ yields $P0 \in K \Leftrightarrow Ps0 \notin K$:

$$eq(0, s0) = 1$$
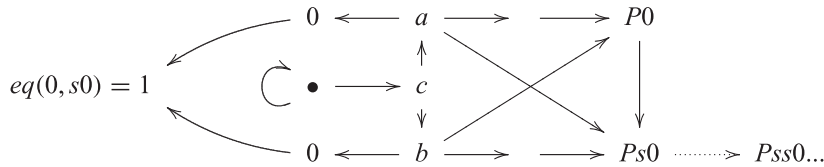
Consequently, this subgraph, and hence the whole graph, has no kernel, reflecting the nonexistence of models of (P) over one element domain.

Taking as the underlying set the natural numbers $\mathbb{N}$, with the standard interpretation of s as $+1$, the graph again becomes the ray $P0 \rightarrow P1 \rightarrow P2 \rightarrow ....$ But now no equality $eq(p, q)$ holds except for $eq(p, p)$. The instance of the subgraph above swaps $\mathbf{0}$ and $\mathbf{1}$, obtaining two kernels, with Ps^n0 for all even $n \geq 0$, or for all odd $n > 0$.

The following extends Fact 2.5 to FOL, showing that $\mathcal{G}_D(\Gamma)$ captures all models of Γ over a set D.

FACT 2.12 (2.5)

For a GNF theory Γ in FOL and any nonempty set D, there are injections

$kr : Mod_D(\Gamma) \rightarrow Ker(\mathcal{G}_D(\Gamma))$ and $md : Ker(\mathcal{G}_D(\Gamma)) \rightarrow Mod_D(\Gamma)$,

such that $M \models \phi \Leftrightarrow kr(M) \models \phi$ and $K \models \phi \Leftrightarrow md(K) \models \phi$, for each $M \in Mod_D(\Gamma), K \in Ker(\mathcal{G}_D(\Gamma))$ and formula ϕ of the language of Γ.

PROOF. The proof has the structure of the proof of Fact 2.5, with the additional treatment of equality, and the more general form (2.8) of axioms. **(kr)** Letting $G = \mathcal{G}_D(\Gamma)$, an injection $kr : Mod_D(\Gamma) \rightarrow Ker(G)$ is obtained by mapping a model $D^+ \models \Gamma$, over a set D, on $kr(D^+) \in Ker(G)$ given by all values induced from:

$$K = \{Bd \in \mathbf{A}_{\mathbf{T}_D} \mid D^+ \models Bd\} \cup \{\overline{B}d \in \overline{\mathbf{A}}_{\mathbf{T}_D} \mid D^+ \not\models Bd\}$$

$$\cup \{\overline{eq}(s,t) \mid D^+ \not\models s = t, \; s, t \in \mathbf{T}_D\} \cup \{eq(s,t) \mid D^+ \models s = t, \; s, t \in \mathbf{T}_D\}.$$

The second summand is empty in case $\overline{\mathbf{A}}_D = \emptyset$ and each Bd obtains a value relatively to its outneighbours $\mathbf{E}_G(Bd)$ determined by the axioms.

Vertices included into $kr(D^+)$ by inducing from K, but not mentioned in the definition of K above, are among the auxiliary vertices in graphs 2.9.4.(c). They do not affect the argument below, so we identify $kr(D^+) = K$.

i. Equality in D^+ is reflected by *eq* in G. Since each term applied to elements of D yields a unique element of D, K determines a unique solution to each subgraph from point 3 of Definition 2.9 with $eq(t,d) = \mathbf{1}$ for $d \in D$ interpreting the term $t \in \mathbf{T}_D$, i.e. $D^+ \models t = d$, which induces $\mathbf{0}$ to $\bullet_t$. The last two summands of K determine unique solutions to all subgraphs from Definition 2.9.4:

(r), since $eq(p,p)$ is a sink,

(s), since $eq(p,q) \Leftrightarrow eq(q,p)$ both $\bullet$ vertices in subgraph (s) obtain induced value $\mathbf{0}$ and

(t), since $eq(p,g)$ and $eq(q,r)$ imply $eq(p,r)$, so $\bullet$ in subgraph (t) obtains induced value $\mathbf{0}$.

For (c), if $eq(t,s) \in K$, i.e. $D^+ \models t = s$, then $D^+ \models ft = fs$ for each function f, and $D^+ \models Bte \leftrightarrow Bse$ for each predicate B and $e \in D^{ar(Bt)} = D^{ar(Bs)}$. Thus, if $eq(t,s) \in K$ then $eq(ft,fs) \in K$, while $Bte \in K \Leftrightarrow Bse \in K$, and the subgraph from 4.(c) obtains a solution since inducing from these values ensures $\mathbf{E}_G(\circ) \cap K \neq \emptyset$ and $\mathbf{E}_G(\bullet) \cap K \neq \emptyset$.

The last two summands of K give also unique solutions to the subgraphs from 2.9.5.

ii. Instances $Bd \in Ins$ defined by (2.8) are treated as in the proof of Fact 2.5, with a small proviso. It may happen that $Bd \notin Ins$, while for some axiom with Btx in LS, $D^+ \models d = tc$, so $D^+ \models Bd \Leftrightarrow D^+ \models Btc$. Then $eq(d,tc) \in K$ and $Bd \in K \Leftrightarrow Btc \in K$, so K solves the subgraph 2.9.4.(c) with Bd, Btc in place of Ps, Pt.

We show first $\mathbf{V}_G \setminus K \subseteq \mathbf{E}_G^-(K)$. This inclusion follows for vertices $eq(\_,\_)$ and $\overline{eq}(\_,\_)$ by **i** (and for *Aux* by inducing from K), so we consider atoms $A \in \mathbf{A}_{\mathbf{T}_D}$ with $A \notin K$, i.e. $D^+ \not\models A$.

Let $A = Btc \in Ins$, i.e. Btc is an instance of LS Btx of some axiom (2.8) F. Since $D^+ \not\models Btc$, so $D^+ \models \neg Btc$ and $D^+ \models B_i ce$ for some instance of some conjunct $B_i xy$ in RS of F. Hence $B_i ce \in K$, and $Btc \in \mathbf{E}_G^-(K)$ since $B_i ce \in \mathbf{E}_G(Btc)$ by 2.9.6.

If $A = Bd \notin Ins$, then $\mathbf{E}_G(Bd) = \{\overline{B}d\}$ by 2.9.7. Since $D^+ \not\models Bd$, so $\overline{B}d \in K$ and $Bd \in \mathbf{E}_G^-(K)$. These two cases, with point **i**, establish $\mathbf{V}_G \setminus K \subseteq \mathbf{E}_G^-(K)$.

iii. For the opposite inclusion, assuming $A \in \mathbf{E}_G^-(K)$, there are two cases.

