Graphical Conjunctive Queries

Filippo Bonchi
University of Pisa
filippo.bonchi@unipi.it

Jens Seeber
IMT School for Advanced Studies Lucca
jens.seeber@imtlucca.it

Paweł Sobociński
University of Southampton
sobocinski@gmail.com

Abstract

The Calculus of Conjunctive Queries (CCQ) has foundational status in database theory. A celebrated theorem of Chandra and Merlin states that CCQ query inclusion is decidable. Its proof transforms logical formulas to graphs: each query has a *natural model*—a kind of *graph*—and query inclusion reduces to the existence of a graph homomorphism between natural models.

We introduce the diagrammatic language *Graphical Conjunctive Queries* (GCQ) and show that it has the same expressivity as CCQ. GCQ terms are *string diagrams*, and their algebraic structure allows us to derive a sound and complete axiomatisation of query inclusion, which turns out to be exactly Carboni and Walters' notion of *cartesian bicategory of relations*. Our completeness proof exploits the combinatorial nature of string diagrams as (certain cospans of) hypergraphs: Chandra and Merlin's insights inspire a theorem that relates such cospans with spans. Completeness and decidability of the (in)equational theory of GCQ follow as a corollary. Categorically speaking, our contribution is a model-theoretic completeness theorem of free cartesian bicategories (on a relational signature) for the category of sets and relations.

2012 ACM Subject Classification Theory of computation → Categorical semantics

Keywords and phrases conjunctive query inclusion, string diagrams, cartesian bicategories

Digital Object Identifier 10.4230/LIPIcs.CSL.2018.12

1 Introduction

Conjunctive queries (CCQ) are first-order logic formulas that use only relation symbols, equality, truth, conjunction, and existential quantification. They are a kernel language of queries to relational databases and are the foundations of several languages: they are select-project-join queries in relational algebra [16], or select-from-where queries in SQL [13]. While expressive enough to encompass queries of practical interest, they admit algorithmic analysis: in [14], Chandra and Merlin showed that the problem of *query inclusion* is NP-complete.

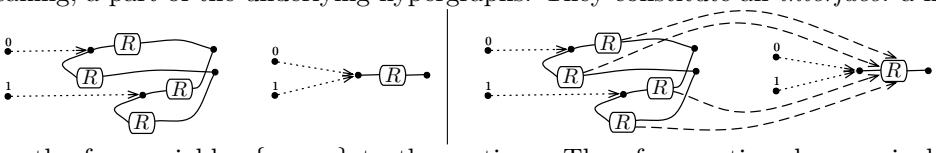
For an example of query inclusion in action, consider formulas

$$\phi = \exists z_0 \colon (x_0 = x_1) \wedge R(x_0, z_0) \text{ and } \psi = \exists z_0, z_1 \colon R(x_0, z_0) \wedge R(x_1, z_0) \wedge R(x_0, z_1) \wedge R(x_1, z_1),$$

with free variables x_0, x_1. Irrespective of model, and thus the interpretation of the relation symbol R, every free variable assignment satisfying ϕ satisfies ψ: i.e. ϕ is included in ψ.

Chandra and Merlin's insight involves an elegant reduction to graph theory, namely the existence of a hypergraph homomorphism from a graphical encoding of ψ to that of ϕ. Below on the left we give a graphical rendering of ψ and ϕ, respectively: vertices represent variables,

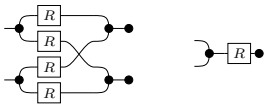
while edges are labelled with relation symbols. The dotted connections are not, strictly speaking, a part of the underlying hypergraphs. They constitute an *interface*: a mapping



from the free variables $\{x_0, x_1\}$ to the vertices. The aforementioned query inclusion is witnessed by an interface-preserving hypergraph homomorphism, displayed above on the right. In category-theoretic terms, hypergraphs-with-interfaces are *discrete cospans*, and the homomorphisms are *cospan homomorphisms*.

In previous work [5], the first and third authors with Gadducci, Kissinger and Zanasi showed that such cospans characterise an important family of *string diagrams*—i.e. diagrammatic representations of the arrows of monoidal categories—namely those equipped with an algebraic structure known as a special Frobenius algebra. This motivated us to study the connection between this fashionable algebraic structure—which has been used in fields as diverse as quantum computing [1, 17, 30, 25], concurrency theory [7, 8, 10, 9], control theory [6, 3] and linguistics [31]—and conjunctive queries.

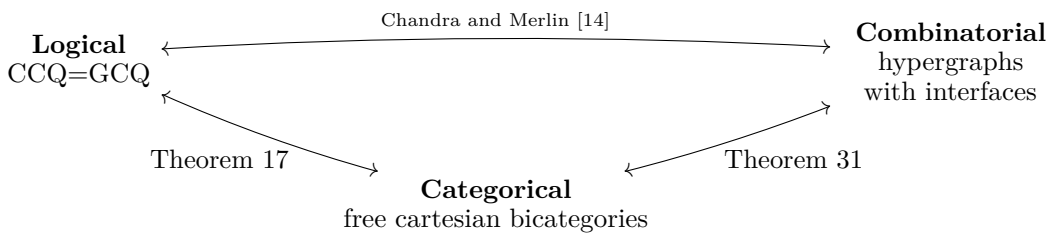
We introduce the logic of Graphical Conjunctive Queries (GCQ). Although superficially unlike CCQ, we show that it is equally expressive. Its syntax lends itself to string-diagrammatic representation and diagrammatic reasoning respects the underlying logical semantics. GCQ string diagrams for ψ and ϕ are drawn below. Note that, while GCQ syntax does not have variables, the concept of CCQ free variable is mirrored by "dangling" wires in diagrams.



While interesting in its own right as an example of a string-diagrammatic representation of a logical language—which has itself become a topic of recent interest [21]—GCQ comes into its own when reasoning about query inclusion, which is characterised by the laws of *cartesian bicategories*. This important categorical structure was introduced by Carboni and Walters [12] who were, in fact, aware of the logical interpretation, mentioning it in passing without giving the details. Our definition of GCQ, its expressivity, and *soundness* of the laws of cartesian bicategories w.r.t. query inclusion is testament to the depth of their insights.

The main contribution of our work is the *completeness* of the laws of cartesian bicategories for query inclusion (Theorem 17).

As a side result, we obtain a categorical understanding of the proof by Chandra and Merlin. This uncovers a beautiful triangle relating logical, combinatorial and categorical structures, similar to the Curry-Howard-Lambek correspondence relating intuitionistic propositional logic, λ-calculus and free cartesian closed categories.



The rightmost side of the triangle (Theorem 31) provides a combinatorial characterisation of free cartesian bicategories as discrete cospans of hypergraphs, with the Chandra and Merlin ordering: the existence of a cospan homomorphism in the opposite direction. This result can also be regarded as an extension of the aforementioned [5] to an enriched setting. The fact that the Chandra and Merlin ordering is not antisymmetric forces us to consider preorder-enrichment as opposed to the usual [12] poset-enrichment of cartesian bicategories.[1]

The step from posets to preorders is actually beneficial: it provides a one-to-one correspondence between hypergraphs and models which we see as functors, following the tradition of categorical logic. The model corresponding to a hypergraph G is exactly the (contravariant) Hom-functor represented by G. By a Yoneda-like argument, we obtain a "preorder-enriched analogue" of Theorem 17 (Theorem 37). With this result, proving Theorem 17 reduces to descending from the preorder-enriched setting down to poset-enrichment.

Working with both poset- and preorder-enriched categories means that there is a relatively large number of categories at play. We give a summary of the most important ones in the table below, together with pointers to their definitions. The remainder of this introduction is a roadmap for the paper, focussing on the roles played by the categories mentioned below.

	preordered	posetal
free categories	$\mathbb{CB}^{\leq}_{\Sigma}$ (Def 29)	$\mathbb{CB}_{\Sigma}$ (Def 21)
semantic domains for the logic	$\mathsf{Span}^{\leq}\,\mathbf{Set}$ (Def 33)	**Rel** (Ex 20)
combinatorial structures	$\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ (Def 26)	-

We begin by justifying the "equation" CCQ=GCQ in the triangle above: we recall CCQ and introduce GCQ in Sections 2 and 3, respectively, and show that they have the same expressivity. We explore the algebraic structure of GCQ in Sections 4 and 5, which—as we previously mentioned—is exactly that of cartesian bicategories. As instances of these, we introduce $\mathbb{CB}_{\Sigma}$, the free cartesian bicategory, and **Rel**, the category of sets and relations.

In Section 6 we introduce *preordered* cartesian bicategories (the free one denoted by $\mathbb{CB}^{\leq}_{\Sigma}$) and the category of discrete cospans of hypergraphs with the Chandra and Merlin preorder, denoted by $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$. Theorem 31 states that these two are isomorphic.

Theorem 37 is proved in Section 7. Rather than **Rel**, the preordered setting calls for models in $\mathsf{Span}^{\leq}\,\mathbf{Set}$, the preordered cartesian bicategory of spans of sets. In Section 8, we explain the passage from preorders to posets, completing the proof of Theorem 17.

We delay a discussion of the ramification of our work, a necessarily short and cursory account—due to space restrictions—of the considerable related work, and directions for future work to Section 9. We conclude with the observation that *(i)* the diagrammatic language for formulas, *(ii)* the semantics, e.g. of composition of diagrams—what we understand in modern terms as the combination of conjunction and existential quantification—and *(iii)* the use of diagrammatic reasoning as a powerful method of logical reasoning actually go back to the pre-Frege work of the 19th century American polymath CS Peirce on *existential graphs*. Interestingly, it is only recently (see, e.g. [29]) that this work has been receiving the attention that it richly deserves.

Preliminaries. We assume familiarity with basic categorical concepts, in particular symmetric monoidal, ordered and preordered categories. We do not assume familiarity with cartesian bicategories: the acquainted reader should note that what we call "cartesian bicategories"

[1] While cartesian bicategories were later generalised [11] to a bona fide higher-dimensional setting, our preorder-enriched variant seems to be an interesting stop along the way.

are "cartesian bicategories of relations" in [12]. A *prop* is a symmetric strict monoidal category where objects are natural numbers, and the monoidal product on objects is addition $m \oplus n := m + n$. Due to space restrictions, most proofs are in the Appendix.

2 Calculus of Conjunctive Queries

Assume a set Σ of relation symbols with arity function $ar : \Sigma \to \mathbb{N}$ and a countable set $Var = \{x_i \mid i \in \mathbb{N}\}$ of variables. The grammar for the calculus of conjunctive queries is:

$$\Phi ::= \top \mid \Phi \wedge \Phi \mid x_i = x_j \mid R(\vec{x}) \mid \exists x.\Phi \tag{CCQ}$$

where $R \in \Sigma$, $ar(R) = n$, and $\vec{x}$ is a list of length n of variables from Var. We assume the standard bound variable conventions and some basic metatheory of formulas; in particular we write $\phi[\vec{x}/\vec{y}]$, where $\vec{x}, \vec{y}$ are variable lists of equal length, for the simultaneous substitution of variables from $\vec{x}$ for variables in $\vec{y}$. We write $\vec{x}_{[m,n]}$, where $m \leq n$, for the list of variables $x_m, x_{m+1}, \ldots, x_n$. Given a formula ϕ, $fv(\phi)$ is the set of its free variables.

The semantics of (CCQ) formulas is standard and inherited from first order logic.

▶ **Definition 1.** A model $\mathcal{M} = (X, \rho)$ is a set X and, for each $R \in \Sigma$, a set $\rho(R) \subseteq X^{ar(R)}$.

Given a model $\mathcal{M} = (X, \rho)$, the semantics $[\![\phi]\!]_{\mathcal{M}}$ is the set of all assignments of elements from X to $fv(\phi)$ that makes it evaluate to truth, given the usual propositional interpretation.

