Codd's Theorem for Databases over Semirings

GUILLERMO BADIA, University of Queensland, Australia
PHOKION G. KOLAITIS, University of California Santa Cruz, USA and IBM Research, USA
CARLES NOGUERA, University of Siena, Italy

Codd's Theorem, a fundamental result of database theory, asserts that relational algebra and relational calculus have the same expressive power on relational databases. We explore Codd's Theorem for databases over semirings and establish two different versions of this result for such databases: the first version involves the five basic operations of relational algebra, while in the second version the division operation is added to the five basic operations of relational algebra. In both versions, the difference operation of relations is given semantics using semirings with monus, while on the side of relational calculus a limited form of negation is used. The reason for considering these two different versions of Codd's theorem is that, unlike the case of ordinary relational databases, the division operation need not be expressible in terms of the five basic operations of relational algebra for databases over an arbitrary positive semiring; in fact, we show that this inexpressibility result holds for bag databases, as well as for databases over the tropical semiring.

CCS Concepts: • **Information systems → Relational database model**; **Relational database model**; • **Theory of computation → Logic**;

Additional Key Words and Phrases: Relational algebra; relational calculus; annotated databases; semirings

ACM Reference Format:
Guillermo Badia, Phokion G. Kolaitis, and Carles Noguera. 2025. Codd's Theorem for Databases over Semirings. *Proc. ACM Manag. Data* 3, 5 (PODS), Article 277 (November 2025), 26 pages. https://doi.org/10.1145/3767713

1 Introduction

Background and Motivation Codd's Theorem [8] is arguably the most fundamental result in relational database theory. Informally, it asserts that relational algebra and relational calculus have the same expressive power on relational databases. The significance of Codd's Theorem is that it demonstrated that a procedural query language based on operations on relations and a declarative query language based on first-order logic can express precisely the same queries. As regards its impact, Codd's Theorem gave rise to the notion of *relational completeness* as a benchmark for the expressive power of query languages. Furthermore, it became the catalyst for numerous subsequent investigations of both sublanguages of relational algebra and relational calculus, such as conjunctive queries, and extensions of relational algebra and relational calculus, such as Datalog.

More recently, there has been an extensive study of databases over semirings, i.e., databases in which the tuples of the relations are annotated with elements from a fixed semiring. Standard relational databases are, of course, databases over the Boolean semiring $\mathbb{B} = (\{0, 1\}, \vee, \wedge, 0, 1)$, while *bag* databases are databases over the semiring $\mathbb{N} = (N, +, \cdot, 0, 1)$ of the natural numbers. The interest in semirings other than the Boolean and the bag ones was sparked by the influential paper [14] in which semirings of polynomials were used to formalize and study the *provenance* of database

Authors' Contact Information: Guillermo Badia, g.badia@uq.edu.au, University of Queensland, Australia; Phokion G. Kolaitis, kolaitis@ucsc.edu, University of California Santa Cruz, USA and IBM Research, USA; Carles Noguera, carles.noguera@unisi.it, University of Siena, Italy.

ACM 2836-6573/2025/11-ART277
https://doi.org/10.1145/3767713

queries. Originally, the semiring-based study of provenance focused on the positive fragment of relational algebra (i.e., unions of conjunctive queries), but it was subsequently extended to the study of provenance in first-order logic [15] and least fixed-point logic [9]. Researchers also investigated a variety of other database problems under semiring semantics, including the query containment problem [13, 20], the evaluation of Datalog queries [19, 28], and the interplay between local consistency and global consistency for relations over semirings [5, 6]. Furthermore, connections between databases over semirings and quantum information theory were established [2].

In view of the aforementioned developments, it is rather surprising that the following questions have not been addressed thus far: Is there a Codd's Theorem for databases over semirings? In other words, do relational algebra and relational calculus have the same expressive power on databases over semirings? In particular, is there a version of Codd's Theorem for bag databases? Our goal in this paper is to investigate these questions and to provide some answers.

Before proceeding further, let us take a closer look at the precise statement of Codd's Theorem. In a literal sense, it is simply *not* true that relational algebra and relational calculus have the same expressive power over relational databases. One of reasons for this is that relational calculus allows for the use of unrestricted negation, which implies that relational calculus formulas can define infinite relations. Codd [8] was aware of this issue, thus he carefully crafted a collection of relational calculus formulas, called *α-expressions*, and showed that α-expressions have the same expressive power as relational algebra. Since the syntax of α-expressions is somewhat convoluted, there was extensive follow-up research on richer syntactically-defined fragments of relational calculus that have the same expressive power as relational algebra (see, e.g., [11]). In a parallel direction, it was also realized that the equivalence between relational algebra and relational calculus holds if the relational calculus formulas are *domain independent*, which means that they return the same answers as long as the universe of discourse contains the *active domain* of the database at hand. This version of Codd's Theorem is the one presented in database theory textbooks (e.g., [1]). We will investigate whether and how this version of Codd's Theorem extends to databases over semirings.

Analysis of Choices A semiring is a structure $\mathbb{K} = (K, +, \cdot, 0, 1)$, where addition $+$ and multiplication $\cdot$ are binary operations on K, and certain identities hold (precise definitions are given in the next section). When positive relational algebra is interpreted over semirings, the operations $+$ and $\cdot$ are used to give semantics to union $\cup$ and Cartesian product $\times$. Similarly, $+$ and $\cdot$ are used to give semantics to disjunction $\vee$ and conjunction $\wedge$, when unions of conjunctive queries are interpreted over semirings. However, relational algebra also contains the difference operation $\setminus$ as a basic operation. Thus, to interpret full relational algebra over semirings, we must expand the semirings with a "difference" operation. It is well understood, though, that there is no unique way to expand a semiring with a well-behaved "difference" operation (see [26] for a thorough coverage of this issue). Thus, a choice has to be made concerning the meaning of the "difference" operation on semirings.

Here, we opt to focus on semirings with a *monus* operation, which is well defined for *naturally ordered* semirings with some additional properties. In the study of provenance of arbitrary relational algebra expressions, the monus operation was adopted in [10] as an interpretation of the difference operation on relations; this choice was criticized in [4] on the grounds that certain desirable algebraic identities involving natural join and difference fail on some semirings with monus. Our decision to focus on semirings with monus is based on two considerations: first, semirings with monus comprise a large class of semirings that, beyond the Boolean semiring, includes the bag semiring, the tropical semiring, various provenance semirings, and every complete bounded distributive lattice (hence, the fuzzy semiring); second, we do not believe that the failure of some useful identities on some semirings with monus is sufficient justification to disregard them. Such identities as $R \cup R = R$ and $R \bowtie (R \bowtie S) = R \bowtie S$ fail on the bag semiring, yet bag semantics is the default SQL semantics.