If $A = Btc \in Ins$ then $Btc \in \mathbf{E}_G^-(K)$ means that $B_i ce \in K$ for some instance $B_i ce$ of some $B_i xy$ in the RS of the axiom $Btx \leftrightarrow ...B_i xy$. Then $D^+ \models B_i ce$, so $D^+ \not\models Btc$ and $Btc \notin K$.

If $A = Bd \notin Ins$ then $\mathbf{E}_G(Bd) = \{\overline{B}d\}$ by 2.9.7 and since $Bd \in \mathbf{E}_G^-(K)$, so $\overline{B}d \in K$, which means that $D^+ \not\models Bd$, so that $Bd \notin K$.

These two cases give $\mathbf{A}_{\mathbf{T}_D} \cap \mathbf{E}_G^-(K) \subseteq \mathbf{V}_G \setminus K$. If $v \in \mathbf{E}_G^-(K)$ is $eq(\_,\_)$ or $\overline{eq}(\_,\_)$, then partial solutions to the subgraphs in point **i** give $v \in \mathbf{V}_G \setminus K$, so that $\mathbf{E}_G^-(K) \subseteq \mathbf{V}_G \setminus K$. With **ii**, this gives $\mathbf{E}_G^-(K) = \mathbf{V}_G \setminus K$, i.e. $K \in Ker(G)$.

iv. Since $D^+ \models Bd \Leftrightarrow K \models Bd$ for each atom $Bd \in \mathbf{A}_{\mathbf{T}_D}$, and $D^+ \models s = t \Leftrightarrow K \models eq(s,t)$ for $s, t \in \mathbf{T}_D$, the equivalence $D^+ \models \phi \Leftrightarrow K \models \phi$ holds for arbitrary formula ϕ of the language of Γ.

v. The so defined *kr* is injective, because two different models $D_1^+ \neq D_2^+$, over any given set D, obviously give two different $kr(D_1^+) \neq kr(D_2^+)$, since by definition $K_1 \neq K_2$.
(md) An injection $md : Ker(G) \rightarrow Mod_D(\Gamma)$ is obtained as follows.

i. Given a $K \in Ker(G)$, we note first that points 2 and 3 of Definition 2.9 ensure well-definedness of the function $i : \mathbf{T}_D \rightarrow D$, given by $i(t) = d$, for $d \in D$ such that $eq(t,d) \in K$. By 2.9.3 such a d is unique for each $t \in \mathbf{T}_D$, while by 2.9.2, the restriction $i|_D$ is the identity on D.

Subgraphs (r), (s) and (t) from 2.9.4 ensure that $eq(\_,\_)$ is an equivalence on D, while (c) that $eq(s,t)$ entails $eq(fs,ft)$ for each function symbol f. Hence, i is a well-defined quotient mapping.

ii. We define the structure $M = md(K)$ on D by interpreting all function symbols using i. For each function f and $a \in D^{ar(f)}$, define its interpretation in M by $f^M a = i(fa) = d \in D$, where $\{eq(fa, d)\} = \mathbf{V}_{C(fa)} \cap K$, for the subgraph $C(fa)$ from 2.9.3.

For each atom $Bd \in \mathbf{A}_{\mathbf{T}_D}$, we define $M \models Bd$ iff $Bd \in K$. This is well-defined, since subgraphs 2.9.4.(c) ensure that for all $t, s \in \mathbf{T}_D$, if $eq(t, s) \in K$ then $Bt \in K \Leftrightarrow Bs \in K$. Suppose that Γ contains two distinct axioms (2.8), $Btx \leftrightarrow ...$ and $Bsy \leftrightarrow ...$, while $eq(ta, d) \in K$ and $eq(sb, d) \in K$ for some $a \in \mathbf{T}_D^{ar(t)}$, $b \in \mathbf{T}_D^{ar(s)}$. Then $eq(ta, sb) \in K$ by the subgraphs (s), (t) from Definition 2.9.3, and $Bta \in K \Leftrightarrow Bsb \in K$ by (c). Hence, $Bta \in K \Leftrightarrow Bsb \in K \Leftrightarrow Bd \in K$, so $M \models Bta \Leftrightarrow M \models Bsb \Leftrightarrow M \models Bd$.

iii. For any atomic equality axiom $s = t \in \Gamma$, subgraph $\curvearrowright \bullet \to eq(sd, td)$ from 2.9.5 forces $eq(sd, td) \in K$ for every instance $sd, td \in \mathbf{T}_D$ of s, t. By **i**, for each such instance $i(sd) = i(td)$, so that $M \models s = t$. Similarly, for the axiom $s \neq t \in \Gamma$, the subgraph $eq(sd, td) \leftrightarrows \overline{eq}(sd, td) \leftarrow \bullet \;\circlearrowright$, for each instance $sd, td \in \mathbf{T}_D$, forces $eq(sd, td) \notin K$, so $i(sd) \neq i(td)$, giving $M \models s \neq t$.

iv. Consider now any axiom (2.8) $F \in \Gamma$, with Btx in LS, and any $d \in D^{ar(Bt)}$. If $M \models Btd$, i.e. $Btd \in K$, then $\mathbf{E}_G(Btd) \cap K = \emptyset$ since $K \in Ker(G)$, so that $B_i dc \notin K$ for all $B_i xy$ in the RS of F and $c \in D^{ar(B_i d)}$. Thus $M \not\models B_i dc$, i.e. $M \models \neg B_i dc$ for all $B_i dc$ in the RS, so M satisfies the implication from left to right of F.

If $M \not\models Btd$ then $Btd \notin K$ and $Btd \in \mathbf{E}_G^-(K)$, since $K \in Ker(G)$. This means that for some $B_i xy$ in the right side of F and $c \in D^{ar(B_i d)}$, $B_i dc \in K$, since such $B_i dc$ form $\mathbf{E}_G(Btd)$ by point 6 of Definition 2.9. Thus, M satisfies also the right to left implication of (2.8), so $M \models F$.

v. Since $M \models Bd \Leftrightarrow K \models Bd$ for each atom Bd and, by **ii**, $M \models s = t \Leftrightarrow K \models eq(s, t)$ for each $s, t \in \mathbf{T}_D$, the equivalence $M \models \phi \Leftrightarrow K \models \phi$ holds for arbitrary formula ϕ of the language of Γ.

vi. Two different kernels $K_1 \neq K_2$ of G differ for at least one atom Bd. This follows because membership of atoms in a kernel K determines uniquely this kernel, as can be seen inspecting the subgraphs in Definition 2.9. For instance, restriction of any kernel K to the atoms in any subgraph 2.9.4.(c), i.e. to Pt, Ps and all $eq(t_i, s_i)$, induces unique values to all remaining (auxiliary) vertices of this subgraph, which therefore must coincide with their (non)membership in K. By Definition 2.9.7, each $K \in Ker(G)$ determines $Bd \in K$ or $\overline{B}d \in K$ also for $Bd \notin Ins$.