In order to facilitate a principled definition of the semantics (Definition 3) and to serve the needs of our diagrammatic approach, we will need to take a closer look at free variables. To this end, we give an alternative, sorted presentation of (CCQ) that features explicit free variable management. As we shall see, the system of judgments below will allow us to derive $n \vdash \phi$ where $n \in \mathbb{N}$, whenever ϕ is a formula of CCQ and $fv(\phi) \subseteq \{x_0, \ldots, x_{n-1}\}$.

$$\frac{}{0 \vdash \top}\,(\top) \qquad \frac{R \in \Sigma \quad ar(R) = n}{n \vdash R(x_0, \ldots, x_{n-1})}\,(\Sigma) \qquad \frac{n \vdash \phi}{n-1 \vdash \exists x_{n-1}.\phi}\,(\exists)$$

$$\frac{}{2 \vdash x_0 = x_1}\,(=) \qquad \frac{m \vdash \phi \quad n \vdash \psi}{m+n \vdash \phi \wedge (\psi[\vec{x}_{[m,m+n-1]}/\vec{x}_{[0,x_{n-1}]}])}\,(\wedge)$$

Note that the above are restrictive: e.g. $(\wedge)$ enforces disjoint sets of variables, and $(\exists)$ allows quantification only over the last variable. To overcome these limitations we include three structural rules that allow us to manipulate (swap, identify, and introduce) free variables.

$$\frac{n \vdash \phi \quad (0 \leq k < n-1)}{n \vdash \phi[x_{k+1}, x_k / x_k, x_{k+1}]}\,(\mathsf{Sw}_{n,k}) \qquad \frac{n \vdash \phi}{n-1 \vdash \phi[x_{n-2}/x_{n-1}]}\,(\mathsf{Id}_n) \qquad \frac{n \vdash \phi}{n+1 \vdash \phi}\,(\mathsf{Nu}_n)$$

Rule Sw allows us to swap two free variables. Alone, Id identifies the final and the penultimate free variable; used together with Sw it allows for the identification of any two. Finally, Nu introduces a free variable. The eight suffice for any CCQ formula, in the following sense:

▶ **Proposition 2.** *ϕ is a formula derived from* (CCQ) *with $fv(\phi) \subseteq \{x_0, \ldots, x_{n-1}\}$ iff $n \vdash \phi$.*

We use the sorted presentation to define the semantics.

$$[\![0 \vdash \top]\!]_{\mathcal{M}} = \{\bullet\} \quad (\top) \qquad (\vec{u}, v, w, \vec{x}) \in [\![n \vdash \phi[x_{k+1}, x_k / x_k, x_{k+1}]]\!]_{\mathcal{M}} \Leftrightarrow (\vec{u}, w, v, \vec{x}) \in [\![n \vdash \phi]\!]_{\mathcal{M}} \quad (\mathsf{Sw}_{n,k})$$

$$[\![n \vdash R(x_0, \ldots, x_{n-1})]\!]_{\mathcal{M}} = \rho(R) \quad (\Sigma) \qquad (\vec{v}, w) \in [\![n-1 \vdash \phi[x_{n-2} / x_{n-1}]]\!]_{\mathcal{M}} \Leftrightarrow (\vec{v}, w, w) \in [\![n \vdash \phi]\!]_{\mathcal{M}} \quad (\mathsf{Id}_n)$$

$$[\![2 \vdash x_0 = x_1]\!]_{\mathcal{M}} = \{(v, v) \mid v \in X\} \quad (=) \qquad \vec{v} \in [\![n-1 \vdash \exists x_{n-1}.\phi]\!]_{\mathcal{M}} \Leftrightarrow \exists w \in X. (\vec{v}, w) \in [\![n \vdash \phi]\!]_{\mathcal{M}} \quad (\exists)$$

$$[\![n+1 \vdash \phi]\!]_{\mathcal{M}} = [\![n \vdash \phi]\!]_{\mathcal{M}} \times X \quad (\mathsf{Nu}_n) \qquad [\![m+n \vdash \phi \wedge (\psi[\ldots])]\!]_{\mathcal{M}} = [\![m \vdash \phi]\!]_{\mathcal{M}} \times [\![n \vdash \psi]\!]_{\mathcal{M}} \quad (\wedge)$$

Figure 1 Semantics of CCQ for a model $\mathcal{M} = (X, \rho)$. We write $\bullet$ for the unique element of X^0.

▶ **Definition 3.** Given a model $\mathcal{M} = (X, \rho)$, the semantics of $n \vdash \phi$ is a set of tuples $[\![n \vdash \phi]\!]_{\mathcal{M}} \subseteq X^n$. We define it in Figure 1 by recursion on the derivation of $n \vdash \phi$.

Finally, we define the concepts that are of central interest: *query equivalence* and *inclusion*.

▶ **Definition 4.** Given $n \vdash \phi$ and $n \vdash \psi$, we say that ϕ and ψ are equivalent and write $\phi \equiv \psi$ if for all models $\mathcal{M}$ we have $[\![n \vdash \phi]\!]_{\mathcal{M}} = [\![n \vdash \psi]\!]_{\mathcal{M}}$. We write $\phi \leqq \psi$ when, for all $\mathcal{M}$, $[\![n \vdash \phi]\!]_{\mathcal{M}} \subseteq [\![n \vdash \psi]\!]_{\mathcal{M}}$. Clearly $\phi \leqq \psi$ and $\psi \leqq \phi$ implies $\phi \equiv \psi$.

3 Graphical conjunctive queries

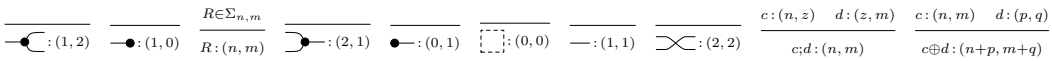


Figure 2 Sort inference rules.

We introduce an alternative logic, called Graphical Conjunctive Queries (GCQ). GCQ and CCQ are—superficially—quite different. Nevertheless, in Propositions 9 and 10 we show that they have the same expressive power. The grammar of GCQ formulas is given below.

$$c ::= \quad \text{—●} \mid \text{—(} \mid \text{)—} \mid \text{●—} \mid \text{⬚} \mid \text{—} \mid \text{⨯} \mid c \oplus c \mid c \,;\, c \mid R \qquad \text{(GCQ)}$$

GCQ syntax is a radical departure from (CCQ). Rather than use CCQ's existential quantification and conjunction, GCQ uses the operations of monoidal categories: composition and monoidal product. There are no variables, thus no assumptions of their countable supply, nor any associated metatheory of capture-avoiding substitution.

The price is a simple sorting discipline. A *sort* is a pair (n, m), with $n, m \in \mathbb{N}$. We consider only terms sortable according to Figure 2. There and in (GCQ), R ranges over the symbols of a *monoidal* signature Σ, a set of relation symbols equipped with both an arity and a *coarity*: $\Sigma_{n,m}$ consists of the symbols in Σ with arity n and coarity m. A GCQ signature plays a similar role to relation symbols in CCQ: we abuse notation for this reason. A simple induction shows sort uniqueness: if $c : (n, m)$ and $c : (n', m')$ then $n = n'$ and $m = m'$.

In (GCQ) we used a graphical rendering of GCQ constants. Indeed, we will not write terms of GCQ as formulas, but instead represent them as 2-dimensional diagrams. The justification for this is twofold: the diagrammatic conventions introduced in this section mean that a diagram is a readable, faithful and unambiguous representation of a sorted (GCQ) term. More importantly, our characterisation of query inclusion in subsequent sections consists of intuitive topological deformations of the diagrammatic representations of formulas.

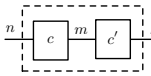
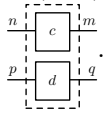
A GCQ term $c : (n, m)$ is drawn as a diagram with n "dangling wires" on the left, and m on the right. Roughly speaking, dangling wires are GCQ's answer to the free variables of CCQ. Composing (;) means connecting diagrams in series and tensoring means

$[\![$  $]\!]_{\mathcal{M}} = \{(x, \binom{x}{x}) \mid x \in X\}$ $\quad [\![$  $]\!]_{\mathcal{M}} = \{(x, \bullet) \mid x \in X\}$ $\quad [\![c \oplus d]\!]_{\mathcal{M}} = [\![c]\!]_{\mathcal{M}} \oplus [\![d]\!]_{\mathcal{M}}$ $\quad [\![$  $]\!]_{\mathcal{M}} = \{(\bullet, \bullet)\}$

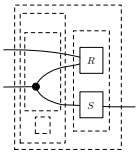
$[\![$  $]\!]_{\mathcal{M}} = \{(\binom{x}{x}, x) \mid x \in X\}$ $\quad [\![$  $]\!]_{\mathcal{M}} = \{(\bullet, x) \mid x \in X\}$ $\quad [\![c\,;d]\!]_{\mathcal{M}} = [\![c]\!]_{\mathcal{M}}\,;[\![d]\!]_{\mathcal{M}}$ $\quad [\![R]\!]_{\mathcal{M}} = \rho(R)$

$[\![$  $]\!]_{\mathcal{M}} = \{(\binom{x}{y}, \binom{y}{x}) \mid x, y \in X\}$ $\quad [\![$  $]\!]_{\mathcal{M}} = \{(x, x) \mid x \in X\}$

Figure 3 Semantics of GCQ for a model $\mathcal{M} = (X, \rho)$. We used the notation $R\,;S = \{(x, z) \mid \exists y \in Y \text{ s.t. } (x, y) \in R \text{ and } (y, z) \in S\}$ and $R \oplus S = \{(\binom{x}{u}, \binom{y}{v}) \mid (x, y) \in R \text{ and } (u, v) \in S\}$. $\bullet$ is the unique element of X^0 and $\begin{pmatrix} x_0 \\ \vdots \\ x_{n-1} \end{pmatrix}$ an element of X^n.

stacking. The shorthand  stands for m wires in parallel. The box  stands for a relation symbol $R \in \Sigma_{n,m}$. Thus, given $c : (n, m)$, $c' : (m, k)$, $c\,;c' : (n, k)$ is drawn  , and given $d : (p, q)$, $c \oplus d : (n + p, m + q)$ is drawn  .

▶ **Example 5.** Consider $((— \oplus$  $) \oplus$  $);(R \oplus S) : (2, 1)$, assuming $R \in \Sigma_{2,0}$, $S \in \Sigma_{1,1}$. Its diagrammatic rendering is on the right. Note that the use of the dotted boxes induces a tree-like quality to diagrams. Indeed, they are a faithful representation for syntactic terms constructed from (GCQ).



We now turn to semantics. First, the notion of model of GCQ is similar to a model of CCQ.

▶ **Definition 6.** A model $\mathcal{M} = (X, \rho)$ is a set X and, for each $R \in \Sigma_{n,m}$, $\rho(R) \subseteq X^n \times X^m$.

Given a model $\mathcal{M} = (X, \rho)$, the semantics of $c : (n, m)$ is the relation $[\![c]\!]_{\mathcal{M}} \subseteq X^n \times X^m$ defined recursively in Figure 3. Armed with a notion of semantics, we can define query equivalence ($\equiv$) and inclusion ($\leqq$) for GCQ terms analogously to Definition 4.

▶ **Example 7.** Consider the GCQ term  of sort $(0, 0)$. For a model $\mathcal{M} = (X, \rho)$, its semantics $[\![$  $]\!]_{\mathcal{M}} \subseteq X^0 \times X^0$ is either the empty relation $\emptyset$, if X is empty, or the relation $\{(\bullet, \bullet)\}$, if X is not empty. Since $\emptyset \subseteq \{(\bullet, \bullet)\}$, and since $[\![$  $]\!]_{\mathcal{M}} = \{(\bullet, \bullet)\}$ for all models $\mathcal{M}$, it holds that  $\leqq$  . Intuitively, the first term corresponds to the CCQ formula $\exists x.\top$, holding in all non empty models, while the second corresponds to the formula $\top$. In the remainder of this section we will make this intuition precise.

3.1 Expressivity

We now give a semantics preserving translation Θ from CCQ to GCQ. For each CCQ relation symbol $R \in \Sigma$ of arity n, we assume a corresponding GCQ symbol $R \in \Sigma_{n,0}$. Using Proposition 2, it suffices to consider judgments $n \vdash \phi$. For each, we obtain a GCQ term $\Theta(n \vdash \phi) : (n, 0)$. The translation Θ, given in Figure 4, is defined by recursion on the derivation of $n \vdash \phi$. Given a CCQ model $\mathcal{M} = (X, \rho)$, let $\Theta(\mathcal{M}) = (X, \rho')$ be the obvious corresponding GCQ model: $\rho'(R) = \rho(R) \times \{\bullet\}$. The following confirms that semantics is preserved.