And, of course, if one wishes to establish a version of Codd's Theorem for bag databases, then the monus operation on bags has to be used, since, this is what is used in the semantics of SQL.

On the side of relational calculus, we have to make a choice concerning negation. In the study of provenance of first-order formulas, Grädel and Tannen [15] allowed for unrestricted negation, but the formulas are always assumed to be in negation normal form (i.e., the negation symbol is pushed in front of atomic formulas). Further, the notion of an *interpretation* is used to give semiring semantics to first-order formulas in negation normal form, where an interpretation is an assignment of semiring values to every atomic and negated atomic statements. Note that Grädel and Tannen [15] do not explicitly define what it means for a database over a semiring $\mathbb{K}$ to satisfy a first-order formula; such a definition, however, could be obtained by fixing a particular *canonical* interpretation associated with the database of interest. We believe that the restriction to formulas in negation normal form is rather severe from a query language point of view. For this reason, we do not require the formulas to be in any normal form instead, we opt to use the binary connective $\prec$ ("but not"), which expresses a limited form of negation and has a long history of uses in algebra and in non-classical logics [12, 22, 24]. We note that, when logics are given semiring semantics, the use of the connective $\prec$ to model negation was criticized in [16]. That criticism was mainly based on the argument that the connective $\prec$ does not give rise to a "reasonable" form of negation, when the negation $\neg\varphi$ of a formula φ is defined as $1 \prec \varphi$, i.e., when $\prec$ is used to define an unrestricted form of negation. As mentioned earlier, however, the connective $\prec$ has been traditionally used to model a restricted form of negation (a "safe" negation, if you will) and not an unrestricted negation. In our setting, we need a restricted form of negation that is the counterpart of the monus operation.

There are two other important issues to bring up at this point. First, when one explores a Codd's Theorem for databases over semirings, one quickly realizes that relational algebra has to be expanded with a *support* operation. This operation takes as input a relation over a semiring, removes all annotations from the semiring, and returns the underlying (ordinary) relation as output. For the Boolean semiring, there is no need to consider the support as an additional operation, since the support of a relation is the relation itself. For relations over the fuzzy semiring, however, one can show that the support is not expressible in terms of the basic operations of relational algebra, thus the support operation has to be added to the syntax of relational algebra for relations over semirings. In view of these considerations, we consider semirings expanded with a support operation s such that $\mathrm{s}(a) = 1$ if $a \neq 0$, while $\mathrm{s}(a) = 0$ if $a = 0$.

The second issue is more subtle and more consequential. Codd showed that the five basic operations of relational algebra (union, difference, Cartesian product, selection, and projection) can express the *division* (or the *quotient*) of two relations [8]. Division is a powerful operation that is used to express universal quantification in the translation of relational calculus to relational algebra. The division operation has natural semantics for relations over semirings. At first, one may expect that division is always expressible in terms of the five basic relational algebra operations when we consider relations over semirings. Unfortunately, this is not true. Using combinatorial arguments, we show that both on the bag semiring $\mathbb{N} = (N, +, \cdot, 0, 1)$ and on the tropical semiring $\mathbb{T} = (\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$, the division operation is *not* expressible in terms of the five basic relational algebra operations. Thus, to obtain a version of Codd's Theorem for relational calculus formulas with universal quantification, relational algebra must be expanded with a division operation.

Contributions With the insights gained from the preceding analysis, we proceed to establish the following two versions of Codd's Theorem for databases over semirings, one for relational algebra without division and one for relational algebra with division.

Let $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ be a commutative semiring with monus and support. We write $\mathcal{BRA}$ to denote the collection of all *basic relational algebra* expressions, that is, all relational algebra expressions over a schema built using the operations union, difference, Cartesian product, selection, projection, and support. We also write $\mathcal{BRC}$ to denote the collection of all *basic relational calculus* formulas, that is, all formulas built from atomic formulas, using the connectives $\wedge$, $\vee$, $\prec$, and existential quantification. We show that a query q on $\mathbb{K}$-databases is expressible by a $\mathcal{BRA}$-expression if and only if it is definable by a domain independent $\mathcal{BRC}$-formula (Theorem 1).

Let $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ be a semiring with monus and support, and no zero-divisors (i.e., if $a \cdot b = 0$, then $a = 0$ or $b = 0$). We write $\mathcal{RA}$ to denote the collection of all *relational algebra* expressions, that is, all expressions over a schema built using the operations union, difference, Cartesian product, selection, projection, support, and division. We also write $\mathcal{RC}$ to denote the collection of all *relational calculus* formulas, that is, all formulas built from atomic formulas, using the connectives $\wedge$, $\vee$, $\prec$, and both existential and universal quantification. We show that a query q on $\mathbb{K}$-databases is expressible by an $\mathcal{RA}$-expression if and only if it is definable by a domain independent $\mathcal{RC}$-formula (Theorem 3), i.e., a formula whose meaning depends only on the active domain. An immediate corollary is a version of Codd's Theorem for bag databases. In our view, this corollary alone justifies the choices of the monus operation and of the connective $\prec$ that we made.

The proofs proceed by induction on the construction of the (basic) relational algebra expressions and the construction of the (basic) relational calculus formulas. Some of the inductive steps, require a rather delicate analysis because of the nuances that arise when $\mathbb{K}$-databases over semirings are considered. Also, in the simulation of (basic) relational calculus by (basic) relational algebra, the proofs use repeatedly the fact that there is a relational algebra expression that uniformly defines the active domain of a $\mathbb{K}$-database, a fact whose proof uses the support operation.

In conclusion, our work makes a contribution to the ongoing study of databases over semirings by investigating Codd's Theorem in this context. Specifically, our work identifies the "right" constructs of relational algebra and relational calculus needed to establish Codd's Theorem for databases over semirings, delineates the features of these constructs, unveils differences between the Boolean semiring and other semirings, and potentially establishes benchmarks for comparing the expressive power of these and other formalisms under semiring semantics.