Thus, $K_1 \neq K_2$ implies for some atom $Bd \in (K_1 \setminus K_2) \cup (K_2 \setminus K_1)$. The respective models differ then on Bd by **ii**, $md(K_1) \neq md(K_2)$, so md is injective. $\qquad\square$

Consequently, kernels represent exactly models of a theory: Fact 2.13 below follows from Fact 2.12 in the same way Fact 2.7.1–2 follows from Fact 2.5.

FACT 2.13 (2.7)
In FOL:

1. For an arbitrary GNF theory $\Gamma : Mod(\Gamma) \leftrightarrows GMod(\Gamma)$.
2. For an arbitrary theory $T : Mod(T) \leftrightarrows GMod(GNF(T))$.

2.3 Some facts about kernels

We gather some relevant facts about kernels. Since existence of a kernel for some graph in $Gr(\Gamma)$ is equivalent to consistency of Γ, we start by quoting a couple of results on kernel existence. A graph is *kernel perfect* if each induced subgraph has a kernel. Often, establishing solvability, one shows

actually kernel perfectness and we, too, will use this stronger notion. The central result in kernel theory is the following theorem of Richardson.

THEOREM 2.14 ([3]).
A graph G without odd cycles is kernel perfect if (a) for each $x \in \mathbf{V}_G : \mathbf{E}_G(x)$ is finite or (b) there are no rays (infinite, simple, outgoing paths).

In particular, a finite graph without odd cycles is kernel perfect. We will also encounter the following notion and fact. A digraph G is *bipartite* if so is its underlying undirected graph (forgetting directions of edges), that is, if $\mathbf{V}_G$ can be partitioned into two independent subsets, so that each $\mathbf{E}_G$-edge connects a vertex in one subset to a vertex in the other.

FACT 2.15 ([6]).
A bipartite graph is kernel perfect.

Kernels can be transferred from homomorphic images to preimages. A *graph homomorphism* from G to H is a function $h : \mathbf{V}_G \to \mathbf{V}_H$ such that

$$\forall x \in \mathbf{V}_G : h(\mathbf{E}_G(x)) = \mathbf{E}_H(h(x)), \tag{2.16}$$

where h is extended pointwise to subsets, i.e. for $X \subseteq \mathbf{V}_G : h(X) = \{h(x) \mid x \in X\}$.

FACT 2.17
For a homomorphism $h : G \to H$, if $K \in Ker(H)$ then $K^- \in Ker(G)$, where $K^- = h^-(K) = \{y \in \mathbf{V}_G \mid h(y) \in K\}$.

PROOF. When K is independent then so is K^-, because $h(\mathbf{E}_G(x)) \subseteq \mathbf{E}_H(h(x))$ by (2.16).

For any $y \in \mathbf{V}_G \setminus K^-$, i.e. $h(y) \notin K$, there is an $h \in \mathbf{E}_H(h(y)) \cap K$, since $\mathbf{V}_H \setminus K \subseteq \mathbf{E}_H^-(K)$. By (2.16) $h(\mathbf{E}_G(y)) \supseteq \mathbf{E}_H(h(y))$, so $h \in h(\mathbf{E}_G(y)) \cap K$, hence for some $g \in \mathbf{E}_G(y) : h(g) = h$. Since $h \in K$, so $g \in K^-$ and $y \in \mathbf{E}_G^-(K^-)$. Thus, $\mathbf{V}_G \setminus K^- \subseteq \mathbf{E}_G^-(K^-)$ and $K^- \in Ker(G)$. $\square$

An isomorphism of graphs G and H, denoted by $G \simeq H$, is a bijective homomorphism either way.

FACT 2.18
For every GNF theory Γ in FOL$^-$ and two sets D, E, if there is a surjection (bijection) $D \to E$, then there is a surjective (bijective) homomorphism $\mathcal{G}_D(\Gamma) \to \mathcal{G}_E(\Gamma)$.

PROOF. A surjection (bijection) $\beta' : D \to E$ gives a surjection (bijection) on atoms $\beta : \mathbf{A}_D \to \mathbf{A}_E$ by $\beta(B(d)) = B(\beta'(d))$. We verify that it is a homomorphism. For $X \in \{D, E\}$, replace indices $\mathcal{G}_X(\Gamma)$ by X, e.g. $\mathbf{V}_X = \mathbf{V}_{\mathcal{G}_X(\Gamma)}$, etc. For each $B(d) \in \mathbf{V}_D$

$\mathbf{E}_E(\beta(B(d))) = \mathbf{E}_E(B(\beta'(d))) = \{B_i(\beta'(d), e) \mid 1 \le i \le n, e \in E^{k_i}\}$, for $k_i = ar(B_i(\beta'(d)))$.
On the other hand, $\mathbf{E}_D(B(d)) = \{B_i(d, c) \mid 1 \le i \le n, c \in D^{k_i}\}$, so that

$\beta(\mathbf{E}_D(B(d))) = \{\beta(B_i(d, c)) \mid 1 \le i \le n, c \in D^{k_i}\} = \{B_i(\beta'(d), \beta'(c)) \mid 1 \le i \le n, c \in D^{k_i}\}$.
Since $\beta' : D \to E$ is surjective (bijective), it gives a surjection (bijection) $D^{k_i} \to E^{k_i}$ for every $k_i \in \omega$, so that $\beta(D^{k_i}) = E^{k_i}$ and $\beta(\mathbf{E}_D(B(d))) = \mathbf{E}_E(\beta(B(d)))$. $\square$

Hence, a bijection $D \leftrightarrows E$ gives an isomorphism $\mathcal{G}_D(\Gamma) \simeq \mathcal{G}_E(\Gamma)$, also for Γ in FOL. We can therefore identify graphs in $Gr(\Gamma)$ by their domain's cardinality and set $Gr(\Gamma) = \{\mathcal{G}_\kappa(\Gamma) \mid \kappa > 0\}$, excluding the empty graph $\mathcal{G}_0(\Gamma)$. The graph in Example 2.3 is thus $\mathcal{G}_2(A)$. By Fact 2.14, $\mathcal{G}_\kappa(\Gamma)$ captures up to isomorphism all models of Γ in FOL over domains with cardinality κ. Incidentally,

a simple version of upward Skolem–Löwenheim theorem follows for FOL$^-$: if Γ in FOL$^-$ has a model of cardinality κ, then for any $\lambda > \kappa$, surjection $\lambda \to \kappa$ gives by Fact 2.18 a homomorphism $\mathcal{G}_\lambda(\Gamma) \to \mathcal{G}_\kappa(\Gamma)$, reflecting latter's kernels by Fact 2.17 and yielding a model over λ by Fact 2.5.

Homomorphisms reflect also bipartitions, according to the (proof of the) following fact.