▶ **Proposition 8.** *For a CCQ model $\mathcal{M} = (X, \rho)$: $\vec{v} \in [\![n \vdash \phi]\!]_{\mathcal{M}}$ iff $(\vec{v}, \bullet) \in [\![\Theta(n \vdash \phi)]\!]_{\Theta(\mathcal{M})}$.*

Furthermore, to characterise query inclusion in CCQ, it is enough to characterise it in GCQ.

▶ **Proposition 9.** *For all CCQ formulas $n \vdash \phi$ and $n \vdash \psi$, $\phi \leqq_{CCQ} \psi$ iff $\Theta(\phi) \leqq_{GCQ} \Theta(\psi)$.*

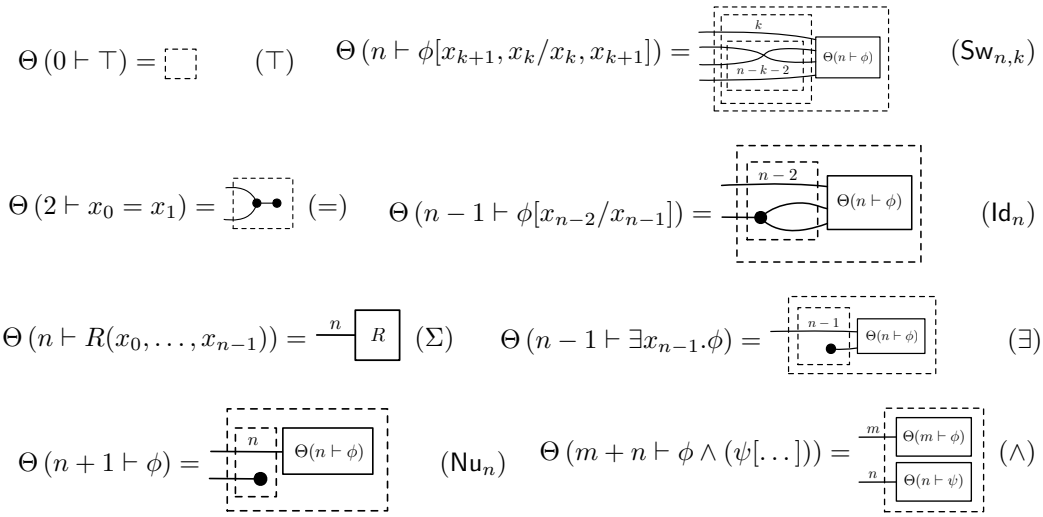


Figure 4 Translation Θ from CCQ to GCQ.

Proposition 8 yields the left-to-right direction. For right-to-left, we give a semantics-preserving translation Λ from GCQ to CCQ in Appendix A. Modulo ≡, Λ is inverse of Θ.

▶ **Proposition 10.** *There exists a semantics preserving translation* Λ *from GCQ to CCQ such that for all GCQ terms* $c, d\colon (n, m)$*, it holds that* $c \leqq_{GCQ} d$ *iff* $\Lambda(c) \leqq_{CCQ} \Lambda(d)$*.*

Propositions 9 and 10 together imply that *CCQ* and *GCQ* have the same expressive power.

▶ **Example 11.** Recall from Example 7, that  is related to $\exists x.\top$. By translating the CCQ formula $0 \vdash \exists x_0.\top$ via Θ, one obtains  . The latter and  are different—syntactically—but they are equal modulo ≡. Note that their diagrams are similar: in the next section, we prove that terms differing only by dashed boxes are equal modulo ≡.

4 From terms to string diagrams

The first step towards an equational characterisation of query inclusion is to move from GCQ, where the graphical notation was a faithful representation of ordinary syntactic terms, to bona fide *string diagrams*; that is, graphical notation for the arrows of a prop, a particularly simple kind of symmetric monoidal category (SMC). This is an advantage of GCQ syntax: its operations are amenable to an elegant axiomatisation. A hint of the good behaviour of GCQ operations is that query inclusion (and, therefore, query equivalence is) a (pre)congruence.

▶ **Lemma 12.**

(i) *Let* $c, c' : (n, m)$ *and* $d, d' : (m, k)$ *with* $c \leqq c'$ *and* $d \leqq d'$*. Then* $(c\,;d) \leqq (c'\,;d')$*.*

(ii) *Let* $c, c' : (n, m)$ *and* $d, d : (p, q)$ *with* $c \leqq c'$ *and* $d \leqq d'$ *Then* $(c \oplus d) \leqq (c' \oplus d')$*.*

We now consider the laws of strict symmetric monoidal categories (Figure 5) and discover that any two GCQ terms identified by them are logically equivalent. This means that we can eliminate the clutter of dashed boxes from our graphical notation.

▶ **Proposition 13.** ≡ *satisfies the axioms in Figure 5.*

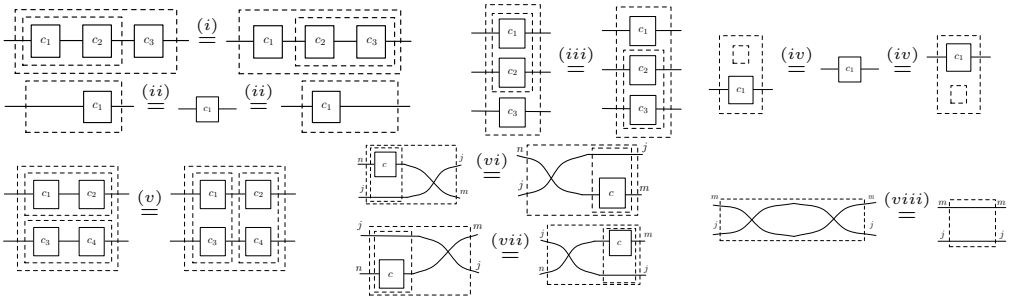
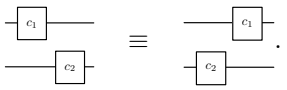


Figure 5 Axioms of strict symmetric monoidal categories. Wire annotations in (i)-(v) have been omitted for clarity.

The terms of GCQ up-to query equivalence, therefore, organise themselves as arrows of a *monoidal category* (axioms (i)-(v)), and the operation of "erasing all dotted boxes" from diagrams is well-defined. The resulting structure is the well-known combinatorial/topological concept of *string diagram*. Equality reduces to the connectivity of their components, and is thus stable under intuitive topological transformations, known as *diagrammatic reasoning*. For instance, axioms (ii) and (v) in Figure 5 imply that for $c_1 : (m_1, n_1)$ and $c_2 : (m_2, n_2)$



Axioms (vi)-$(viii)$ assert that GCQ terms modulo $\equiv$ form a *symmetric* monoidal category (SMC). Therein,  stands for the crossing of n wires over m wires. This has a standard recursive definition, using  , — and the operations of GCQ. Intuitively, boxes "slide over" wire crossings. Moreover, it is well-known that (vi) and (vii) of Figure 5 imply the Yang-Baxter equation for crossings, which—with $(viii)$—implies that in diagrammatic reasoning wires do not "tangle" and crossings act like permutations of finite sets.

5 Axiomatisation

We have seen that, up-to query equivalence, GCQ enjoys the structural properties of SMCs. Here we give further properties that *characterise* query equivalence ($\equiv$) and inclusion ($\leqq$).

Our first observation is that  and form, modulo $\equiv$, a *commutative monoid*, i.e., they satisfy axioms (A), (C) and (U) in Figure 6. Similarly,  and form a *cocommutative comonoid* (axioms (A^{op}), (C^{op}) and (U^{op})). Monoid and comonoid together give rise to a *special Frobenius bimonoid* (axioms (S) and (F)), a well-known algebraic structure that is important in various domains [1, 17, 7, 6].

▶ **Proposition 14.** $\equiv$ *satisfies the axioms in Figure 6.*

Figure 7 shows a set of properties of query inclusion. The two axioms on the left state that  is the left adjoint of  and the central axioms assert that  is the left adjoint of  . For the rightmost ones, it is convenient to introduce some syntactic sugar:  ,  ,  and stand for the n-fold versions of monoid and comonoid. Now, axiom (L_1) asserts that  laxly commutes with  , while axiom (L_2) states that it laxly commutes with  . In a nutshell,  is required to be a *lax comonoid morphism*.

▶ **Proposition 15.** $\leqq$ *satisfies the axioms of Figure 7.*

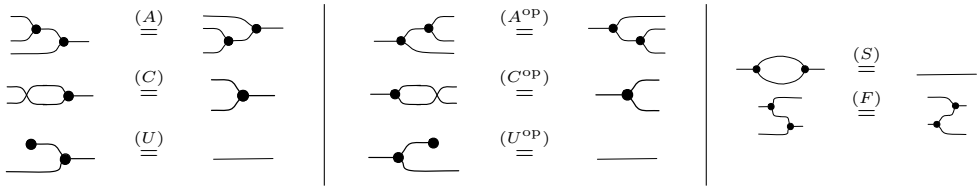


Figure 6 Axioms for special Frobenius bimonoids.

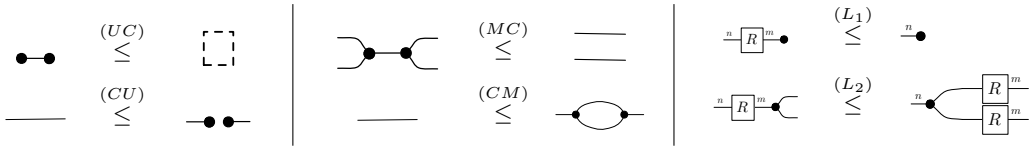


Figure 7 Axioms for adjointness of —● and ●— (left) adjointness of —◖ and ◗— (center) lax comonoid morphism (right).

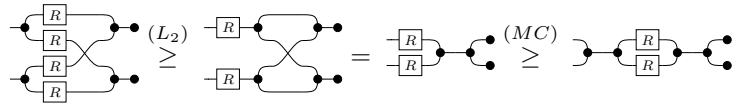
Interestingly, the observations we made so far suffice to characterise query equivalence and inclusion. This is the main theorem which we will prove in the remainder of this paper.

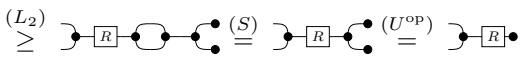
▶ **Definition 16.** The relation $\leq_{\mathbb{CB}_\Sigma}$ on the terms of GCQ is the smallest precongruence containing the equalities in Figures 5, 6, their converses and the inequalities in Figure 7. The relation $=_{\mathbb{CB}_\Sigma}$ is the intersection of $\leq_{\mathbb{CB}_\Sigma}$ and its converse.

▶ **Theorem 17.** $\leq_{\mathbb{CB}_\Sigma} = \leqq$

▶ Remark. There is an apparent redundancy in Figure 7: (CM) follows immediately from (S) in Figure 6, while (S) can by derived from (CU), (U^{op}) and (U) for one inclusion and (CM) for the other. We kept both (CM) and (S) because, as we shall see in §6, it is important to keep the algebraic structures of Figures 6 and 7 separate.

▶ **Example 18.** Recall the example from the Introduction. We can now prove the inclusion of queries using diagrammatic reasoning, as shown below. In the unlabeled equality we make use of the well-known *spider theorem*, which holds in every special Frobenius algebra [27].





5.1 Cartesian bicategories

The structure in Figures 6 and 7 is not arbitrary: these are exactly the laws of *cartesian bicategories*, a concept introduced by Carboni and Walters [12], that we recall below.

▶ **Definition 19.** A cartesian bicategory is a symmetric monoidal category $\mathcal{B}$ with tensor $\oplus$ and unit I, enriched over the category of partially ordered sets, such that:

1. every object X has a special Frobenius bimonoid: a monoid  $: X \oplus X \to X$,  $: I \to X$, a comonoid $: X \to X \oplus X$, $: X \to I$ satisfying the axioms in Figure 6;
2. the monoid and comonoid on X are adjoint (axioms in Figure 7, left and center);
3. every arrow $R \colon X \to Y$ is a lax comonoid morphism (axioms in Figure 7, right).

Furthermore, a morphism $\mathcal{F}$ of cartesian bicategories is a functor $\mathcal{F} \colon \mathcal{B}_1 \to \mathcal{B}_2$ preserving the tensor, the partial orders and the monoid and comonoid on every object.