2 Semirings with Monus and Support

Definition 1. *A* semiring *is a structure* $\mathbb{K} = (K, +, \cdot, 0, 1)$*, where addition* $+$ *and multiplication* $\cdot$ *are binary operations on* K*, and* 0*,* 1 *are elements of* K *such that*

- $(K, +, 0)$ *is a commutative monoid,* $(K, \cdot, 1)$ *is a monoid, and* $0 \neq 1$*;*
- *for all* a, b, c *in* K*, we have that* $a \cdot (b + c) = a \cdot b + a \cdot c$ *and* $(b + c) \cdot a = b \cdot a + c \cdot a$*;*
- *for all* a *in* K*, we have that* $a \cdot 0 = 0 \cdot a = 0$.

Let $\mathbb{K} = (K, +, \cdot, 0, 1)$ be a semiring. We will consider several additional properties.

- $\mathbb{K}$ is *commutative* if the monoid $(K, \cdot, 1)$ is commutative.
- $\mathbb{K}$ is *zero-sum-free* if for all elements $a, b \in K$ with $a + b = 0$, we have that $a = 0$ and $b = 0$.
- $\mathbb{K}$ has *no zero divisors* if for all elements $a, b \in K$ with $a \cdot b = 0$, we have that $a = 0$ or $b = 0$.
- $\mathbb{K}$ is a *positive* semiring if it is commutative, zero-sum-free, and has no zero divisors.

A *ring* is a semiring $\mathbb{K} = (K, +, \cdot, 0, 1)$ such that $(K, +, 0)$ is a commutative group, i.e., a ring is a commutative semiring in which for every $b \in K$, there is an element $-b \in K$ such that $b + (-b) = 0$. Therefore, on every ring, a binary *difference* operation $-$ can be defined by setting $a - b = a + (-b)$. This raises the question: is it possible to expand a semiring with a binary operation that possesses some of the properties of the difference operation on rings. As discussed in [26], there are several alternatives to defining such an operation. Here, we will focus on the *monus* operation [3, 10], which, however, requires that the semiring at hand is *naturally ordered.*

If $\mathbb{K} = (K, +, \cdot, \mathrm{s}, 0, 1)$ is a semiring, then the *natural preorder* of $\mathbb{K}$ is the binary relation $\preceq$ on K defined as follows: $a \preceq b$ if there is an element $c \in K$ such that $a + c = b$. Clearly, $\preceq$ is a reflexive and transitive relation on K, i.e., $\preceq$ is indeed a preorder. If $\preceq$ happens to be a partial order (that is, $a \preceq b$ and $b \preceq a$ imply $a = b$), then $\preceq$ is called the *natural order* of the semiring $\mathbb{K}$. In this case, we say that $\mathbb{K}$ is a *naturally ordered* semiring.

Note that if $\mathbb{K}$ is naturally ordered, then $\mathbb{K}$ is zero-sum-free. Indeed, if $a + b = 0$, then $a \preceq 0$, but also $0 \preceq a$ (since $0 + a = a$), hence $a = 0$ because $\preceq$ is a partial order; similarly, we get $b = 0$.

Definition 2. *A* semiring with monus *is an algebraic structure* $\mathbb{K} = (K, +, \cdot, -, 0, 1)$ *such that:*

(1) $(K, +, \cdot, 0, 1)$ *is a naturally ordered commutative semiring;*
(2) *For all elements $a, b \in K$, there is a smallest (in the sense of the natural order $\preceq$) element c in K, denoted by $a - b$, such that $a \preceq b + c$.*

The next proposition provides different characterizations of the monus operation (see [3, 4]).

Proposition 1. *Let $\mathbb{K} = (K, +, \cdot, 0, 1)$ be a naturally ordered semiring and let $-'$ be a binary operation on K. Then, the following statements are equivalent:*

(1) *For all $a, b \in K$, we have $a -' b = a - b$, i.e., $a -' b$ is the smallest element c such that $a \preceq b + c$.*
(2) *For all $a, b, c \in K$, we have: $a -' b \preceq c$ if and only if $a \preceq b + c$.*
(3) *The following four identities hold for $-'$:*

$$a -' a = 0; \quad 0 -' a = 0; \quad a + (b -' a) = b + (a -' b); \quad a -' (b + c) = (a -' b) -' c.$$

We now present several examples of important semirings with monus.

Example 1. *The two-element Boolean semiring $\mathbb{B} = (\{\top, \bot\}, \wedge, \vee, \bot, \top)$ can be expanded to a semiring with monus in which $a - b := a \wedge b^*$, where $*$ is the standard Boolean complement.*

Example 2. *The semiring $\mathbb{N} = (N, +, \cdot, 0, 1)$ of the natural numbers, also known as the* bag *semiring, is a naturally ordered semiring in which the natural order coincides with the standard order $\leq$ on N. The bag semiring can be expanded to a semiring with monus in which the monus operation is the* truncated subtraction *$a \dot{-} b$, that is, $a \dot{-} b = a - b$ if $a \geq b$, while $a \dot{-} b = 0$ if $a < b$.*

Example 3. *Every bounded distributive lattice $\mathbb{D} = (L, \vee, \wedge, \bot, \top)$ is a naturally ordered semiring in which $a \preceq b$ if and only if $a \vee b = b$. Such a lattice is* complete *if every non-empty subset of L has a smallest element. Every complete and bounded distributive lattice can be expanded to a semiring with monus in which $a - b = \min\{c : a \preceq b \vee c\}$. Complete bounded distributed lattices expanded with monus are called* Brouwerian algebras*, a term coined by McKinsey and Tarski [22]. A prominent example of a Brouwerian algebra is the* fuzzy *semiring $\mathbb{F} = ([0, 1], \max, \min, 0, 1)$ expanded with monus; in this case, we have that $a - b = a$ if $a > b$, while $a - b = 0$ if $a \leq b$.*

Example 4. *The* Łukasiewicz *semiring $\mathbb{L} = ([0, 1], \max, \odot, 0, 1)$, where $a \odot b = \max(a + b - 1, 0)$, is a naturally ordered semiring in which the natural order coincides with the standard order on $[0, 1]$. It can be expanded to a semiring with monus in which $a - b = a$ if $a > b$, while $a - b = 0$ if $a \leq b$.*

Example 5. *The* tropical *semiring $\mathbb{T} = (\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$ of the real numbers is a semiring in which the natural order is the reverse standard order on $\mathbb{R}$, i.e., $a \preceq b$ if and only if $a \geq b$. It can be expanded to a semiring with monus by letting $a - b = a$ if $a < b$ (in the standard order of the reals), while $a - b = \infty$ if $a \geq b$ (in the standard order of the reals).*