FACT 2.19
If $h : G \to H$ is a homomorphism and H is bipartite, then so is G.

PROOF. We show that if $\langle A_1, A_2 \rangle$ is a bipartition of H, then $\langle A_1^-, A_2^- \rangle$ is a bipartition of G, where $A_i^- = \{x \in \mathbf{V}_G \mid h(x) \in A_i\}$. Let $\{i,j\} = \{1, 2\}$. If $h(x) = y \in A_i$ then $\mathbf{E}_H(y) \cap A_i = \emptyset$, so there is no $z \in \mathbf{E}_G(x) \cap A_i^-$, since otherwise $h(z) \in h(\mathbf{E}_G(x) \cap A_i^-) \subseteq h(\mathbf{E}_G(x)) \cap h(A_i^-) = \mathbf{E}_H(y) \cap A_i$. So $\mathbf{E}_G(A_i^-) \subseteq A_j^-$, and dually, $\mathbf{E}_G(A_j^-) \subseteq A_i^-$, i.e. $\langle A_i^-, A_j^- \rangle$ is a bipartition of G. □

Finally, we sometimes start with a partial assignment of boolean values (select a part of a kernel) and propagate its consequences, that is, induce values to some other vertices. Briefly, a vertex must be assigned **0** if it has an edge to a vertex assigned **1**, while if all outneighbours are assigned **0**, the vertex itself must be **1**. More formally, given a partial assignment σ with domain $X \subseteq \mathbf{V}_G$, we start with $V_0 = \mathbf{V}_G$, set

$\sigma_0^\mathbf{1} = \{x \in X \mid \sigma(x) = \mathbf{1}\} \cup (sinks(G) \setminus X)$ and $\sigma_0^\mathbf{0} = \{x \in X \mid \sigma(x) = \mathbf{0}\} \cup (\mathbf{E}_G^-(\sigma_0^\mathbf{1}) \setminus X)$, and iterate the following:

$$V_{i+1} = V_i \setminus (\sigma_i^\mathbf{1} \cup \sigma_i^\mathbf{0}), \text{ and in limits } \lambda : V_\lambda = \bigcap_{i < \lambda} V_i,$$

$$G_j \text{ is the subgraph of } G \text{ induced by } V_j \tag{2.20}$$

$$\sigma_{i+1}^\mathbf{1} = sinks(G_{i+1}) \text{ and } \sigma_{i+1}^\mathbf{0} = \mathbf{E}_G^-(\sigma_{i+1}^\mathbf{1}) \cap V_{i+1}.$$

We stop when $V_{\kappa+1} = V_\kappa$, obtaining the induced $(\bigcup_{i \le \kappa} \sigma_i^\mathbf{0} \times \{\mathbf{0}\}) \cup (\bigcup_{i \le \kappa} \sigma_i^\mathbf{1} \times \{\mathbf{1}\})$. Depending on σ, this may not be a function, but we induce only when it is. More details can be found in [6].

3 Extensions

Transforming manually a given theory into GNF may be a cumbersome task worthwhile only in special cases. However, GNF is often encountered directly, as in definitional extensions or fixed point definitions, where a new predicate B is introduced by a form $Bx \leftrightarrow RS$, from which GNF can easily be obtained, transforming RS. Often, extension should be conservative or allow a (unique) expansion of every model of the theory under extension, and various syntactic restrictions are utilized to ensure this. Graph representation gives a new perspective on such situations, in particular, when circularity is involved.

We therefore view now extensions as the primary objects. An extension is simply a GNF theory Δ, with undefined predicates (if any) marking connections to any theory Γ which it may extend. This is how extensions are often used: as a generic possibility of augmenting a wide range of theories. Just like transitive closure can be applied to any binary relation, we can think of an extension as potentially applicable to any theory possessing predicates with the arities of the undefined predicates of the extension.

Given Δ as an independent object, its undefined predicates may need renaming to match the appropriate predicates of Γ, avoiding such an identification of predicates defined in Δ. Ensuring

this, when applying Δ to Γ, is straightforward, so we assume the naming details are always resolved and write the result of such an extension as $\Delta(\Gamma)$.[3] For every (set of) cardinality κ, the graph equality $\mathcal{G}_\kappa(\Delta(\Gamma)) = \mathcal{G}_\kappa(\Delta) \cup \mathcal{G}_\kappa(\Gamma)$ holds because of this assumption, so that each sink of $\mathcal{G}_\kappa(\Delta)$, representing an undefined predicate application Pd, is identified with vertex Pd of $\mathcal{G}_\kappa(\Gamma)$ (or, if no Pd occurs in $\mathcal{G}_\kappa(\Gamma)$, acquires 2-cycle in $\mathcal{G}_\kappa(\Delta(\Gamma))$). Whenever the choice of κ is inessential, we speak about the graph $\mathcal{G}(\Delta)$ of the extension and the graph $\mathcal{G}(\Gamma)$ of the extended theory.

Consequently, while sinks for undefined atoms of $\mathcal{G}(\Delta)$ end up among vertices of $\mathcal{G}(\Gamma)$, there are no edges from $\mathcal{G}(\Gamma)$ to $\mathcal{G}(\Delta)$—the theory being extended does not use any predicates defined by the extension.

Sinks of $\mathcal{G}(\Delta)$ obtain thus a dual status. Some may represent *LS*s of axioms with empty *RS*s, which are simply true. Belonging to every kernel, they affect the graph in a unique way, inducing some consequences (following (2.3), with $\sigma_0^1 = sinks$). Such sinks must be distinguished from those which represent undefined atoms, to be identified with identical atoms of the extended theory. We call the latter *u-sinks*. Kernels of $\mathcal{G}(\Delta)$ are to be investigated under arbitrary valuations of u-sinks, as their values are (to be) determined by the theory which is being extended. Formally, this comes closer to equipping them with 2-cycles, but name 'u-sinks' marks that the question concerns now existence of a kernel under arbitrary—and not only some appropriate—valuation of these atoms. Extension Δ in most examples below contains no atomic axioms (sinks) and only undefined atoms (u-sinks) matched by atoms of Γ.

Definitional extension implies conservativity, but we distinguish also two other intermediary notions. Given languages $\mathcal{L} \subseteq \mathcal{L}^+$, an $\mathcal{L}^+$ structure M^+ is an *expansion of* $\mathcal{L}$-structure M, if the $\mathcal{L}$-reduct of M^+ equals M, i.e. $M^+|_{\mathcal{L}} = M$. Recall also that an *explicit definition*, of a predicate $B \notin \mathcal{L}$ in a language $\mathcal{L}$, is $Bx \leftrightarrow \phi$, where ϕ is an $\mathcal{L}$-formula with free variables $\mathcal{V}(\phi) \subseteq x$.