▶ **Example 20.** The archetypal cartesian bicategory is the category of sets and relations **Rel**, with cartesian product $\times$ as tensor and $1 = \{\bullet\}$ as unit. To be precise, **Rel** has sets as objects and relations $R \subseteq X \times Y$ as arrows $X \to Y$. Composition and tensor are defined as in Figure 3. For each set X, the monoid and comonoid structure is:  $= \{(x, \binom{x}{x}) \mid x \in X\}$, $= \{(x, \bullet) \mid x \in X\}$, $= \{(\binom{x}{x}, x) \mid x \in X\}$, $= \{(\bullet, x) \mid x \in X\}$.

Cartesian bicategories allow us to employ the usual construction from categorical logic: the arrows of the cartesian bicategory freely generated from Σ are GCQ terms modulo $=_{\mathbb{CB}_\Sigma}$, and morphisms from this cartesian bicategory to **Rel** are exactly GCQ models.

▶ **Definition 21.** The ordered prop $\mathbb{CB}_\Sigma$ has GCQ terms of sort (n, m) modulo $=_{\mathbb{CB}_\Sigma}$ as arrows $n \to m$. These are ordered by $\leq_{\mathbb{CB}_\Sigma}$.

▶ **Lemma 22.** *$\mathbb{CB}_\Sigma$ is a cartesian bicategory.*

▶ **Proposition 23.** *Models of GCQ (Definition 6) are in bijective correspondence with morphisms of cartesian bicategories $\mathbb{CB}_\Sigma \to$ **Rel**.*

6 Discrete cospans of hypergraphs

In order to prove Theorem 17, in this section we give a combinatorial characterisation of free cartesian bicategories as hypergraphs-with-interfaces, formalised as a (bi)category of *cospans* equipped with an ordering inspired by Merlin and Chandra [14].

Indeed, the appearance of graph-like structures in the context of conjunctive queries should not come as a shock. Merlin and Chandra, to compute inclusion $\varphi \leqq \psi$ of CCQ queries, translate them into hypergraphs G_φ, G_ψ with "interfaces" that represent free variables. Then $\varphi \leqq \psi$ iff there exists an interface-preserving homomorphism from G_ψ to G_ϕ.

6.1 Hypergraphs and Cospans

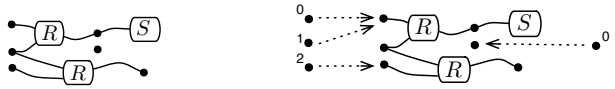
Our goal in this part is the characterisation of GCQ diagrams as certain combinatorial structures. We start by introducing the notion of Σ-hypergraph.

▶ **Definition 24** (Σ-hypergraph)**.** Let Σ be a monoidal signature. A Σ-hypergraph G is a set G_V of vertices and, for each $R \in \Sigma_{n,m}$, a set of R-labeled hyperedges G_R, with source and target functions $s_R \colon G_R \to (G_V)^n, t_R \colon G_R \to (G_V)^m$. A morphism $f \colon G \to G'$ is a function $f_V \colon G_V \to G'_V$ and a family $f_R \colon G_R \to G'_R$, for each $R \in \Sigma_{n,m}$, s.t. the following commutes.

$$\begin{array}{ccccc} (G_V)^n & \xleftarrow{s_R} & G_R & \xrightarrow{t_R} & (G_V)^m \\ {\scriptstyle f_V}\downarrow & & {\scriptstyle f_R}\downarrow & & \downarrow{\scriptstyle f_V} \\ (G'_V)^n & \xleftarrow[s'_R]{} & G'_R & \xrightarrow[t'_R]{} & (G'_V)^m \end{array}$$

A Σ-hypergraph G is finite if G_V and G_R are finite. Σ-hypergraphs and morphisms form the category $\mathbf{Hyp}_\Sigma$. Its full subcategory of finite Σ-hypergraphs is denoted by $\mathbf{FHyp}_\Sigma$.

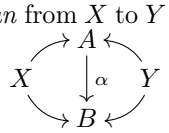
We visualise hypergraphs as follows: • is a vertex and  is a hyperedge with ordered tentacles. An example is shown below left, where $S \in \Sigma_{1,0}$ and $R \in \Sigma_{2,1}$.



(1)

In order to capture GCQ diagrams, we need to equip hypergraphs with interfaces, as illustrated in (1) on the right. Roughly speaking, an interface consists of two sets, called the left boundary and the right boundary. Each has an associated function to the underlying set of hypergraph vertices, depicted by the dotted arrows. Graphical structures with interfaces are common in computer science, (e.g., in automata theory [22], graph rewriting [18], Petri nets [32]). Categorically speaking, they are (discrete) cospans.

▶ **Definition 25** (Cospan)**.** Let $\mathcal{C}$ be a finitely cocomplete category. A *cospan* from X to Y is a pair of arrows $X \to A \leftarrow Y$ in $\mathcal{C}$. A morphism $\alpha\colon (X \to A \leftarrow Y) \Rightarrow (X \to B \leftarrow Y)$ is an arrow $\alpha\colon A \to B$ in $\mathcal{C}$ s.t. the diagram on the right commutes. Cospans $X \to A \leftarrow Y$ and $X \to B \leftarrow Y$ are *isomorphic* if there exists an isomorphism $A \to B$. For $X \in \mathcal{C}$, the *identity cospan* is $X \xrightarrow{\mathrm{id}_X} X \xleftarrow{\mathrm{id}_X} X$. The composition of $X \to A \xleftarrow{f} Y$ and $Y \xrightarrow{g} B \leftarrow Z$ is $X \to A +_{f,g} B \leftarrow Z$, obtained by taking the pushout of f and g. This data is the bicategory [4] $\mathsf{Cospan}(\mathcal{C})$: the objects are those of $\mathcal{C}$, the arrows are cospans and 2-cells are homomorphisms. Finally, $\mathsf{Cospan}(\mathcal{C})$ has monoidal product given by the coproduct in $\mathcal{C}$, with unit the initial object $0 \in \mathcal{C}$.



(2)

To avoid the complications of non-associative composition, it is common to consider a *category* of cospans, where isomorphic cospans are equated: let therefore $\mathsf{Cospan}^{\leq}\mathcal{C}$ be the monoidal category that has isomorphism classes of cospans as arrows. Note that, when going from bicategory to category, after identifying isomorphic arrows it is usual to simply discard the 2-cells. Differently, we consider $\mathsf{Cospan}^{\leq}\mathcal{C}$ to be locally preordered with $(X \to A \leftarrow Y) \leq (X \to B \leftarrow Y)$ if there exists a morphism α going *the other way*: $\alpha\colon (X \to B \leftarrow Y) \Rightarrow (X \to A \leftarrow Y)$. It is an easy exercise to verify that this (pre)ordering is well-defined and compatible with composition and monoidal product. Note that, in general, $\leq$ is a genuine preorder: i.e. it is possible that both $(X \to A \leftarrow Y) \leq (X \to B \leftarrow Y)$ and $(X \to B \leftarrow Y) \leq (X \to A \leftarrow Y)$ without the cospans being isomorphic.

Armed with the requisite definitions, we can be rigorous about hypergraphs with interfaces.

▶ **Definition 26.** The preorder-enriched category $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ is the full subcategory of $\mathsf{Cospan}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ with objects the finite ordinals and arrows (isomorphism classes of) finite hypergraphs, inheriting the preorder. We call its arrows discrete cospans.

The above deserves an explanation: an ordinal n can be considered as the discrete hypergraph with vertices $\{0, \ldots, n-1\}$. An arrow $n \to m$ in $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ is thus a cospan $n \to G \leftarrow m$ where G is a hypergraph and $n \to G$ and $m \to G$ are functions to its vertices. The picture in (1) shows a discrete cospan $3 \to 1$, with dotted lines representing the two morphisms.

6.2 Preordered cartesian bicategories

Here we explore the algebraic structure of $\mathsf{Cospan}^{\leq}\mathcal{C}$. It is closely related to that of cartesian bicategories, yet—given the discussion above—it is more natural to consider $\mathsf{Cospan}^{\leq}\mathcal{C}$ as a locally *pre*ordered category. We therefore need a slight generalisation of Definition 19.

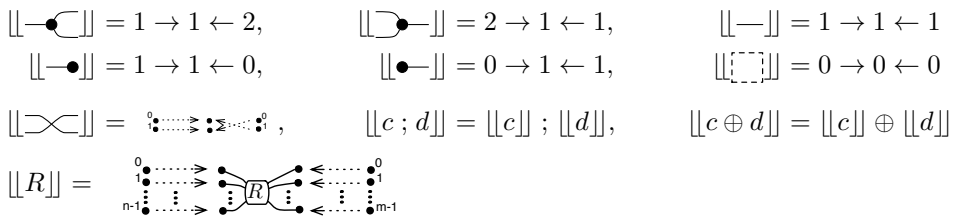


Figure 8 Inductive definition of the isomorphism $\llbracket\cdot\rrbracket : \mathbb{CB}^{\leq}_{\Sigma} \to \mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$. In the first two lines, the finite ordinal n denotes the discrete hypergraph with n vertexes, and the functions between ordinals are uniquely determined by initiality of 0 and finality of 1.

▶ **Definition 27.** A *preordered cartesian bicategory* has the same structure as a cartesian bicategory (Definition 19), with one difference: the ordering is not required to be a partial order, merely a preorder – it is for this reason we separated the equational and inequational theories in Figures 6 and 7. The definition of morphism is as expected.

▶ **Proposition 28.** $\mathsf{Cospan}^{\leq}\mathcal{C}$ *is a preordered cartesian bicategory.*

As a consequence, $\mathsf{Cospan}^{\leq}\,\mathbf{FHyp}_{\Sigma}$, and thus also $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$, are preordered cartesian bicategories. The latter is of particular interest: the main result of this section, Theorem 31, states that $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ is the *free* preordered cartesian bicategory on Σ, defined as follows.

▶ **Definition 29.** The preordered prop $\mathbb{CB}^{\leq}_{\Sigma}$ has, as arrows $n \to m$, GCQ terms of sort (n, m) modulo the smallest congruence generated by $=$ in Figures 5 and 6. These are ordered by the smallest precongruence generated by $\leq$ in Figure 7.

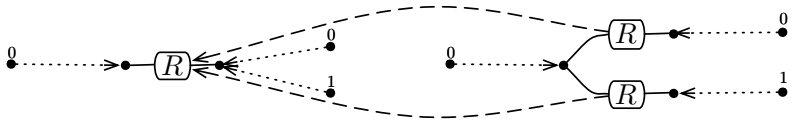
▶ Remark. Intuitively, the ordered prop $\mathbb{CB}_{\Sigma}$ of Definition 21 is the "poset reduction" of the preordered prop $\mathbb{CB}^{\leq}_{\Sigma}$ introduced above. We will make this formal in Section 8.

Theorem 3.3 in [5] states that $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ and $\mathbb{CB}^{\leq}_{\Sigma}$ are isomorphic as mere categories, i.e. forgetting the preorders. We thus need only to prove that the preorder of the two categories coincides, that is for all c, d in $\mathbb{CB}^{\leq}_{\Sigma}$,

$$c \leq d \text{ iff } \llbracket c\rrbracket \leq \llbracket d\rrbracket \tag{3}$$

where $\llbracket\cdot\rrbracket : \mathbb{CB}^{\leq}_{\Sigma} \to \mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$ is the isomorphism from [5] recalled in Figure 8. The 'only-if' part is immediate from Proposition 28. An alternative proof consists of checking, for each of the inclusions $c \leq d$ in Figure 7, that there exists a morphism of cospans from $\llbracket d\rrbracket$ to $\llbracket c\rrbracket$, as illustrated by the following example.

▶ **Example 30.** The left and the right hand side of (L_2) in Figure 7 for $R \in \Sigma_{1,1}$ are translated via $\llbracket\cdot\rrbracket$ into the cospans on the left and right below. The morphism from the rightmost hypergraph to the leftmost one, depicted by the dashed lines, witnesses the preorder.



The 'if' part of (3) requires some work. Its proof is given in full detail in Appendix B.2.

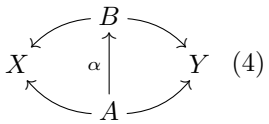
▶ **Theorem 31.** $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma} \cong \mathbb{CB}^{\leq}_{\Sigma}$ *as preordered cartesian bicategories.*

▶ **Example 32.** Recall Example 18. The derivation corresponds via $\llbracket\cdot\rrbracket$ to the homomorphism of cospans of hypergraphs illustrated in the Introduction.