Example 6. *Let $\mathbb{N}[X]$ be the semiring of all polynomials with variables from a set $X = \{x_1, \ldots, x_n\}$ and with coefficients from the set N of non-negative integers. This semiring is called the* provenance *semiring because of its use in the study of database provenance [15]. As pointed out in [10, Example 9], the provenance semiring $\mathbb{N}[X]$ can be expanded to a semiring with monus in the following way. For every tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in N^n$, let x^α denote the monomial $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$, where we set $x_i^0 = 1$.*

For a finite set $I \subseteq \mathbb{N}^n$, let $f[X] = \sum_{\alpha \in I} f_\alpha x^\alpha$ and $g[X] = \sum_{\alpha \in I} g_\alpha x^\alpha$ be two polynomials in $\mathbb{N}[X]$. The monus is defined by letting $f[X] - g[X] = \sum_{\alpha \in I}(f_\alpha \dot{-} g_\alpha)x^\alpha$, where $\dot{-}$ is the truncated difference on N, that is, $f_\alpha \dot{-} g_\alpha = f_\alpha - g_\alpha$ if $f_\alpha \geq g_\alpha$, and $f_\alpha \dot{-} g_\alpha = 0$ if $f_\alpha < g_\alpha$.

Several other well-studied provenance semirings, such as $\mathrm{Why}(X)$, $\mathrm{Trio}(X)$, *and* $\mathrm{Lin}(X)$ *can also be expanded to semirings with monus (for the definitions of these semirings, see, e.g., [13]). In fact, the restriction of the monus on $\mathbb{N}(X)$ to each of these semirings forms a monus for that semiring.*

Not every naturally ordered semiring can be expanded to a semiring with monus. An example of such a semiring can be found in [23].

Definition 3. *A* semiring with support *is a structure* $\mathbb{K} = (K, +, \cdot, \mathrm{s}, 0, 1)$ *such that* $(K, +, \cdot, 0, 1)$ *is a semiring and* s *is the unary operation on K with $s(a) = 1$ if $a \neq 0$ and $s(a) = 0$ if $a - 0$, where a in K.*

Unlike monus, every semiring can be expanded (in the sense of model theory) to a semiring with support, and in a unique way.

Definition 4. *A* semiring with monus and support *is a structure* $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ *such that* $(K, +, \cdot, -, 0, 1)$ *is a semiring with monus and* $(K, +, \cdot, \mathrm{s}, 0, 1)$ *is a semiring with support.*

3 Codd's Theorem for Relational Algebra with Monus and Support

3.1 Semiring-annotated Relations and Basic Relational Algebra

In what follows, we assume that D is a fixed denumerable set, which will be the *domain* of elements occurring in databases. For $n \geq 1$, we write D^n to denote the set of all n-tuples from D. We also write D^0 to denote the singleton $\{<>\}$, where $<>$ is the empty tuple (we assume that $<> \notin D$).

Definition 5. *Let $\mathbb{K} = (K, +, \cdot, 0, 1)$ be a semiring and let $n \geq 0$ be a non-negative integer.*

1. An n-ary $\mathbb{K}$-relation R is a function $R\colon D^n \longrightarrow K$ of finite support, that is to say, the set $\mathrm{supp}(R) := \{(d_1, \ldots, d_n) \in D^n : R(d_1, \ldots, d_n) \neq 0\}$ *is finite. We call n the* arity *of R.*

2. A $\mathbb{K}$-relation *is an n-ary $\mathbb{K}$-relation R, for some $n \geq 0$.*

3. If $n \geq 1$ and $A \subseteq D$, then an n-ary $\mathbb{K}$-relation on A *is an n-ary $\mathbb{K}$-relation R with* $\mathrm{supp}(R) \subseteq A^n$.

Note that there are $|K|$-many 0-ary $\mathbb{K}$-relations. Indeed, for every $a \in K$, we have the 0-ary relation R^0_a such that $R^0_a(<>) = a$. If $a \neq 0$, then $\mathrm{supp}(R^0_a) = \{<>\}$, while $\mathrm{supp}(R^0_0) = \emptyset$.

Next, we will define several different operations on $\mathbb{K}$-relations. Before doing so, we need to introduce some useful terminology and notation.

- Let $\mathbf{t} = (d_1, \ldots, d_n) \in D^n$ be an n-tuple of elements from D, where $n \geq 1$. If $v = (i_1, \ldots, i_k)$ is a k-tuple of distinct indices from $\{1, \ldots, n\}$, then $\mathbf{t}[v]$ denotes the *projection of* $\mathbf{t}$ *on* v, that is to say, $\mathbf{t}[v] = (d_{i_1}, \ldots, d_{i_k}) \in D^k$. Observe that, if $k = n$, then $\mathbf{t}[v]$ is a permutation of $\mathbf{t}$.
- An *n-ary selection condition* θ is an expression built by using (some or all of) the variables $x_1 \ldots, x_n$, equalities $x_i = x_j$, negated equalities $x_i \neq x_j$, conjunctions, and disjunctions, where $1 \leq i, j \leq n$. Every n-ary selection condition θ gives rise to an n-ary *selection function* $P_\theta : D^n \longrightarrow \{0, 1\}$ such that for every n-tuple $\mathbf{t} = (d_1, \ldots, d_n) \in D^n$, we have that $P_\theta(\mathbf{t}) = 1$, if the assignment $x_i \mapsto d_i$, $1 \leq i \leq n$, satisfies θ, while $P_\theta(\mathbf{t}) = 0$, otherwise. Note that every n-ary selection condition is also an m-ary selection condition, for every $m \geq n$.

Definition 6 (Operations on $\mathbb{K}$-relations). *Let $\mathbb{K} = (K, +, \cdot, -, s, 0, 1)$ be a semiring with monus and support. We define the following operations on $\mathbb{K}$-relations.*

- **Union:** *If R_1 and R_2 are two n-ary $\mathbb{K}$-relations, then the* union $R_1 \cup R_2$ *is the n-ary $\mathbb{K}$-relation defined by $(R_1 \cup R_2)(\mathbf{t}) = R_1(\mathbf{t}) + R_2(\mathbf{t})$, for all $\mathbf{t} \in D^n$.*
- **Difference:** *If R_1 and R_2 are two n-ary $\mathbb{K}$-relations, then the* difference $R_1 \setminus R_2$ *is the n-ary $\mathbb{K}$-relation defined by $(R_1 \setminus R_2)(\mathbf{t}) = R_1(\mathbf{t}) - R_2(\mathbf{t})$, for all $\mathbf{t} \in D^n$.*