DEFINITION 3.1
Let $\Gamma \subseteq \Gamma^+$ be theories over languages $\mathcal{L} \subseteq \mathcal{L}^+$. $\Delta = \Gamma^+ \setminus \Gamma$ is a

- *definitional extension of* Γ, according to definition before Fact 1.5,
- *model unique extension of* Γ if each model of Γ has a unique expansion to a model of Γ^+,
- *model extension of* Γ if each model of Γ has an expansion to a model of Γ^+,
- *conservative extension of* Γ if for every $\mathcal{L}$-sentence $\phi : \Gamma \vdash \phi \Leftrightarrow \Gamma^+ \vdash \phi$.

The notions are listed with decreasing strength: every definitional extension is a model unique extension, which is a model extension, and every model extension is conservative. In general, none of these inclusions can be reversed.

We begin with a few simple examples. An extension with a predicate B, even if not in GNF, has typically the form $Bx \leftrightarrow RS$. An equivalent extension in GNF can be then obtained more easily, than by following Definitions 1.3 and 1.4, by reformulating appropriately RS. This is typically done below.

EXAMPLE 3.2
Definitional extension Δ given by $Dx \leftrightarrow Fx \vee \neg Hx$, as written in GNF to the left, has graph $\mathcal{G}(\Delta)$ to the right:

---

[3]Strictly speaking, application $\Delta(\Gamma)$ is relative to a renaming τ of Δ's undefined predicate symbols to predicate symbols (of the same arities) of Γ, and of Δ's defined symbols to symbols not occurring in Γ. The result is $\Delta(\Gamma, \tau) = \tau(\Delta) \cup \Gamma$, but we drop τ which is identity in the examples. Every theory, written in GNF, is thus an extension. If it has no undefined predicates matching (under τ) predicates of the actual argument theory Γ, the application yields disjoint union of the two.

$$\Delta: \quad \begin{aligned} Dx &\leftrightarrow \neg Bx \\ Bx &\leftrightarrow \neg\overline{H}x \wedge \neg Fx \\ \overline{H}x &\leftrightarrow \neg Hx \end{aligned}$$

An application of Δ to any actual theory Γ amounts to matching F and H to some unary predicates of Γ. For any κ, the graph $\mathcal{G}_\kappa(\Delta)$ has a copy of the above $\mathcal{G}(\Delta)$ for each $x \in \kappa$. The u-sinks Fx and Hx of $\mathcal{G}_\kappa(\Delta)$ obtain in $\mathcal{G}_\kappa(\Delta(\Gamma))$ the edges which Fx, Hx have in $\mathcal{G}_\kappa(\Gamma)$.

EXAMPLE 3.3

The extension $Nx \leftrightarrow Fx \vee (Hx \wedge \neg Nx)$, i.e. $Nx \leftrightarrow \neg(\neg Fx \wedge \neg(Hx \wedge \neg Nx))$, given in GNF to the left, has graph $\mathcal{G}(N)$ to the right:

$$(N) \quad \begin{aligned} Nx &\leftrightarrow \neg Bx \\ Bx &\leftrightarrow \neg Fx \wedge \neg Cx \\ Cx &\leftrightarrow \neg\overline{H}x \wedge \neg Nx \\ \overline{H}x &\leftrightarrow \neg Hx \end{aligned}$$

The graph has no solution for $Fx = \mathbf{0}, Hx = \mathbf{1}$, so (N) does not model extend any Γ consistent with $\exists x(\neg Fx \wedge Hx)$. From (N), we can actually prove $Fx \vee \neg Hx$, so this extension is not even conservative for any such Γ.[4]

EXAMPLE 3.4

Let (N') be as (N) in Example 3.3, but with Bx replaced by the following $B'x$:

$$(N') \quad B'x \leftrightarrow \neg Fx \wedge \neg Cx \wedge \neg Nx$$

The new edge $B'x \to Nx$ makes the induced subgraph $\{B'x, Cx, Nx\}$ kernel perfect. Consequently, it is solvable for every valuation of its u-sinks, so (N') is a model extension of every theory.

The following more complex example illustrates also the effects of quantifiers.

3.1 Transitive closure

Given a binary relation E, a natural attempt to define its transitive closure is by adding the axiom

(TC) $\quad TCxy \leftrightarrow Exy \vee \exists z : Exz \wedge TCzy$.

We ask first about the general relation between possible models of E and models of E extended with (TC). In GNF, the definition of TC becomes the four equivalences in (3.5). The dotted edges, marked

---

[4]In chapter 4.I of [2], viewing (N) as a 'definition', the authors remark that it 'allows us to prove *a priori* that all H's are F's.'. Well, all H's are F's *assuming* (N), but while a definition may seem *a priori*, there is nothing *a priori* about an extension with nonlogical axioms. Since every theory can be written in the GNF format, typical if not required for definitions, there seems little reason to distinguish here between definitions and axioms.

with $\forall z$ on the sketch of $\mathcal{G}(TC)$ below, signal branching to all instances of the target formula, with some marked explicitly on the dotted edges leading to them. The edges $\overline{Exy} \to Exy$ to the u-sinks Exy of $\mathcal{G}(TC)$ are left implicit.

$$
\begin{aligned}
TCxy &\leftrightarrow \neg Bxy & Cxzy &\leftrightarrow \neg \overline{Exz} \wedge \neg \overline{TCzy} \\
Bxy &\leftrightarrow \neg Exy \wedge \bigwedge_z \neg Cxzy & \overline{Exy} &\leftrightarrow \neg Exy \ \text{ and } \ \overline{TCzy} \leftrightarrow \neg TCzy
\end{aligned}
\tag{3.5}
$$
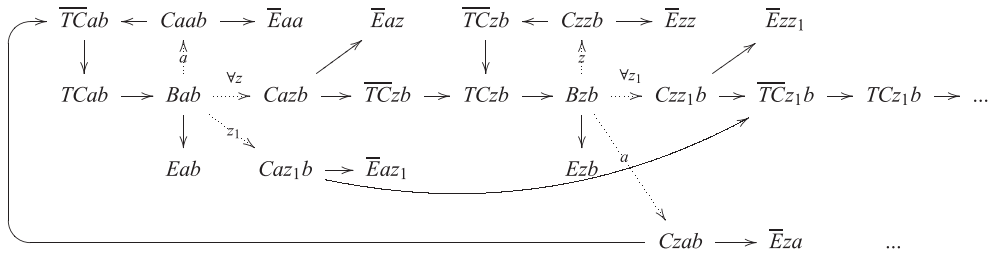
i. First, if $Eaa = \mathbf{1}$, i.e. $\overline{Eaa} = \mathbf{0}$, then we can obtain a model with $TCab = \mathbf{1}$ by simply choosing $\overline{TCab} = \mathbf{0}$, which yields $Caab = \mathbf{1}$ and hence $Bab = \mathbf{0}$, independently of the choice of assignments further down the graph. This corresponds to the fact that, when $Eaa = \mathbf{1}$, the instance $TCab \leftrightarrow Eab \vee (Eaa \wedge TCab)$ of (TC) with $z = a$, becomes trivially satisfiable by merely choosing $TCab = \mathbf{1}$.