7 Completeness for spans

Having established a combinatorial characterisation of the free preordered cartesian bicategory, here we prove our central completeness result, Theorem 37. In the preordered setting, completeness holds for "multirelational" models: the role of the poset-enriched category **Rel** of sets and relations is taken by a (preorder-enriched) bicategory of spans of functions.

▶ **Definition 33** (Span, Span$^{\leq}$)**.** Given a finitely complete category $\mathcal{C}$, the bicategory $\mathsf{Span}(\mathcal{C})$ is dual to that of cospans of Definition 25: it can be defined as $\mathsf{Cospan}(\mathcal{C}^{\mathrm{op}})$. More explicitly, objects are those of $\mathcal{C}$, arrows of type $X \to Y$ are spans $X \leftarrow A \to Y$, composition ; is defined by pullback and $\oplus$ by categorical product. The 2-cells from $X \leftarrow A \to Y$ to $X \leftarrow B \to Y$ are span homomorphisms, that is arrows $\alpha\colon A \to B$ such that the diagram on the right commutes. As before, the bicategory $\mathsf{Span}(\mathcal{C})$ can be seen as a category by identifying isomorphic spans. We obtain a category $\mathsf{Span}^{\leq}\mathcal{C}$, on which we define a preorder in a similar way to $\mathsf{Cospan}^{\leq}\mathcal{C}$, but in the *reverse* direction: $(X \to A \leftarrow Y) \leq (X \to B \leftarrow Y)$ when there is a homomorphism (4).



(4)

▶ **Lemma 34.** $\mathsf{Span}^{\leq}\mathcal{C}$ *is a preordered cartesian bicategory.*

Models are now morphisms $\mathcal{M}\colon \mathbb{CB}^{\leq}_{\Sigma} \to \mathsf{Span}^{\leq}\,\mathbf{Set}$ of preordered cartesian bicategories. Observe that, since the interpretation of the monoid and comonoid structure is predetermined, a morphism is uniquely determined by its value on the object 1 and on $R \in \Sigma$. In other words, a model consists of a set $\mathcal{M}(1)$ and, for each $R \in \Sigma_{n,m}$, a span $\mathcal{M}(1)^n \leftarrow Y \to \mathcal{M}(1)^m$. This data is exactly the definition of a (possibly infinite) Σ-hypergraph (Definition 24).

▶ **Proposition 35.** *Morphisms* $\mathcal{M}\colon \mathbb{CB}^{\leq}_{\Sigma} \to \mathsf{Span}^{\leq}\,\mathbf{Set}$ *are in bijective correspondence with* Σ*-hypergraphs.*

Given this correspondence and the fact that $\mathbb{CB}^{\leq}_{\Sigma} \cong \mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$, each hypergraph G induces a morphism $\mathcal{U}_G\colon \mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma} \to \mathsf{Span}^{\leq}\,\mathbf{Set}$. Moreover, G acts like a representing object of a contravariant Hom-functor, in the following sense: $\mathcal{U}_G$ maps $n \xrightarrow{\iota} G' \xleftarrow{\omega} m$ to

$$\mathbf{Hyp}_{\Sigma}[n, G] \xleftarrow{\iota;-} \mathbf{Hyp}_{\Sigma}[G', G] \xrightarrow{\omega;-} \mathbf{Hyp}_{\Sigma}[m, G]$$

where $\mathbf{Hyp}_{\Sigma}[G', G]$ is the set of hypergraph homomorphisms from G' to G, and $(\iota\,;-)$ and $(\omega\,;-)$ are defined, given $f \in \mathbf{Hyp}_{\Sigma}[G', G]$, by $(\iota\,;-)(f) = \iota\,;f$ and $(\omega\,;-)(f) = \omega\,;f$.

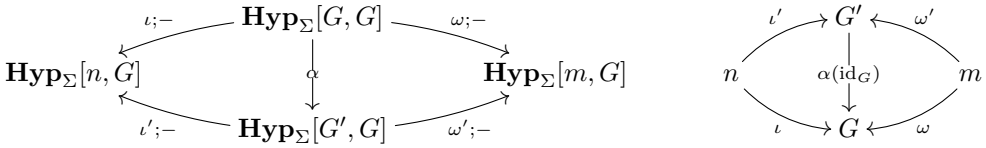
▶ **Proposition 36.** *Suppose that* $n \xrightarrow{\iota} G' \xleftarrow{\omega} m$ *a discrete cospan in* $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$*. Then*

$$\mathcal{U}_G(n \xrightarrow{\iota} G' \xleftarrow{\omega} m) = \mathbf{Hyp}_{\Sigma}[n, G] \xleftarrow{\iota;-} \mathbf{Hyp}_{\Sigma}[G', G] \xrightarrow{\omega;-} \mathbf{Hyp}_{\Sigma}[m, G].$$

Proof. The conclusion of Theorem 31 allows us to use induction on $n \xrightarrow{\iota} G' \xleftarrow{\omega} m$. The inductive cases follow since the contravariant Hom-functor sends colimits to limits. Four of the base cases, $\llbracket$ $\rrbracket$, $\llbracket$  $\rrbracket$, $\llbracket$  $\rrbracket$ and $\llbracket$ $\rrbracket$, follow by the same argument, and the others ($\llbracket$  $\rrbracket$, $\llbracket$  $\rrbracket$ and $\llbracket R \rrbracket$) are easy to check. The details are in Appendix B.3. ◀

▶ **Theorem 37** (Completeness for $\mathsf{Span}^{\leq}\,\mathbf{Set}$)**.** *Let* $n \xrightarrow{\iota} G \xleftarrow{\omega} m$ *and* $n \xrightarrow{\iota'} G' \xleftarrow{\omega'} m$ *be arrows in* $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma}$*. If, for all morphisms* $\mathcal{M}\colon \mathsf{Csp}^{\leq}\,\mathbf{FHyp}_{\Sigma} \to \mathsf{Span}^{\leq}\,\mathbf{Set}$*, we have* $\mathcal{M}(n \xrightarrow{\iota} G \xleftarrow{\omega} m) \leq \mathcal{M}(n \xrightarrow{\iota'} G' \xleftarrow{\omega'} m)$*, then* $(n \xrightarrow{\iota} G \xleftarrow{\omega} m) \leq (n \xrightarrow{\iota'} G' \xleftarrow{\omega'} m)$*.*

Proof. If the inequality holds for all morphisms, it holds for $\mathcal{U}_G$. By the conclusion of Proposition 36, there is a function $\alpha\colon \mathbf{Hyp}_\Sigma[G,G] \to \mathbf{Hyp}_\Sigma[G',G]$ making the diagram on



the left commute. We take the identity $\mathrm{id}_G \in \mathbf{Hyp}_\Sigma[G,G]$ and consider $\alpha(\mathrm{id}_G)\colon G' \to G$. By the commutativity of the left diagram, we have that $\iota = \iota' \,;\, \alpha(\mathrm{id}_G)$ and $\omega = \omega' \,;\, \alpha(\mathrm{id}_G)$. This means that the right diagram commutes, that is $(n \xrightarrow{\iota} G \xleftarrow{\omega} m) \leq (n \xrightarrow{\iota'} G' \xleftarrow{\omega'} m)$. ◀

▶ Remark. The reader may have noticed that, in the above proof, $\mathcal{U}_G$ plays a role analogous to Chandra and Merlin's [14] *natural model* for the formula corresponding to $n \xrightarrow{\iota} G \xleftarrow{\omega} m$.

Given the completeness theorem of this section, proving completeness for models of $\mathbb{CB}_\Sigma$ in **Rel** is a simple step that we illustrate in the next section.

8 Completeness for relations

We conclude by showing how Theorem 37 leads to a proof of Theorem 17. The key observation lies in the tight connection between the preordered setting and the posetal one.

▶ **Definition 38.** Let $\mathcal{C}$ be a preorder-enriched category. The poset-reduction of $\mathcal{C}$ is the category $\mathcal{C}^\sim$ having the same objects as $\mathcal{C}$ and morphisms in $\mathcal{C}^\sim$ are equivalence classes of those in $\mathcal{C}$ modulo $\sim = \leq \cap \geq$. Composition is inherited from $\mathcal{C}$; this is well-defined as $\sim$ is a congruence wrt composition.

This assignment extends to a functor $(\cdot)^\sim$ from the category of preorder-enriched categories and functors to the category of poset-enriched ones. See Appendix B.4 for details.

We have already seen, although implicitly, an example of this construction in passing from $\mathbb{CB}^{\leq}_\Sigma$ (Definition 29) to $\mathbb{CB}_\Sigma$ (Definition 21): it is indeed immediate to see that $\left(\mathbb{CB}^{\leq}_\Sigma\right)^\sim = \mathbb{CB}_\Sigma$. Another crucial instance is provided by the following observation, where $\mathsf{Span}^\sim\mathcal{C}$ is a shorthand for $\left(\mathsf{Span}^{\leq}\mathcal{C}\right)^\sim$.

▶ **Proposition 39.** $\mathsf{Span}^\sim\,\mathbf{Set} \cong \mathbf{Rel}$ *as cartesian bicategories.*

The above proposition implicitly makes use of the following fact.

▶ **Proposition 40.** *The functor $(\cdot)^\sim$ maps preorder-enriched cartesian bicategories and morphisms into poset-enriched cartesian bicategories and morphisms.*

To conclude, it is convenient to establish a general theory of completeness results.

▶ **Definition 41.** Let $\mathcal{C}, \mathcal{D}$ be preorder-enriched categories and let $\mathcal{F}$ be a class of preordered functors $\mathcal{C} \to \mathcal{D}$. We say that $\mathcal{C}$ is $\mathcal{F}$-complete for $\mathcal{D}$ if for all arrows x, y in $\mathcal{C}$, $\mathcal{M}(x) \leq \mathcal{M}(y)$ for all $\mathcal{M} \in \mathcal{F}$ entails that $x \leq y$.

▶ **Lemma 42** (Transfer lemma). *Let $\mathcal{C}, \mathcal{D}$ be preorder-enriched categories and $\mathcal{F}$ a class of preordered functors $\mathcal{C} \to \mathcal{D}$. Assume $\mathcal{C}$ to be $\mathcal{F}$-complete for $\mathcal{D}$.*

1. *Then $\mathcal{C}^\sim$ is $\mathcal{F}^\sim$-complete for $\mathcal{D}^\sim$, where $\mathcal{F}^\sim = \{F^\sim \mid F \in \mathcal{F}\}$.*
2. *If $\mathcal{F} \subseteq \mathcal{F}'$, then $\mathcal{C}$ is $\mathcal{F}'$-complete for $\mathcal{D}$.*

All the pieces are now in place for a

Proof of Theorem 17. We need to show completeness—that is—assuming $c \leqq c'$, we need to prove $c \leq_{\mathbb{CB}_\Sigma} c'$ for all GCQ terms c and c'. Observe that $c \leq_{\mathbb{CB}_\Sigma} c'$ if and only if

$$c \leq c' \text{ as arrows of } \mathbb{CB}_\Sigma \text{ (Definition 21).} \tag{†}$$

Moreover, using Proposition 23, $c \leqq c'$ iff

$$\mathcal{M}c \leq \mathcal{M}c', \text{ for all morphisms of cartesian bicategories } \mathcal{M} \colon \mathbb{CB}_\Sigma \to \mathbf{Rel}\,. \tag{‡}$$

Our task becomes, therefore, to show that (‡) implies (†). In other words, we need to prove $\mathbb{CB}_\Sigma$ to be $\mathcal{G}$-complete for **Rel**, where $\mathcal{G}$ is the class of morphisms of cartesian bicategories of type : $\mathbb{CB}_\Sigma \to \mathbf{Rel}$. Let $\mathcal{F}$ be the class of morphisms of preorder-enriched cartesian bicategories from $\mathbb{CB}^{\leqq}_\Sigma$ to $\mathsf{Span}^{\leq}\,\mathbf{Set}$. Since, by Theorem 37, $\mathbb{CB}^{\leqq}_\Sigma$ is $\mathcal{F}$-complete for $\mathsf{Span}^{\leq}\,\mathbf{Set}$, we can conclude by Lemma 42.1 that $\left(\mathbb{CB}^{\leqq}_\Sigma\right)^{\sim}$ is $\mathcal{F}^{\sim}$-complete for $\left(\mathsf{Span}^{\leq}\,\mathbf{Set}\right)^{\sim}$. By Proposition 39, this is equivalent to $\mathbb{CB}_\Sigma$ being $\mathcal{F}^{\sim}$-complete for **Rel**. Now, by Proposition 40 $\mathcal{F}^{\sim} \subseteq \mathcal{G}$, so the claim follows by Lemma 42.2. ◀

9 Discussion, related and future work

We introduced a string diagrammatic language for conjunctive queries and demonstrated a sound and complete axiomatisation for query equivalence and inclusion. To prove completeness, we showed that our language provides an algebra able to express all hypergraphs and that our axioms characterise both hypergraph isomorphisms and existence of hypergraph morphisms. A recent result [19] introduced an extension of the allegorical fragment of the algebra of relations [33] that is able to express all graphs with tree-width at most 2. Furthermore, the isomorphism of these graphs can be axiomatised. The algebra in [19], which is clearly less expressive than ours, can be elegantly encoded into our string diagrams. The same holds for the representable allegories by Freyd and Scedrov [20].