- **Cartesian Product:** *If* R_1 *is an* m*-ary* $\mathbb{K}$*-relation and* R_2 *is an* n*-ary* $\mathbb{K}$*-relation, then the* Cartesian product $R_1 \times R_2$ *is the* $(m+n)$*-ary* $\mathbb{K}$*-relation defined by* $(R_1 \times R_2)(\mathbf{t}) = R_1(\mathbf{t}[v_1]) \cdot R_2(\mathbf{t}[v_2])$*, for all* $\mathbf{t} \in D^{m+n}$*, where* $v_1 = (1, \ldots, m)$ *and* $v_2 = (m+1, \ldots, m+n)$*.*
- **Projection:**
(i) If R *is an* n*-ary* $\mathbb{K}$*-relation with* $n \geq 1$ *and* $v = (i_1, \ldots, i_k)$ *is a non-empty sequence of distinct indices from* $\{1, \ldots, n\}$*, then the* projection $\pi_v(R)$ *is the* k*-ary* $\mathbb{K}$*-relation with* $\pi_v(R)(\mathbf{t}) = \sum_{\mathbf{w}[v]=\mathbf{t} \textit{ and } R(\mathbf{w}) \neq 0} R(\mathbf{w})$*, for all* $\mathbf{t} \in D^n$*. (We take the empty sum to be* 0*.)*
(ii) If R *is a* $\mathbb{K}$*-relation, then the* projection of R on the empty sequence *is the* 0*-ary relation* $\pi_{[\,]}(R)$ *defined by* $\pi_{[\,]}(R)(<>) = \sum_{R(\mathbf{t}) \neq 0} R(\mathbf{t})$*.*
- **Selection:** *If* R *is an* n*-ary* $\mathbb{K}$*-relation and* θ *is an* n*-ary selection condition, then the* selection $\sigma_\theta(R)$ *is the* n*-ary* $\mathbb{K}$*-relation defined by* $\sigma_\theta(R)(\mathbf{t}) = R(\mathbf{t}) \cdot P_\theta(\mathbf{t})$*, for all* $\mathbf{t} \in D^n$*, where* P_θ *is the* n*-ary selection function arising from* θ*. Note that* $\sigma_\theta(R)$ *is indeed a* $\mathbb{K}$*-relation, since* $\mathrm{supp}(\sigma_\theta(R)) \subseteq \mathrm{supp}(R)$*, and* $\mathrm{supp}(R)$ *is a finite set.*
- **Support:** *If* R *is an* n*-ary* $\mathbb{K}$*-relation, then the* support $\mathrm{supp}(R)$ of R *is the* n*-ary* $\mathbb{K}$*-relation defined by* $\mathrm{supp}(R)(\mathbf{t}) = \mathrm{s}(R(\mathbf{t}))$*, for all* $\mathbf{t} \in D^n$*.*

In Definition 5, we defined the support $\mathrm{supp}(R)$ of an n-ary $\mathbb{K}$-relation as the set of all n-tuples in D^n on which R takes a non-zero value, while here we just defined $\mathrm{supp}(R)$ to be another $\mathbb{K}$-relation. There is no inconsistency between these two definitions, because $\mathrm{supp}(R)$ viewed as a $\mathbb{K}$-relation takes values 0 and 1 only, so it coincides with the ordinary relation consisting of all n-tuples on which R takes a non-zero value. We will use these two views of $\mathrm{supp}(R)$ interchangeably.

We now introduce the syntax of basic relational algebra, which is a formalism for expressing combinations of the operations on $\mathbb{K}$-relations that were described in Definition 6. As usual, a *relational schema* or, simply, a *schema* is a finite sequence $\tau = (\hat{R}_1, \ldots, \hat{R}_p)$ of relation symbols, where each relation symbol $\hat{R}_i$ has a designated positive integer r_i as its *arity*.

Definition 7 (Basic Relational Algebra). *Let* $\tau = (\hat{R}_1, \ldots, \hat{R}_p)$ *be a schema. The* basic relational algebra language $\mathcal{BRA}(\tau)$ over τ *is the set of all expressions defined recursively as follows:*

- *Each relation symbol* $\hat{R}_i$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* r_i*,* $1 \leq i \leq p$*.*
- *If* E_1, E_2 *are* $\mathcal{BRA}(\tau)$*-expressions of the same arity* n*, then* $E_1 \cup E_2$ *and* $E_1 \setminus E_2$ *are* $\mathcal{BRA}(\tau)$*-expressions of arity* n*.*
- *If* E_1, E_2 *are* $\mathcal{BRA}(\tau)$*-expressions of arities* n, m*, respectively, then* $E_1 \times E_2$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* $n+k$*.*
- *If* E *is a* $\mathcal{BRA}(\tau)$*-expression of arity* $n \geq 1$ *and* $v = (i_1, \ldots, i_k)$ *is a sequence of distinct indices from* $\{1, \ldots, n\}$*, then* $\pi_v(E)$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* k*. Furthermore, if* E *is a* $\mathcal{BRA}(\tau)$*-expression of arity* $n \geq 0$*, then* $\pi_{[\,]}(E)$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* 0*.*
- *If* E *is a* $\mathcal{BRA}(\tau)$*-expression with arity* n *and* θ *is an* n*-ary selection condition, then* $\sigma_\theta(E)$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* n*,*
- *If* E *is a* $\mathcal{BRA}(\tau)$*-expression of arity* n*, then* $\mathrm{supp}(E)$ *is a* $\mathcal{BRA}(\tau)$*-expression of arity* n*.*

Definition 8 (Semantics of Basic Relational Algebra). *Let* $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ *be a semiring with monus and support, and let* $\tau = (\hat{R}_1, \ldots, \hat{R}_p)$ *be a schema.*

- *A* $\mathbb{K}$-database over τ *is a finite sequence* $\mathbf{I} = (R_1, \ldots, R_p)$ *of* $\mathbb{K}$*-relations such that the arity of each* $\mathbb{K}$*-relation* R_i *matches the arity of the relation symbol* $\hat{R}_i$*,* $1 \leq i \leq p$*.*
We say that $\mathbf{I}$ *is* non-trivial *if at least one of the relations* R_i *of* $\mathbf{I}$ *has non-empty support. In what follows, we make the blanket assumption that all* $\mathbb{K}$*-databases considered are non-trivial.*
- *If* E *is a* $\mathcal{BRA}(\tau)$*-expression of arity* n *and* $\mathbf{I} = (R_1, \ldots, R_p)$ *is a* $\mathbb{K}$*-database, then* $E^{\mathbf{I}}$ *is the* n*-ary* $\mathbb{K}$*-relation obtained by evaluating* E *on* $\mathbf{I}$ *according to the definitions of the operations on* $\mathbb{K}$*-relations. For completeness, we spell out the definition of the semantics of* $\mathcal{BRA}$*-expressions:*