ii. Considering irreflexive E, (TC) still does not capture transitive closure, though this is less obvious. The case $\mathcal{G}_\omega(TC)$ of $\mathcal{G}(TC)$ has namely a kernel including $TCab$ when there is an infinite chain $R = \{z_0, z_1, z_2, ...\}$ with $z_0 = a$, and $Ez_ib = \mathbf{0} = \overline{Ez_iz_{i+1}}$ for each $z_i \in R$. The last condition means $Ez_iz_{i+1} = \mathbf{1}$, i.e. R is a ray (or enters a cycle), which gives a very specific and unintended meaning to any b being E-reachable from a, as if an infinite walk reached every vertex (then TCz_ib for each $z_i \in R$). Still, in this situation there is also another model in which $TCab = \mathbf{0}$.

The aim of this example is not to reexamine undefinability of transitive closure in FOL but to note that the graph above has a kernel for every valuation of E-vertices: (TC) is a model extension of an arbitrary theory of E. This follows from the observation that $\mathcal{G}(TC)$ is bipartite. We justify it by the following general argument.

For a GNF theory Γ, a simpler schematic graph $\mathcal{S}(\Gamma)$ conveys often much information. Its vertices are (labeled by) the predicate symbols alone, ignoring the arguments. Each equivalence gives thus edges from the predicate in its LS, to each predicate occurring in its RS. Each of the graphs in Examples 3.2, 3.3 and 3.4 is isomorphic to such a schematic graph of its theory. For the extension (TC) from (3.5), the schematic graph $\mathcal{S}(TC)$ is

$$
\begin{array}{ccccccc}
TC & \rightrightarrows & B & \longrightarrow & C & \longrightarrow & \overline{TC}. \\
 & & \downarrow & & \downarrow & & \\
 & & E & \longleftarrow & \overline{E} & &
\end{array}
\tag{3.6}
$$

Every path in the graph $\mathcal{G}_\kappa(\Gamma)$, for any κ, results from unfolding some path in such a schematic graph $\mathcal{S}(\Gamma)$. The latter is also the homomorphic image of the former under the *canonical* homomorphism, identifying all vertices with the same predicate symbol.

The graph $\mathcal{S}(TC)$ is trivially bipartite (since so is its underlying undirected graph, having no odd undirected cycles). Using the just described canonical homomorphism and Fact 2.19, we

conclude that $\mathcal{G}(TC)$ is bipartite and, by Fact 2.15, kernel perfect. This holds generally. Whenever the schematic graph $\mathcal{S}(\Gamma)$ is bipartite, then each graph $\mathcal{G}_\kappa(\Gamma)$ is kernel perfect.

The simplified graph $\mathcal{S}(TC)$ allows thus to conclude that every $\mathcal{G}(TC)$ is solvable for every assignment to its u-sinks *Exy*. Consequently, (TC) is a model extension of any theory of E. This triviality about (TC) becomes a useful fact, when formulated generally.

FACT 3.7
An extension Δ, of any theory Γ, is a model (unique) extension of Γ if for every cardinality κ, the graph $\mathcal{G}_\kappa(\Delta)$ is (uniquely) solvable for every assignment to its u-sinks.

This follows because, given a theory Γ, each kernel M of $\mathcal{G}_\kappa(\Gamma)$ determines values of (some) *u-sinks*$(\mathcal{G}_\kappa(\Delta))$. Since no edges go from $\mathcal{G}_\kappa(\Gamma)$ to $\mathcal{G}_\kappa(\Delta)$, while the latter is solvable for every valuation of its u-sinks, kernel M of $\mathcal{G}_\kappa(\Gamma)$ can be extended to a kernel of $\mathcal{G}_\kappa(\Delta(\Gamma))$. Section 3.3 addresses the fact that the condition, requiring solvability of $\mathcal{G}(\Delta)$, is independent of Γ.

In Example 3.4, (N') is a model extension. As observed there, this follows because its graph is kernel perfect so, in particular, has a solution for every valuation of its u-sinks.

In Example 3.2, (D) is a definitional extension, while its graph $\mathcal{G}(D)$ is a dag without any rays. By the first result in kernel theory from [4], such a graph has a unique kernel and is actually kernel perfect. The effect of the syntactic restrictions on a definitional extension is, in graph terms, that its graph becomes a rayless dag—uniqueness of its kernel yields model uniqueness. It is a special case of Fact 3.7 which implies, more generally, that Δ is a model extension whenever $\mathcal{G}(\Delta)$ is kernel perfect. This general statement will be used below.
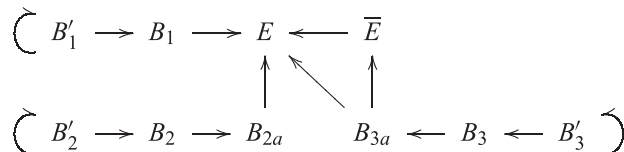
Before that, it may be useful to point out some limitations of using schematic graph $\mathcal{S}(\Delta)$ instead of $\mathcal{G}(\Delta)$, which is attractive whenever applicable. One such limitation is that even though bipartition, and solvability in general, are reflected by homomorphisms, solvability for all valuations of sinks is not. For instance, graph $\bigcirc\, a \rightrightarrows\, \rightarrow s_1 \rightarrow s_2$ is unsolvable if $s_1 = \mathbf{1}$ and $s_2 = \mathbf{0}$, but its homomorphic image $\bigcirc\, a \rightrightarrows\, \rightrightarrows s$ has a solution for each valuation of s. When $\mathcal{S}(\Delta)$ is solvable for all valuations of its u-sinks, solvability of $\mathcal{G}(\Delta)$ for all valuations of its u-sinks may fail and requires additional argument, as illustrated also in Example 3.13 further ahead.

A dual problem is exemplified in the language with constants a, b and predicate Q. Its extension with predicate P and axioms $Pa \leftrightarrow \neg Qa$ and $Pb \leftrightarrow \neg Pa$ is now model unique for every valuation of Q, but the schematic graph $\bigcirc\, P \rightarrow Q$, is solvable only when $Q = \mathbf{1}$.