We also prove completeness with respect to $\mathsf{Span}^{\leq}\,\mathbf{Set}$, the structure of which is closely related to the *bag semantics* of conjunctive queries in SQL. Indeed, the join of two SQL-tables is given by composition in $\mathsf{Span}^{\leq}\,\mathbf{Set}$ and not in **Rel**: in the resulting table the same row can occur several times. As we have seen, with the relational semantics, query inclusion can be decided with Chandra and Merlin's algorithm [14] and its reduction to existence of a hypergraph homomorphism. On the other hand, decidability of inclusion for the bag semantic is, famously, open. Originally posed by Vardi and Chaudhuri [15], it has been studied for different fragments and extensions of conjunctive queries [23, 2, 24]. It is worth mentioning that it is known [26] that there is a reduction to the homomorphism domination problem, which seems intimately related with our Proposition 36. Unfortunately, the preorder in $\mathsf{Span}^{\leq}\,\mathbf{Set}$—the existence of a span morphism—does *not* directly correspond to bag inclusion: one must restrict to the existence of an *injective* morphism. We leave this promising path for future work.

References

1 Samson Abramsky and Bob Coecke. Categorical quantum mechanics. *CoRR*, abs/1401.4973, 2008.

2 Foto N. Afrati, Matthew Damigos, and Manolis Gergatsoulis. Query containment under bag and bag-set semantics. *Inf. Process. Lett.*, 110(10):360–369, 2010. URL: `https://doi.org/10.1016/j.ipl.2010.02.017`, `doi:10.1016/j.ipl.2010.02.017`.

3 John Baez and Jason Erbele. Categories in control. *Theory and Application of Categories*, 30:836–881, 2015.

4 Jean Bénabou. Introduction to bicategories. In *Reports of the Midwest Category Seminar*, pages 1–77. Springer, 1967.

5 Filippo Bonchi, Fabio Gadducci, Aleks Kissinger, Paweł Sobociński, and Fabio Zanasi. Rewriting modulo symmetric monoidal structure. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 710–719. ACM, 2016.

6 Filippo Bonchi, Pawel Sobocinski, and Fabio Zanasi. Full abstraction for signal flow graphs. In *POPL 2015*, pages 515–526. ACM, 2015.

7 Roberto Bruni, Ivan Lanese, and Ugo Montanari. A basic algebra of stateless connectors. *Theoretical Computer Science*, 366(1–2):98–120, 2006.

8 Roberto Bruni, Hernán C. Melgratti, and Ugo Montanari. A connector algebra for P/T nets interactions. In *CONCUR 2011*, volume 6901 of *LNCS*, pages 312–326. Springer, 2011. doi:10.1093/jigpal/6.2.349.

9 Roberto Bruni, Hernán C. Melgratti, Ugo Montanari, and Paweł Sobociński. Connector algebras for C/E and P/T nets' interactions. *Log Meth Comput Sci*, 9(16), 2013.

10 Roberto Bruni, Ugo Montanari, Gordon D. Plotkin, and Daniele Terreni. On hierarchical graphs: Reconciling bigraphs, gs-monoidal theories and gs-graphs. *Fundam. Inform.*, 134(3-4):287–317, 2014.

11 Aurelio Carboni, G Max Kelly, Robert FC Walters, and Richard J Wood. Cartesian bicategories ii. *Theory and Applications of Categories*, 19(6):93–124, 2008.

12 Aurelio Carboni and Robert FC Walters. Cartesian bicategories i. *Journal of pure and applied algebra*, 49(1-2):11–32, 1987.

13 Donald D Chamberlin and Raymond F Boyce. Sequel: A structured english query language. In *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) workshop on Data description, access and control*, pages 249–264. ACM, 1974.

14 Ashok K Chandra and Philip M Merlin. Optimal implementation of conjunctive queries in relational data bases. In *Proceedings of the ninth annual ACM symposium on Theory of computing*, pages 77–90. ACM, 1977.

15 Surajit Chaudhuri and Moshe Y. Vardi. Optimization of *Real* conjunctive queries. In Catriel Beeri, editor, *Proceedings of the Twelfth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 25-28, 1993, Washington, DC, USA*, pages 59–70. ACM Press, 1993. URL: http://doi.acm.org/10.1145/153850.153856, doi:10.1145/153850.153856.

16 Edgar F Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

17 B. Coecke and A. Kissinger. *Picturing Quantum Processes. A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2016.

18 A. Corradini, U. Montanari, F. Rossi, H. Ehrig, R. Heckel, and M. Loewe. Algebraic approaches to graph transformation, part i: Basic concepts and double pushout approach. In *Handbook of Graph Grammars*, pages 163–246. 1997.

19 Enric Cosme-Llópez and Damien Pous. K4-free graphs as a free algebra. 2017.

20 Peter J Freyd and Andre Scedrov. *Categories, allegories*, volume 39. Elsevier, 1990.

21 Dan R. Ghica and Aliaume Lopez. A structural and nominal syntax for diagrams. *CoRR*, abs/1702.01695, 2017. URL: http://arxiv.org/abs/1702.01695, arXiv:1702.01695.

22 Victor Mikhaylovich Glushkov. The abstract theory of automata. *Russian Mathematical Surveys*, 16(5):1, 1961.

23 Yannis E. Ioannidis and Raghu Ramakrishnan. Containment of conjunctive queries: Beyond relations as sets. *ACM Trans. Database Syst.*, 20(3):288–324, 1995. URL: http://doi.acm.org/10.1145/211414.211419, doi:10.1145/211414.211419.

24 T. S. Jayram, Phokion G. Kolaitis, and Erik Vee. The containment problem for REAL conjunctive queries with inequalities. In Stijn Vansummeren, editor, *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 26-28, 2006, Chicago, Illinois, USA*, pages 80–89. ACM, 2006. URL: `http://doi.acm.org/10.1145/1142351.1142363`, `doi:10.1145/1142351.1142363`.

25 Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart. A complete axiomatisation of the zx-calculus for clifford+ t quantum mechanics. *arXiv preprint arXiv:1705.11151*, 2017.

26 Swastik Kopparty and Benjamin Rossman. The homomorphism domination exponent. *Eur. J. Comb.*, 32(7):1097–1114, 2011. URL: `https://doi.org/10.1016/j.ejc.2011.03.009`, `doi:10.1016/j.ejc.2011.03.009`.

27 Stephen Lack. Composing props. *Theory and Applications of Categories*, 13(9):147–163, 2004.

28 Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 2013.

29 Paul-André Melliès and Noam Zeilberger. A bifibrational reconstruction of lawvere's presheaf hyperdoctrine. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, pages 555–564, 2016. URL: `http://doi.acm.org/10.1145/2933575.2934525`, `doi:10.1145/2933575.2934525`.

30 Kang Feng Ng and Quanlong Wang. A universal completion of the zx-calculus. *arXiv preprint arXiv:1706.09877*, 2017.

31 Mehrnoosh Sadrzadeh, Stephen Clark, and Bob Coecke. The frobenius anatomy of word meanings I: subject and object relative pronouns. *CoRR*, abs/1404.5278, 2014.

32 Vladimiro Sassone and Paweł Sobociński. A congruence for Petri nets. *Electronic Notes in Theoretical Computer Science*, 127(2):107–120, 2005.

33 Alfred Tarski. On the calculus of relations. *The Journal of Symbolic Logic*, 6(3):73–89, 1941.

A A translation from GCQ to CCQ

To translate GCQ diagrams to CCQ formulas we need to introduce a minor syntactic variant of CCQ, this time assuming two countable sets of variables $Var_l = \{x_i \mid i \in \mathbb{N}\}$ and $Var_r = \{y_i \mid i \in \mathbb{N}\}$. The idea is that a diagram $c : (n, m)$ will translate to a formula that has its free variables in $\{x_0, \ldots, x_{n-1}\} \cup \{y_0, \ldots, y_{m-1}\}$, i.e. there are "left" free variables $\vec{x}$ and "right" free variables $\vec{y}$.

▶ **Definition 43.** We write $n, m \vdash \phi$ if $fr(\phi) \subseteq \{x_0, \ldots, x_{n-1}\} \cup \{y_0, \ldots, y_{m-1}\}$ and $n+m \vdash \phi[x_{[n,n+m-1]}/y_{[0,m-1]}]$.

Next, for $R \in \Sigma_{n,m}$ we assume a CCQ signature in which R is a relation symbol with arity $n + m$. Then, given a GCQ model $\mathcal{M} = (X, \rho)$ we can obtain a CCQ model $\Lambda(\mathcal{M}) = (X, \rho')$ in the obvious way. With this in place, we can give a recursive translation Λ from GCQ terms to CCQ formulas. The details are given in Figure 9. and Λ preserves the semantics:

▶ **Proposition 44.** *Fix a GCQ model $\mathcal{M} = (X, \rho)$ and suppose that $c : (n, m)$ is a GCQ formula. Then $(\vec{v}, \vec{w}) \in [\![c : (n, m)]\!]_{\mathcal{M}}$ iff $(\vec{v}, \vec{w}) \in [\![n + m \vdash \Lambda(c)]\!]_{\Lambda(\mathcal{M})}$.*

Proof. Induction on the derivation of $c : (n, m)$. ◀

The following is immediate from the definition of the translations Θ and Λ.

▶ **Lemma 45.** *Let $n \vdash \phi$ and $\mathcal{M}$ be a CCQ model. Then $[\![n \vdash \phi]\!]_{\mathcal{M}} = [\![\Lambda\Theta(n \vdash \phi)]\!]_{\Lambda\Theta(\mathcal{M})}$.* ◀

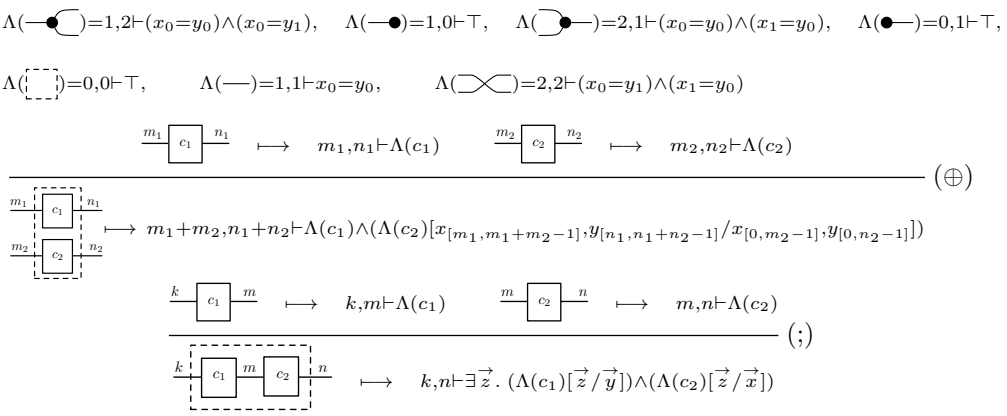


Figure 9 Translation Λ from GCQ to CCQ.

▶ **Example 46.** An interesting case is  . It is the translation, via Θ, of $2 \vdash x_0 = x_1$. Returning to CCQ via Λ, we obtain $2, 0 \vdash \exists z.\ (x_0 = z) \wedge (x_1 = z) \wedge \top$. The formulas are quite different, but they are logically equivalent. The case of GCQ terms  and  is also interesting. The first translates via Λ to $0, 0 \vdash \top$, the second to $0, 0 \vdash \exists z_0. \top \wedge \top$.