$(\hat{R}_i)^{\mathbf{I}} = R_i$, *where* $1 \leq i \leq p$; $(E_1 \cup E_2)^{\mathbf{I}} = E_1^{\mathbf{I}} \cup E_2^{\mathbf{I}}$ *and* $(E_1 \setminus E_2)^{\mathbf{I}} = E_1^{\mathbf{I}} \setminus E_2^{\mathbf{I}}$, *where* E_1 *and* E_2 *are* $\mathcal{BRA}$*-expressions of the same arity;* $(E_1 \times E_2)^{\mathbf{I}} = E_1^{\mathbf{I}} \times E_2^{\mathbf{I}}$; $(\pi_v(E))^{\mathbf{I}} = \pi_v(E^{\mathbf{I}})$ *and* $(\pi_{[\,]}(E))^{\mathbf{I}} = \pi_{[\,]}(E^{\mathbf{I}})$; $(\sigma_\theta(E))^{\mathbf{I}} = \sigma_\theta(E^{\mathbf{I}})$; $(\mathrm{supp}(E))^{\mathbf{I}} = \mathrm{supp}(E^{\mathbf{I}})$.

To simplify the notation in relational algebra expressions, we will often use R_i to denote a relation symbol, instead of $\hat{R}_i$. Also, when the schema τ is understood from the context, we will write $\mathcal{BRA}$, instead of $\mathcal{BRA}(\tau)$, and will talk about $\mathcal{BRA}$-expressions, instead of $\mathcal{BRA}(\tau)$-expressions.

Remark 1 (Necessity of Empty Projections). *Codd [8] did not allow for empty projections* $\pi_{[\,]}(R)$ *in the definition of relational algebra. If empty projections are disallowed, then all relational algebra expressions have positive arity, thus they cannot express* 0*-ary queries; in particular, they cannot express queries expressible by first-order sentences. In subsequent expositions of relational algebra (e.g., in [1]), empty projections are allowed, which is the choice we made here.*

Remark 2 (Necessity of Support). *In Codd's original paper on the relational data model [7], the support operation was not among the operations considered. There is a good reason for this: on the Boolean semiring,* $\mathrm{supp}(E) = E$ *holds for every relational algebra expression* E*. On an arbitrary semiring* $\mathbb{K}$*, though, the support is needed as a separate operation, because the support may not be expressible using the other relational algebra operations. For example, for the fuzzy semiring* $\mathbb{F} = ([0,1], \max, \min, 0, 1)$*, the support operation cannot be expressed in terms of the other five relational algebra operations. Indeed, assume that* τ *consists of a single relation symbol* R *of arity* 1*. Let* $\mathbf{I}$ *be the* $\mathbb{F}$*-database in which* R *is the* $\mathbb{F}$*-relation with* $R(a) = 0.5$ *and* $R(b) = 0$*, for all* $b \neq a$*. A straightforward induction shows that if* E *is a relational algebra expression (not involving support) over* τ*, then* $E(\mathbf{I})$ *is an* $\mathbb{F}$*-relation such that* $E^{\mathbf{I}}(\mathbf{t}) \in [0, 0.5]$ *if* $\mathbf{t} = (a, \ldots, a)$*, and* $E^{\mathbf{I}}(\mathbf{t}) = 0$ *if* $\mathbf{t} \neq (a, \ldots, a)$*. Thus,* $E^{\mathbf{I}} \neq \mathrm{supp}(R)$*. Later on, the support will be used to express the active domain of a* $\mathbb{K}$*-database.*

3.2 Basic Relational Calculus

We will now introduce a first-order language called *basic relational calculus*, which has existential quantification and a limited form of negation, but not universal quantification. Let $\tau = (\hat{R}_1, \ldots \hat{R}_p)$ be a schema. The building blocks of the *basic relational calculus over* τ, denoted by $\mathcal{BRC}(\tau)$, consist of a countable set $V = \{x_1, \ldots, x_n, \ldots\}$ of variables, a distinguished binary relation $=$ (equality), the relation symbols of τ, three binary connectives $\wedge$ ('and'), $\vee$ ('or') and $\prec$ ('but not'), a unary connective ∇ ('it is not false that'), and the existential quantifier $\exists$ ('there exists'). In what follows, we will write $y_1, \ldots, y_i, \ldots$ to denote *meta-variables*, i.e., each y_i is one of the variables in the set V.

Definition 9 (Basic Relational Calculus). *Let* $\tau = (\hat{R}_1, \ldots \hat{R}_p)$ *be a schema. The* basic relational calculus language $\mathcal{BRC}(\tau)$ over τ *is the set of all expressions defined recursively as follows:*

- *If* y_i *and* y_j *are variables, then the expression* $y_i = y_j$ *is a* $\mathcal{BRC}(\tau)$*-formula with* y_i *and* y_j *as its free variables.*
- *If* $\hat{R}_i$ *is a relation symbol of arity* r_i *and if* $y_1, \ldots, y_{r_i}$ *are variables, then the expression* $\hat{R}_i(y_1, \ldots, y_{r_i})$ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_1, \ldots, y_{r_i}$ *as its free variables.*
- *If* φ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_1, \ldots, y_n$ *as its free variables and if* ψ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_{n+1}, \ldots, y_{n+k}$ *as its free variables, then the expressions* $\varphi \vee \psi$*,* $\varphi \wedge \psi$*, and* $\varphi \prec \psi$ *are* $\mathcal{BRC}(\tau)$*-formulas, and each of them has* $y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+k}$ *as its free variables.*
- *If* φ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_1, \ldots, y_n$ *as its free variables, then the expression* $\nabla\varphi$ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_1, \ldots, y_n$ *as its free variables.*
- *If* φ *is a* $\mathcal{BRC}(\tau)$*-formula with* $y_1, \ldots, y_n$ *as its free variables, then the expression* $\exists y_i \varphi$ *is a* $\mathcal{BRC}(\tau)$*-formula with the variables* $y_1, \ldots, y_n$ *other than* y_i *as its free variables.*

The notation $\varphi(y_1, \ldots, y_n)$ *denotes a* $\mathcal{BRC}(\tau)$*-formula* φ *whose free variables are* $y_1, \ldots, y_n$*.*