In the examples above, schematic graphs of consistent extensions are still solvable, but they need not be. Let Δ be the following extension of a theory having a binary predicate symbol E:

$$\Delta: \quad \begin{aligned} B_1'x &\leftrightarrow \neg B_1 x \wedge \neg B_1' x & B_2 x &\leftrightarrow \neg B_{2a} x \\ B_2'x &\leftrightarrow \neg B_2 x \wedge \neg B_2' x & B_{2a} x &\leftrightarrow \bigwedge_y \neg Exy \\ B_3'xyz &\leftrightarrow \neg B_3 xyz \wedge \neg B_3' xyz & B_3 xyz &\leftrightarrow \neg B_{3a} xyz \\ B_1 x &\leftrightarrow \neg Exx & B_{3a} xyz &\leftrightarrow \neg \overline{E}xy \wedge \neg \overline{E}yz \wedge \neg Exz. \end{aligned}$$

More comprehensibly: $\Delta \Leftrightarrow GNF(\{\forall x \neg Exx, \forall x \exists y Exy, \forall xyz(Exy \wedge Eyz \rightarrow Exz)\})$, which forces domain to be infinite. Although Δ is consistent, its schematic graph $\mathcal{S}(\Delta)$ is unsolvable:

$$\bigcirc\, B_1' \rightarrow B_1 \rightarrow E \leftarrow \overline{E}$$
$$\bigcirc\, B_2' \rightarrow B_2 \rightarrow B_{2a} \qquad B_{3a} \leftarrow B_3 \leftarrow B_3' \,\bigcirc$$

When the only terms are variables, $\mathcal{S}(\Delta)$ is isomorphic to the graph $\mathcal{G}_1(\Delta)$ over domain with one element. As Δ forces here domain to be infinite, it does not appear possible to represent it by such a graph. It does not even seem possible to represent it by retaining an infinite number of u-sinks *Exy*, while collapsing distinct instances of internal vertices, e.g. identifying B_2d for all d in actual domain, etc. For each d, $B_2d = \mathbf{1}$ must hold; then also $B_{2a}d = \mathbf{0}$, requiring some e, distinct from d, with $Ede = \mathbf{1}$. Multiplicity of distinct vertices Ede may require multiplicity of distinct vertices B_2d.

In short, simplification offered by the schematic graph is far from universal. It seems highly improbable that any single schema could replace the whole class $Gr(\Delta)$, for arbitrary Δ, but the range of Δs for which schema $\mathcal{S}(\Delta)$ is applicable, or its generalizations, might deserve clarification.

3.2 Fixed points and positive occurrences

Restrictions on fixed point definitions provide another example, besides definitional extension, of syntactic means ensuring condition of Fact 3.7. One defines a predicate by (*) $Bx \leftrightarrow \phi x$, with B occurring in ϕ (and $\mathcal{V}(Bx) = x = \mathcal{V}(\phi)$). If M is a structure interpreting the symbols from ϕ, let (M, X) denote its expansion with the interpretation of B as $X \subseteq M$. A model of (*) over a given M is then a fixed point of the operator $\mathbb{B}^M(X) = \{m \in M \mid (M, X) \models \phi m\}$. Often, one chooses only least or greatest fixed points, but we address only the consistency conditions, that is, the mere existence of fixed points.

A simple restriction, ensuring monotonicity of $\mathbb{B}$ and existence of fixed points, forbids negative occurrences of B in ϕ. In terms of GNF, this amounts to forbidding any occurrence of B under an odd number of negations, when replacing predicates in the RS of B's equivalence, by the RSs of their equivalences. Such a substitution, performed in Example 1.2 and below, provides a procedure for identifying negative occurrences in GNF.

EXAMPLE 3.8
We repeat definitions from Example 3.3:

$$
\begin{array}{ll}
(N) & Nx \leftrightarrow \neg Bx \\
& Bx \leftrightarrow \neg Fx \wedge \neg Cx \\
& Cx \leftrightarrow \neg \overline{H}x \wedge \neg Nx \\
& \overline{H}x \leftrightarrow \neg Hx
\end{array}
$$

Each equivalence below marks one step of the successive substitution:
$Nx \leftrightarrow \neg Bx \leftrightarrow \neg(\neg Fx \wedge \neg Cx) \leftrightarrow \neg(\neg Fx \wedge \neg(\neg \overline{H}x \wedge \neg Nx))$.

In the last formula, N occurs under three negations, displaying thus its negative occurrence.

Each $\neg$ in GNF amounts to an edge in the corresponding graph, so a negative occurrence of Nx, in the RS of (some such substitution instance of) the equivalence for Nx, amounts to an odd cycle in the corresponding graph. Each odd cycle signals negative occurrence of all predicates in its vertices. In (N) above, N, B and C all have such occurrences. Forbidding negative occurrences amounts to excluding odd cycles.

Strictly speaking, what must not occur negatively is the same atom, say Td, for some d in the domain, and not merely the predicate symbol T. Negative occurrences of T in $T1 \leftrightarrow \neg T2$ or $Tx \leftrightarrow \neg Tsx$ may appear circular, but they do not create any cycles in the graph, as long as $1 \neq 2$ and $x \neq sx$. An (odd) cycle emerges only from a (negative) occurrence in RS of an atom, like Td, which is an instance of Tx in LS, from which Td is reached in the substitution process.

The traditional restriction, forbidding negative occurrences of the mere predicate symbol, ensures economy of applications. This excludes odd cycles from the theory's simplified graph $\mathcal{S}(\Gamma)$, enabling an argument similar to that lifting bipartition of $\mathcal{S}(TC)$ in (2.17) to any graph $\mathcal{G}(TC)$. Having no odd cycles, the graph $\mathcal{S}(\Gamma)$ is kernel perfect if it is finite, by Theorem 2.14. Since it is a homomorphic image of each $\mathcal{G}(\Gamma)$, every such graph is solvable by Fact 2.17. Consequently, a definition without negative occurrences of the defined predicate, and with a finite simplified graph, has fixed points over interpretation domain of each cardinality. Usefulness of such syntactic criteria is easily associated with their limitations. Kernel perfectness or, more generally, the condition of Fact 3.7, provides means for establishing existence of fixed points also in cases with negative occurrences, as illustrated by the following example.

EXAMPLE 3.9
The extension (A) to the left has the simplified graph $\mathcal{S}(A)$ to the right

$$
\begin{aligned}
(A) \quad & Ax \leftrightarrow \neg Bx \\
& Bx \leftrightarrow \neg Cx \wedge \neg Hx \\
& Cx \leftrightarrow \neg Ax \wedge \neg Bx
\end{aligned}
\qquad
A \longleftarrow C \rightleftarrows B \longrightarrow H
$$

Substituting, as described above, yields $Ax \leftrightarrow \neg(\neg(\neg Ax \wedge \neg Bx) \wedge \neg Hx)$, so the predicate A, and even the same atom Ax, occurs negatively, which is reflected by the odd cycle in the graph. All predicates from the odd cycle have negative occurrences but notwithstanding this, the graph is kernel perfect. Thus, fixed points of (A) exist and, moreover, (A) is a model unique extension: $Hx = \mathbf{1}$ forces $Bx = \mathbf{0} = Cx$ and $Ax = \mathbf{1}$, while $Hx = \mathbf{0}$ forces $Bx = \mathbf{1}$ and $Ax = \mathbf{0} = Cx$.