B Proofs

B.1 Proofs of Sections 2, 3, 4, and 5

Proof of Proposition 2. The 'only if' direction is a straightforward induction on the derivation of a formula generated by (CCQ). The 'if' is a trivial induction on derivations obtained from $\{(\top), (R), (\exists), (=), (\wedge), (\mathsf{Sw}_{n,k}), (\mathsf{Id}_n), (\mathsf{Nu}_n)\}$. ◀

Proof of Proposition 8. Easy induction on the derivation of $n \vdash \phi$. ◀

Proof of Proposition 9 and 10. The translation is given in Figure 9. The rest follows easily from Propositions 8 and 44 and Lemma 45. ◀

Proof of Lemma 12. Follows immediately from the definition of semantics and relational composition / tensor in Figure 3. ◀

Proof of Proposition 13. For each fixed model the axioms of Figure 5 are satisfied because the category of relations with monoidal product $\times$ is symmetric monoidal. ◀

Proof of Lemma 22. For every object n, the monoid and comonoid are given by  ,  ,  and  , standing for the "stacking" of n of these diagrams respectively in the usual manner. An easy induction shows that these satisfy the required laws. The definition of $\mathbb{CB}_\Sigma$ asserts that for every $R \in \Sigma_{n,m}$,  is a lax comonoid morphism in $\mathbb{CB}_\Sigma$, but cartesian bicategories require this for *all* arrows. This follows by another induction. ◀

Proof of Lemma 23. In the easy direction, to extract a model $\mathcal{M} = (X, \rho)$ from a morphism of cartesian bicategories $\mathcal{F} \colon \mathbb{CB}_\Sigma \to \mathbf{Rel}$, define $X := \mathcal{F}(1)$ and let $\rho(R) := \mathcal{F}(R)$ for $R \in \Sigma$.

Conversely, given a model $\mathcal{M} = (X, \rho)$, we observe that the semantics map $[\![\cdot]\!]_{\mathcal{M}}$ (Figure 3) gives rise to a morphism of cartesian bicategories $[\![\cdot]\!]_{\mathcal{M}} \colon \mathbb{CB}_\Sigma \to \mathbf{Rel}$. To prove that it is well defined and preserves the ordering, one can easily see that the axioms of $=_{\mathbb{CB}_\Sigma}$ and $\leq_{\mathbb{CB}_\Sigma}$ are sound. By the inductive definition, $[\![\cdot]\!]_{\mathcal{M}}$ preserves composition ; and tensor $\oplus$. Finally, we observe that, by definition, $[\![\cdot]\!]_{\mathcal{M}}$ maps the monoids and comonoids of $\mathbb{CB}_\Sigma$ into those of **Rel**. ◀

B.2 Proofs of Section 6

We first prove Proposition 28 and then we work towards a proof of Theorem 31.

Proof of Proposition 28. We endow every object with a monoid and comonoid structure, prove these structures to be adjoint and satisfy the special Frobenius property.

1. Define the monoid/comonoid structure on every object: Considering $\mathcal{C}$ as a cocartesian monoidal category via its coproduct, it is well known, that every object comes with a natural monoid structure. There is a monoidal functor $F \colon \mathcal{C} \to \mathsf{Cospan}^{\leq}\mathcal{C}$ sending $f \colon X \to Y$ to the cospan $X \xrightarrow{f} Y \xleftarrow{\mathrm{id}} Y$ and mapping the natural monoid structure on every object through F yields the monoid structures in $\mathsf{Cospan}^{\leq}\mathcal{C}$. Furthermore, there is a duality operation $\bullet^{\mathrm{op}}$ on $\mathsf{Cospan}^{\leq}\mathcal{C}$ given by mapping $X \xrightarrow{f} Z \xleftarrow{g} Y$ to $Y \xrightarrow{g} Z \xleftarrow{f} X$. Now define the comonoid structure on every object as the dual of the monoid. It is easy to see that every morphism in $\mathsf{Cospan}^{\leq}\mathcal{C}$ is a lax comonoid homomorphism, which follows from the fact that every morphism in $\mathcal{C}$ preserves the monoid structure.
2. The monoid and comonoid structures are adjoint: In general, for $f \colon X \to Y$ a morphism in $\mathcal{C}$, we have $F(f) \mathbin{;} F(f)^{\mathrm{op}} \leq \mathrm{id}_X$ and $\mathrm{id}_Y \leq F(f)^{\mathrm{op}} \mathbin{;} F(f)$. This follows easily from the definition and implies the adjointness of monoid and comonoid.
3. The monoid and comonoid enjoy the special Frobenius property: The Frobenius law is a consequence of associativity of the natural monoid and its definition. The special law follows from the multiplication being epi.

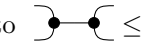
◀

We will prove in Theorem 31 that $\mathsf{Csp}^{\leq}\,\mathbf{FHyp}_\Sigma \cong \mathbb{CB}^{\leq}_\Sigma$. It is convenient to begin with $\Sigma = \varnothing$. Consider the category $\mathbb{F}$: objects are finite ordinals $n = \{0, \ldots, n-1\}$ and arrows all functions. Then $\mathsf{Cospan}^{\leq}\mathbb{F}$ is the free preordered cartesian bicategory on the empty signature.

▶ **Theorem 47.** $\mathsf{Cospan}^{\leq}\mathbb{F} \cong \mathbb{CB}^{\leq}_\emptyset$ *as preordered cartesian bicategories.*

Proof. The translation in Figure 8 defines an isomorphism $\lfloor\!\lfloor \cdot \rfloor\!\rfloor \colon \mathbb{CB}^{\leq}_\emptyset \to \mathsf{Cospan}^{\leq}\mathbb{F}$ (first three lines). The translation $\lceil\!\lceil \cdot \rceil\!\rceil \colon \mathsf{Cospan}^{\leq}\mathbb{F} \to \mathbb{CB}^{\leq}_\emptyset$ can be found in [5, Theorem 3.3], where it is proved that it defines an isomorphism between categories, i.e. forgetting the ordering. It thus suffices to prove, that both translations preserve the ordering. For $c, d \in \mathbb{CB}^{\leq}_\emptyset$, we have $c \leq d$ implies $\lfloor\!\lfloor c \rfloor\!\rfloor \leq \lfloor\!\lfloor d \rfloor\!\rfloor$ by Proposition 28. Consider a morphism of cospans $n \to S \xleftarrow{\alpha} m \to T \leftarrow m$.

We want to prove $\lceil\!\lceil n \to T \leftarrow m \rceil\!\rceil \leq \lceil\!\lceil n \to S \leftarrow m \rceil\!\rceil$. Since every function $\alpha \colon S \to T$ can be decomposed into sums and compositions of $2 \to 1$ and $0 \to 1$ as demonstrated for example in [28, VII.5], we can consider only these cases. In the case $\alpha \colon 0 \to 1$, we have $n = m = 0$ and we have to prove $\bullet\!\!-\!\!\bullet \leq$  which is axiom (UC). The case $\alpha \colon 2 \to 1$, can be

further reduced by the following observation: Given a diagram $n_1 + n_2 \to 2 \leftarrow m_1 + m_2$, one easily computes the composite of spans $n_1 + n_2 \longrightarrow 2 \xleftarrow{\text{id}} 2 \xrightarrow{\text{id}} 2 \xleftarrow{\text{id}} 2 \xrightarrow{\text{id}} 2 \longleftarrow m_1 + m_2$ to be $n_1 + n_2 \to 2 \leftarrow m_1 + m_2$ and the composite $n_1 + n_2 \longrightarrow 2 \xleftarrow{\text{id}} 2 \longrightarrow 1 \longleftarrow 2 \xrightarrow{\text{id}} 2 \longleftarrow m_1 + m_2$ to be $n_1 + n_2 \to 1 \leftarrow m_1 + m_2$. By compositionality, it suffices to consider the case $2 \to 2 \to 1 \leftarrow 2 \leftarrow 2$ which corresponds to  which is axiom (MC). ◀

A cospan of hypergraphs is said to be *disconnected* if it is of the form $\lfloor\!\lfloor R_0 \rfloor\!\rfloor \oplus \lfloor\!\lfloor R_1 \rfloor\!\rfloor \oplus \ldots \lfloor\!\lfloor R_n \rfloor\!\rfloor$ for $R_0, \ldots R_n \in \Sigma$.

▶ **Lemma 48.** *Let $n \xrightarrow{\iota} E \xleftarrow{\omega} m$ and $n' \xrightarrow{\iota'} E' \xleftarrow{\omega'} m'$ be disconnected cospans. If there are functions $f\colon n \to n'$, $g\colon m \to m'$ and $h\colon E \to E'$ s.t. the following commutes*

$$\begin{array}{ccccc} n & \xrightarrow{\quad\iota\quad} & E & \xleftarrow{\quad\omega\quad} & m \\ {\scriptstyle f}\downarrow & & {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle g} \\ n' & \xrightarrow[\quad\iota'\quad]{} & E' & \xleftarrow[\quad\omega'\quad]{} & m' \end{array} \tag{5}$$

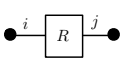
then $\lceil\!\lceil n \xrightarrow{f;\iota'} E' \xleftarrow{g;\omega'} m \rceil\!\rceil \leq \lceil\!\lceil n \xrightarrow{\iota} E \xleftarrow{\omega} m \rceil\!\rceil$.

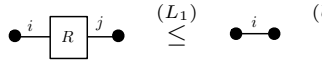
Proof. First note that in the case of disconnected cospans, h uniquely determines f and g. To give a hypergraph homomorphism $h\colon E \to E'$ is the same as giving a label-preserving function between their sets of hyperedges, so we identify E and E' with their sets of hyperedges. We can now consider the labels separately, so assume to have only one label. Furthermore, we can consider each fiber over elements of E' separately, so assume $E' = 1$. So $n' \to E' \leftarrow m'$ consist of a single hyperedge with label $R \in \Sigma_{i,j}$, yielding

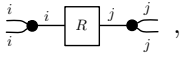
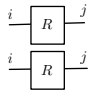
$\lceil\!\lceil n' \xrightarrow{\iota'} E' \xleftarrow{\omega'} m' \rceil\!\rceil =$

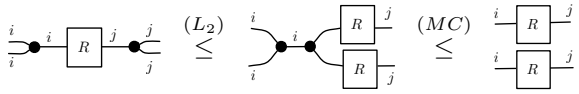
and thus $n' = i$ and $m' = j$. It now suffices to consider cases where the size of E is either 0 or 2, yielding diagrams $\begin{array}{ccccc} 0 & \longrightarrow & 0 & \longleftarrow & 0 \\ {\scriptstyle i}\downarrow & & {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i} \\ i & \xrightarrow[\iota']{} & 1 & \xleftarrow[\omega']{} & j \end{array}$ (6) and $\begin{array}{ccccc} i+i & \xrightarrow{\iota'+\iota'} & 2 & \xleftarrow{\omega'+\omega'} & j+j \\ {\scriptstyle\nabla}\downarrow & & {\scriptstyle !}\downarrow & & \downarrow{\scriptstyle\nabla} \\ i & \xrightarrow[\iota']{} & 1 & \xleftarrow[\omega']{} & j \end{array}$. (7)

The result for $|E| \geq 2$ can be obtained inductively from these base cases. For (6),

$\lceil\!\lceil 0 \to 0 \leftarrow 0 \rceil\!\rceil$ =  , and $\lceil\!\lceil 0 \to 1 \leftarrow 0 \rceil\!\rceil$ =  .

The following derivation thus suffices:  $\overset{(L_1)}{\leq}$ $\overset{(UC)}{\leq}$  . For (7),

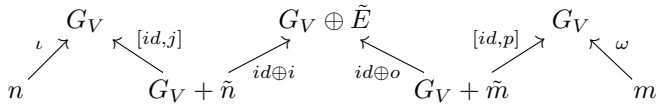
$\lceil\!\lceil i + i \to 1 \leftarrow j + j \rceil\!\rceil$ =  , $\lceil\!\lceil i + i \xrightarrow{\iota'+\iota'} 2 \xleftarrow{\omega'+\omega'} j + j \rceil\!\rceil$ =  . This

derivation thus completes the proof: 

◀

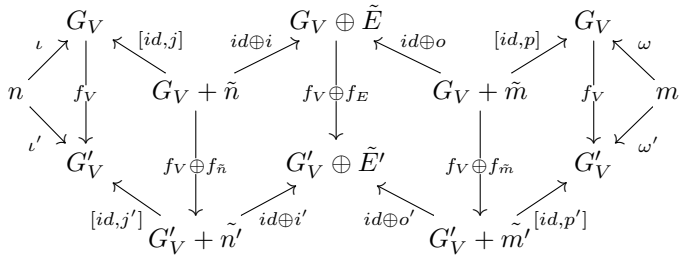
We have now all the ingredients to prove Theorem 31.