Definition 10. *Let $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ be a semiring with monus and support, and let $\tau = (\hat{R}_1, \ldots, \hat{R}_p)$ be a schema.*

- *A* $\mathbb{K}$-structure *over τ is a tuple $\mathbf{A} = (A, R_1, \ldots, R_p)$, where A is a non-empty subset of D, called the* universe *of $\mathbf{A}$, and each R_i is a $\mathbb{K}$-relation on A whose arity matches the arity of the relation symbol $\hat{R}_i$, $1 \le i \le p$. We say that $\mathbf{A}$ is a* finite $\mathbb{K}$-structure *if its universe A is a finite set.*
- *An* assignment *on $\mathbf{A}$ is a function $\alpha \colon V \longrightarrow A$, assigning to every variable an element in A.*
- *If α is an assignment on $\mathbf{A}$ and b is an element of A, then we write $\alpha[x_i/b]$ for the assignment β on $\mathbf{A}$ such that $\beta(x_i) = b$ and $\beta(x_j) = \alpha(x_j)$, for every variable $x_j \neq x_i$.*

In what follows, we make the blanket assumption that all $\mathbb{K}$-structures considered are finite.

Definition 11. *Let $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ be a semiring with monus and support, let $\tau = (\hat{R}_1, \ldots, \hat{R}_p)$ be a schema, and let φ be a $\mathcal{BRC}(\tau)$-formula. The* semantics of φ on $\mathbb{K}$ *is a function $\|\varphi\|$ that takes as input a $\mathbb{K}$-structure $\mathbf{A} = (A, R_1, \ldots, R_p)$ and an assignment α on $\mathbf{A}$, and returns a value $\|\varphi\|(\mathbf{A}, \alpha)$ in K that is defined recursively and simultaneously for all assignments as follows:*

- $\|x_i = x_j\|(\mathbf{A}, \alpha) = 1$, *if* $\alpha(x_i) = \alpha(x_j)$; $\|x_i = x_j\|(\mathbf{A}, \alpha) = 0$, *if* $\alpha(x_i) \neq \alpha(x_j)$.
- $\|\hat{R}_i(x_{i_1}, \ldots, x_{i_n})\|(\mathbf{A}, \alpha) = R_i(\alpha(x_{i_1}), \ldots \alpha(x_{i_n}))$, *for* $1 \le i \le p$.
- $\|\varphi_1 \wedge \varphi_2\|(\mathbf{A}, \alpha) = \|\varphi_1\|(\mathbf{A}, \alpha) \cdot \|\varphi_2\|(\mathbf{A}, \alpha)$; $\|\varphi_1 \vee \varphi_2\|(\mathbf{A}, \alpha) = \|\varphi_1\|(\mathbf{A}, \alpha) + \|\varphi_2\|(\mathbf{A}, \alpha)$.
- $|\varphi_1 \prec \varphi_2\|(\mathbf{A}, \alpha) = \|\varphi_1\|(\mathbf{A}, \alpha) - \|\varphi_2\|(\mathbf{A}, \alpha)$; $\|\nabla\varphi\|(\mathbf{A}, \alpha) = \mathrm{s}(\|\varphi\|(\mathbf{A}, \alpha))$.
- $\|\exists x_i \varphi\|(\mathbf{A}, \alpha) = \sum_{b \in A} \|\varphi\|(\mathbf{A}, \alpha[x_i/b])$.

The intent of the first bullet is to incorporate standard equality in the syntax and semantics of basic relational calculus. Note that the semantics of the existential quantifier is well defined because of our blanket assumption that $\mathbf{A}$ has a finite universe. The semantics is still meaningful for $\mathbb{K}$-structures with infinite universes, provided the semiring $\mathbb{K}$ considered has additional closure properties. In particular, this holds true when $\mathbb{K}$ is a complete bounded distributive lattice.

Using a straightforward induction, one can show that the value $\|\varphi\|(\mathbf{A}, \alpha)$ depends only on the values of the assignment α on the free variables of φ, according to the following lemma.

Lemma 1 (Relevance Lemma). *Let τ be a schema and $\varphi(y_1, \ldots, y_n)$ be a $\mathcal{BRC}(\tau)$-formula whose free variables are $y_1, \ldots, y_n$, and let $\mathbf{A}$ be a (τ)-structure. If α and β are two assignments on $\mathbf{A}$ such that $\alpha(y_i) = \beta(y_i)$ for every free variable y_i of φ, then $\|\varphi\|(\mathbf{A}, \alpha) = \|\varphi\|(\mathbf{A}, \beta)$.*

If $\varphi(y_1, \ldots, y_n)$ is a formula and $(a_1, \ldots, a_n)$ is a sequence of elements from the universe A of a $\mathbb{K}$-structure $\mathbf{A}$, then we will write $\|\varphi\|(\mathbf{A}, (a_1, \ldots, a_n))$ to denote the value $\|\varphi\|(\mathbf{A}, \alpha)$, where α is any assignment on $\mathbf{A}$ such that $\alpha(y_1) = a_1, \ldots, \alpha(y_n) = a_n$. By the preceding Relevance Lemma, this value is well defined. Furthermore, if φ is a *sentence* (i.e., φ has no free variables), we will write $\|\varphi\|(\mathbf{A}, <>)$ to denote $\|\varphi\|(\mathbf{A}, \alpha)$, where α is an arbitrary assignment; note that in this case, $\|\varphi\|(\mathbf{A}, \alpha)$ is the same value, independently of the assignment α.

Definition 12. *Let $\mathbb{K}$ be a semiring with monus and support, and let τ be a schema.*

- *If $\varphi(y_1, \ldots, y_n)$ is a $\mathcal{BRC}(\tau)$-formula whose free variables are $y_1, \ldots, y_n$ and if $\mathbf{A}$ is a $\mathbb{K}$-structure, then $\varphi^{\mathbf{A}} \colon D^n \longrightarrow K$ is the n-ary $\mathbb{K}$-relation defined as follows:*
$$\varphi^{\mathbf{A}}(a_1, \ldots, a_n) = \begin{cases} \|\varphi\|(\mathbf{A}, (a_1, \ldots, a_n)) & \text{if } (a_1, \ldots, a_n) \in A^n \\ 0 & \text{if } (a_1, \ldots, a_n) \notin A^n \end{cases}$$
Note that $\varphi^{\mathbf{A}}$ is actually a $\mathbb{K}$-relation on A since, obviously, $\mathrm{supp}(\varphi^{\mathbf{A}}) \subseteq A^n$.
- *If φ is a $\mathcal{BRC}(\tau)$-sentence and if $\mathbf{A}$ is a $\mathbb{K}$-structure, then $\varphi^{\mathbf{A}} \colon D^0 \longrightarrow K$ is the 0-ary $\mathbb{K}$-relation such that $\varphi^{\mathbf{A}}(<>) = \|\varphi\|(\mathbf{A}, <>)$.*

The next proposition describes some useful properties of the semantics of $\mathcal{BRC}$-formulas.