3.3 Universal extensions

The examples, and the remark after Fact 3.7, suggest a generalization of Definition 3.1.

DEFINITION 3.10
An extension Δ is *universal* model unique/model/conservative if it is such for every theory Γ.

This presupposes an appropriate renaming of predicate symbols of Δ, when applying it to any actual theory Γ. Fact 3.7, although formulated relatively to an arbitrary theory Γ, requires only some property of graphs $\mathcal{G}(\Delta)$. It states actually that Δ is a universal model (unique) extension and can be strengthened to the present, more general context.

FACT 3.11
An extension Δ is a universal model (unique) extension iff, for every cardinality κ, the graph $\mathcal{G}_\kappa(\Delta)$ is (uniquely) solvable for every assignment to its u-sinks.
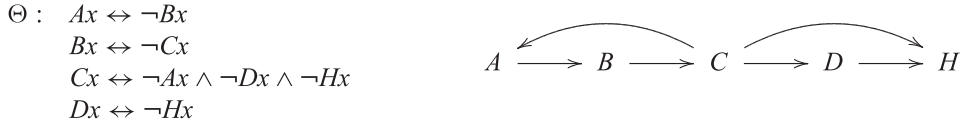
The if direction follows as in Fact 3.7, while for the opposite, any assignment to the u-sinks Γ of $\mathcal{G}_\kappa(\Delta)$, taken as the extended theory, which makes $\mathcal{G}_\kappa(\Delta)$ unsolvable, provides also a model of the theory Γ having no expansion to a model of $\Delta(\Gamma)$. Also, any assignment to the u-sinks Γ for which $\mathcal{G}_\kappa(\Delta)$ has two kernels, provides a model for Γ with two different expansions.

The extension (TC) is a universal model extension, because its graph $\mathcal{G}_\kappa(TC)$ is kernel perfect for every κ, having a homomorphisms onto the bipartite $\mathcal{S}(TC)$ in (3.7). Hence, every model of E can be extended to a model of $E \cup (TC)$. It is not, however, a universal model unique extension, as observed in point **i** under (3.5).

A universal model unique extension occurs, for instance, when its graph is a rayless dags, as with definitional extensions. Every assignment to u-sinks induces then a unique valuation of all vertices. But model unique extensions occur also in many other situations, as illustrated by the concluding examples.
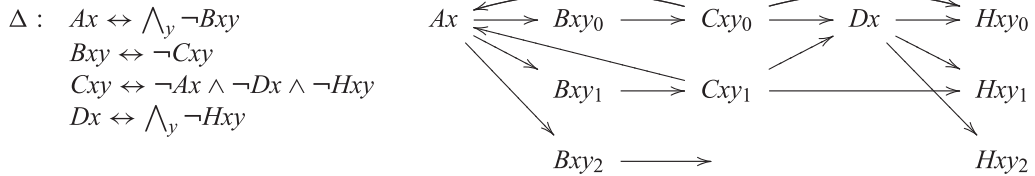
EXAMPLE 3.12
The extension Θ to the left has the simplified graph $\mathcal{S}(\Theta)$ to the right (for every κ, the graph $\mathcal{G}_\kappa(\Theta)$ consists of κ copies of this graph):

$$\Theta: \quad \begin{aligned} &Ax \leftrightarrow \neg Bx \\ &Bx \leftrightarrow \neg Cx \\ &Cx \leftrightarrow \neg Ax \wedge \neg Dx \wedge \neg Hx \\ &Dx \leftrightarrow \neg Hx \end{aligned}$$

$$A \longrightarrow B \longrightarrow C \longrightarrow D \longrightarrow H$$

The graph is not kernel perfect, as witnessed by the odd cycle $\{A, B, C\}$, showing also negative occurrences of its atoms. Still, $\mathcal{S}(\Theta)$ is uniquely solvable for every assignment to its u-sink H: $H = \mathbf{1}$ gives $C = D = A = \mathbf{0}$ and $B = \mathbf{1}$, while $H = \mathbf{0}$ yields $D = \mathbf{1} = B$ and $C = A = \mathbf{0}$. In spite of the negative occurrences, Θ is a universal model unique extension.

EXAMPLE 3.13
Below, in a more complicated version Δ of Θ from Example 3.12, some predicates have different arities:

$$\Delta: \quad \begin{aligned} &Ax \leftrightarrow \bigwedge_y \neg Bxy \\ &Bxy \leftrightarrow \neg Cxy \\ &Cxy \leftrightarrow \neg Ax \wedge \neg Dx \wedge \neg Hxy \\ &Dx \leftrightarrow \bigwedge_y \neg Hxy \end{aligned}$$

The graph $\mathcal{G}_\kappa(\Delta)$ has κ copies of the above graph, one for each $x \in \kappa$, and in each such copy, Ax (Dx) has edges to κ vertices Bxy_i (Hxy_i), for each $y_i \in \kappa$. Each vertex Bxy_i starts a copy of the subgraph following Bxy_0 (and Bxy_1), with an edge from each Cxy_i to the same Ax and Dx.

Now, $\mathcal{S}(\Delta) = \mathcal{S}(\Theta)$. The canonical homomorphism from any $\mathcal{G}_\kappa(\Delta)$ onto $\mathcal{S}(\Delta)$ reflects kernels of $\mathcal{S}(\Delta)$ by Fact 2.17. So $\mathcal{G}_\kappa(\Delta)$ is solvable whenever, for each $x \in \kappa$, all $Hxy_i = \mathbf{0}$ or all $Hxy_i = \mathbf{1}$. To conclude that Δ is a universal model extension, we have to consider also the case of $Hxy_i = \mathbf{1}$ only for some y_i. Then $Dx = Cxy_i = \mathbf{0}$, making $Bxy_i = \mathbf{1}$ and $Ax = \mathbf{0}$. All other Cxy_j, Bxy_j are then determined by the respective Hxy_j. Thus, Δ is a universal model extension (in fact, unique), all negative occurrences and odd cycles notwithstanding.

References

[1] M. Bezem, C. Grabmayer and M. Walicki. Expressive power of digraph solvability. *Annals of Pure and Applied Logic*, **163**, 200–212, 2012.

[2] A. Gupta and N. Belnap. *The Revision Theory of Truth*. The MIT Press, 1993.

[3] M. Richardson. Solutions of irreflexive relations. *The Annals of Mathematics, Second Series*, **58**, 573–590, 1953.

[4] von J. Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[5] M. Walicki. Resolving infinitary paradoxes. *Journal of Symbolic Logic*, **82**, 709–723, 2017.

[6] M. Walicki. Kernels of digraphs with finitely many ends. *Discrete Mathematics*, **342**, 473–486, 2019.