Proof of Theorem 31. The proof relies on a result appearing in the proof of Theorem 3.3 in [5]: every discrete cospan of hypergraphs $n \xrightarrow{\iota} G \xleftarrow{\omega} m$ can be written as the composition



where $\tilde{n} \xrightarrow{i} \tilde{E} \xleftarrow{o} \tilde{m}$ is disconnected, G_V is the set of vertices of G[2], $j\colon \tilde{n} \to G_V$ and $j\colon \tilde{m} \to G_V$ maps the vertices of $\tilde{n} \to \tilde{E} \leftarrow \tilde{m}$ into those of G. We only need to prove the right-to-left implication of (3). We will show that if $n \to G' \leftarrow m \leq n \to G \leftarrow m$ then $\llbracket n \to G' \leftarrow m \rrbracket \leq \llbracket n \to G \leftarrow m \rrbracket$.

Assume now that $n \xrightarrow{\iota'} G' \xleftarrow{\omega'} m \leq n \xrightarrow{\iota} G \xleftarrow{\omega} m$, i.e., there exists an $f\colon G \to G'$ such that $f\iota = \iota'$ and $f\omega = \omega'$. The morphism f induces $f_V\colon G_V \to G'_V$, $f_E\colon \tilde{E} \to \tilde{E}'$, $f_{\tilde{n}}\colon \tilde{n} \to \tilde{n'}$ and $f_{\tilde{m}}\colon \tilde{m} \to \tilde{m'}$ making the following commute.



From the commutativity of the above diagram, one has:

$$\begin{array}{lrcll}
(\gamma_1 :=) & n \to G'_V \leftarrow G_V + \tilde{n} & \leq & n \to G_V \leftarrow G_V + \tilde{n} & (=: \delta_1) \\
(\gamma_2 :=) & G_V + \tilde{m} \to G'_V \leftarrow m & \leq & G_V + \tilde{m} \to G_V \leftarrow m & (=: \delta_2) \\
(\gamma_3 :=) & G_V \to G'_V \leftarrow G_V & \leq & G_V \to G_V \leftarrow G_V & (=: \delta_3) \\
(\gamma_4 :=) & \tilde{n} \to \tilde{E}' \leftarrow \tilde{m} & \leq & \tilde{n} \to \tilde{E} \leftarrow \tilde{m} & (=: \delta_4)
\end{array}$$

Since the first three inequations only involve sets and functions, one can use the conclusion of Theorem 47 and deduce that: $\llbracket \gamma_i \rrbracket \leq \llbracket \delta_i \rrbracket$ for $i \in \{1,2,3\}$. From the fourth inequation, via Lemma 48, one obtains furthermore $\llbracket \gamma_4 \rrbracket \leq \llbracket \delta_4 \rrbracket$ and concludes as follows.

$$\begin{aligned}
&\llbracket n \to G' \leftarrow m \rrbracket = \llbracket \gamma_1 \,;\, (\gamma_3 \oplus \gamma_4) \,;\, \gamma_2 \rrbracket = \llbracket \gamma_1 \rrbracket \,;\, (\llbracket \gamma_3 \rrbracket \oplus \llbracket \gamma_4 \rrbracket) \,;\, \llbracket \gamma_2 \rrbracket \\
&\leq \llbracket \delta_1 \rrbracket \,;\, (\llbracket \delta_3 \rrbracket \oplus \llbracket \delta_4 \rrbracket) \,;\, \llbracket \delta_2 \rrbracket = \llbracket \delta_1 \,;\, (\delta_3 \oplus \delta_4) \,;\, \delta_2 \rrbracket = \llbracket n \to G \leftarrow m \rrbracket
\end{aligned}$$

◀

[2] Since cospans are taken up-to isomorphism and since G is finite one can always assume, without loss of generality, that G_V is a finite ordinal.

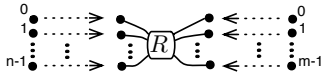
B.3 Proofs of Section 7

Proof of Lemma 34. Immediate from Proposition 28 by duality. ◀

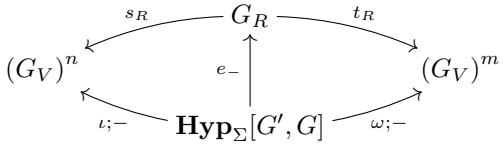
Proof of Proposition 35. As stated in the main text, $\mathcal{M}$ is uniquely determined by the set $\mathcal{M}(1)$ and, for each $R \in \Sigma_{n,m}$, a span $\mathcal{M}(R)\colon \mathcal{M}(1)^n \to \mathcal{M}(1)^m$. This data is that of a (possibly infinite) hypergraph (Definition 24). ◀

Proof of Proposition 36. By definition, $\mathcal{U}_G(1) = G_V$ and $\mathcal{U}_G(\llparenthesis R \rrparenthesis) = (G_V)^n \overset{s_R}{\leftarrow} G_R \overset{t_R}{\rightarrow} (G_V)^m$ for each $R \in \Sigma_{n,m}$. Below, we also use the fact that $(G_V)^n$ is $\mathbf{Hyp}_\Sigma[n, G]$.

The conclusion of Theorem 31 allows us to argue by induction on $n \overset{\iota}{\rightarrow} G' \overset{\omega}{\leftarrow} m$. The base cases are $\llparenthesis$$\rrparenthesis$, $\llparenthesis$$\rrparenthesis$, $\llparenthesis$$\rrparenthesis$, $\llparenthesis$$\rrparenthesis$, $\llparenthesis$$\rrparenthesis$, $\llparenthesis$$\rrparenthesis$ and $\llparenthesis R \rrparenthesis$. Let us consider the last of these, where $n \overset{\iota}{\rightarrow} G' \overset{\omega}{\leftarrow} m$ is

$$\llparenthesis R \rrparenthesis = \quad$$


Any homomorphism $f\colon G' \to G$ maps its single hyperedge to an R-hyperedge of G, call it e_f, the n vertices in the image of ι to the source of e_f ($\iota\,;\, f = s_R(e_f)$) and the m vertices in the image of ω to the target of e_f ($\omega\,;\, f = t_R(e_f)$). This means that the following commutes:



The function $e_-\colon \mathbf{Hyp}_\Sigma[G', G] \to G_R$ is clearly an isomorphism of spans. The other base cases are simpler or, as stated in the main text, follow from the fact that $\mathbf{Hyp}_\Sigma[\_, G]$ maps colimits to limits, which also immediately implies the inductive case. ◀

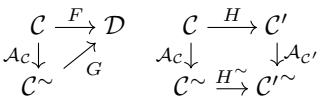
B.4 Proofs of Section 8

Observe that we have a canonical identity-on-objects-functor $\mathcal{A}_\mathcal{C}\colon \mathcal{C} \to \mathcal{C}^\sim$ that sends a morphism in $\mathcal{C}$ to its $\sim$-equivalence class in $\mathcal{C}^\sim$. We will omit the subscript on $\mathcal{A}$ whenever possible. An immediate consequence of the definition is that $\mathcal{A}$ preserves and reflects the ordering in the following sense:

▶ **Lemma 49.** *For $\mathcal{C}$ a preorder-enriched category, and x, y morphisms in $\mathcal{C}$, we have $\mathcal{A}(x) \leq \mathcal{A}(y)$ if and only if $x \leq y$.* ◀

The functors $\mathcal{A}$ exhibit the following universal property:

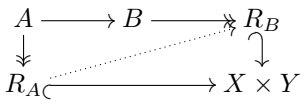
▶ **Lemma 50.** *For every preordered functor $F\colon \mathcal{C} \to \mathcal{D}$ between the preordered category $\mathcal{C}$ and poset-enriched category $\mathcal{D}$, there is a unique poset-enriched functor $G\colon \mathcal{C}^\sim \to \mathcal{D}$ making the left diagram commute. Hence, for every preordered functor $H\colon \mathcal{C} \to \mathcal{C}'$, $\mathcal{C}'$ preorder-enriched, there is a unique functor $H^\sim\colon \mathcal{C}^\sim \to \mathcal{C}'^\sim$ making the right diagram commute.*



Proof. For a morphism $f \in \mathcal{C}$ let $[f]$ denote the equivalence class of f modulo $\sim$. Then setting $G([f]) = F(f)$ is well-defined, since $\mathcal{D}$ is a poset-enriched category. G defines a functor since $\sim$ is a congruence, hence compatible with composition. Since A_C is surjective on objects and morphisms, there can be at most one such functor G, hence G is unique. ◀

In other words, we get a function, $(\cdot)^{\sim}$, that turns functors between preorder-enriched categories into functors between the associated poset-enriched ones.

Proof of Proposition 39. We recall a well-known construction of the ordinary category of relations: a span $X \xleftarrow{f} A \xrightarrow{g} Y$ induces a relation $R_A \subseteq X \times Y$ by factorising $A \xrightarrow{[f,g]} X \times Y$ as a surjection followed by an injection; the injection, when composed with the projections, yields a *jointly-injective* span. These, up-to span isomorphism, are the same thing as subsets $R_A \subseteq X \times Y$. This procedure respects composition and monoidal product, yielding a functorial mapping $\mathsf{Span}^{\leq}\,\mathbf{Set} \to \mathbf{Rel}$ on objects and arrows. Given the above, it suffices to show that there exists a span homomorphism $(X \leftarrow A \rightarrow Y) \rightarrow (X \leftarrow B \rightarrow Y)$ iff $R_A \subseteq R_B$ as relations. The 'only if' direction is implied by the dotted function below, which is an injection since it is the first part of a factorisation of an injection.



For the 'if' part, since (by the axiom of choice) surjective functions split, we obtain $R_B \to B$. Then $A \twoheadrightarrow R_A \to R_B \to B$ is easily shown to be a homomorphism of spans. ◀

Proof of Proposition 40. We stated the axioms of preordered cartesian bicategories and cartesian bicategories in a way that makes the first part obvious. Given a morphism $F\colon \mathcal{B}_1 \to \mathcal{B}_2$ of preorder-enriched cartesian bicategories, clearly $F^{\sim}$ is still an order-preserving monoidal functor. It also preserves the monoid and comonoid structures. ◀

Proof of Lemma 42. The second item is trivial. For the first one, let x, y be morphisms in $\mathcal{C}^{\sim}$ such that $G(x) \leq G(y)$ for all $\mathcal{G} \in \mathcal{F}^{\sim}$. We want to prove $x \leq y$. Now let $F \in \mathcal{F}$ be arbitrary. Then $F^{\sim}(x) \leq F^{\sim}(y)$ by assumption on x, y. Since morphisms in $\mathcal{C}^{\sim}$ are just equivalence classes of morphisms in $\mathcal{C}$, choose representatives, i.e. morphisms f, g in $\mathcal{C}$ such that $\mathcal{A}(f) = x$ and $\mathcal{A}(g) = x$. Since the diagram

$$\begin{array}{ccc} \mathcal{C} & \xrightarrow{F} & \mathcal{D} \\ {\scriptstyle\mathcal{A}}\downarrow & & \downarrow{\scriptstyle\mathcal{A}} \\ \mathcal{C}^{\sim} & \xrightarrow{F^{\sim}} & \mathcal{D}^{\sim} \end{array}$$

commutes, we get $\mathcal{A}(F(f)) = F^{\sim}(\mathcal{A}(f)) = F^{\sim}(x) \leq F^{\sim}(y) = F^{\sim}(\mathcal{A}(g)) = \mathcal{A}(F(g))$. Since $\mathcal{A}$ reflects the ordering (Lemma 49), we get $F(f) \leq F(g)$. But $F \in \mathcal{F}$ was arbitrary, therefore $f \leq g$, since $\mathcal{C}$ is $\mathcal{F}$-complete for $\mathcal{D}$. But therefore $x = \mathcal{A}(f) \leq \mathcal{A}(g) = y$. ◀