Proposition 2. *Let $\mathbb{K}$ be a semiring with monus and support, τ a schema, and $\mathbf{A} = (A, R_1 \ldots, R_p)$ a $\mathbb{K}$-structure. The following statements are true.*

(1) *Consider the atomic formula $\hat{R}_i(y_1, \ldots, y_n)$, where $\hat{R}_i$ is one of the relation symbols of the schema τ, n is the arity of $\hat{R}_i$, and the y_i's are distinct variables. Then $(\hat{R}_i(y_1, \ldots, y_n))^{\mathbf{A}} = R_i$.*
(2) *Let $\varphi_1(y_1, \ldots, y_n)$ and $\varphi_2(y_1, \ldots, y_n)$ be two $\mathcal{BRC}(\tau)$-formulas with the same free variables. Then $(\varphi_1 \vee \varphi_2)^{\mathbf{A}} = \varphi_1^{\mathbf{A}} \cup \varphi_2^{\mathbf{A}}$ and $(\varphi_1 \prec \varphi_2)^{\mathbf{A}} = \varphi_1^{\mathbf{A}} \setminus \varphi_2^{\mathbf{A}}$.*
(3) *Let $\varphi_1(y_1, \ldots, y_n)$ and $\varphi_2(y_{n+1}, \ldots, y_{n+m})$ be two $\mathcal{BRC}(\tau)$-formulas, where $y_1, \ldots, y_{n+m}$ are distinct variables. Then $(\varphi_1 \wedge \varphi_2)^{\mathbf{A}} = \varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}}$.*
(4) *Let $\varphi_1(y_1, \ldots, y_n)$ be a $\mathcal{BRC}(\tau)$-formula. Then $(\nabla\varphi)^{\mathbf{A}} = \mathrm{supp}(\varphi^{\mathbf{A}})$.*
(5) *Let $\varphi(y_1, \ldots, y_n)$ be a $\mathcal{BRC}(\tau)$-formula, i an index with $1 \leq i \leq n$, and $v = (i_1, \ldots, i_{n-1})$ the sequence of distinct indices from $\{1, \ldots, n\} \setminus \{i\}$ ordered by the standard ordering of the natural numbers. Then $(\exists y_i \varphi)^{\mathbf{A}} = \pi_v(\varphi^{\mathbf{A}})$.*

Proof. The first part about the atomic formula $\hat{R}_i(y_1, \ldots, y_n)$ follows from Definitions 11 and 12.

For the second part, let φ_1 and φ_2 be two $\mathcal{BRC}(\tau)$-formulas with $y_1, \ldots, y_n$ as their free variables, and let $\mathbf{A}$ be a $\mathbb{K}$-structure. We have to show that $(\varphi_1 \vee \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_n) = (\varphi_1^{\mathbf{A}} \cup \varphi_2^{\mathbf{A}})(a_1, \ldots, a_n)$ holds, for every tuple $(a_1, \ldots, a_n) \in D^n$. We distinguish two cases for the tuple $(a_1, \ldots, a_n) \in D^n$, namely, whether or not $(a_1, \ldots, a_n) \in A^n$, where A is the universe of $\mathbf{A}$.

If $(a_1, \ldots, a_n) \in A^n$, then we have that

$$(\varphi_1 \vee \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_n) = \|\varphi_1 \vee \varphi_2\|(\mathbf{A}, (a_1, \ldots, a_n)) = \|\varphi_1\|(\mathbf{A}, (a_1, \ldots, a_n)) + \|\varphi_2\|(\mathbf{A}, (a_1, \ldots, a_n)) =$$
$$\varphi_1^{\mathbf{A}}(a_1, \ldots, a_n) + \varphi_2^{\mathbf{A}}(a_1, \ldots, a_n) = (\varphi_1^{\mathbf{A}} \cup \varphi_2^{\mathbf{A}})(a_1, \ldots, a_n),$$

where the first equation follows from Definition 12, the second follows from the semantics of $\mathcal{BRC}(\tau)$-formulas in Definition 11, the third follows from Definition 12, and the last follows from the definition of operations on $\mathbb{K}$-relations in Definition 6.

If $(a_1, \ldots, a_n) \notin A^n$, then $(\varphi_1 \vee \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_n) = 0$ and also $\varphi_1^{\mathbf{A}}(a_1, \ldots, a_n) = 0 = \varphi_1^{\mathbf{A}}(a_1, \ldots, a_n)$ by Definition 12. Therefore,

$$(\varphi_1 \vee \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_n) = 0 = \varphi_1^{\mathbf{A}}(a_1, \ldots, a_n) + \varphi_2^{\mathbf{A}}(a_1, \ldots, a_n) = (\varphi_1^{\mathbf{A}} \cup \varphi_2^{\mathbf{A}})(a_1, \ldots, a_n).$$

The proof that $(\varphi_1 \prec \varphi_2)^{\mathbf{A}} = \varphi_1^{\mathbf{A}} \setminus \varphi_2^{\mathbf{A}}$ is entirely analogous to the preceding one, but with the connective $\prec$ in the role of $\vee$, and with the connective $\setminus$ in the role of $\cup$.

For the third part, assume that $\varphi_1(y_1, \ldots, y_n)$ and $\varphi_2(y_{n+1}, \ldots, y_{n+m})$ are two $\mathcal{BRC}(\tau)$-formulas, where $y_1, \ldots, y_{n+m}$ are distinct variables and let $\mathbf{A}$ be a $\mathbb{K}$-structure. To show that $(\varphi_1 \wedge \varphi_2)^{\mathbf{A}} = \varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}}$, we have to show that for every tuple $(a_1, \ldots, a_{n+m}) \in D^n$, we have that

$$(\varphi_1 \wedge \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_{n+m}) = (\varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}})(a_1, \ldots, a_{n+m}).$$

We distinguish two cases for the tuple $(a_1, \ldots, a_n) \in D^n$. First, if $(a_1, \ldots, a_{n+m}) \in A^{n+m}$, then

$$(\varphi_1 \wedge \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_{n+m}) = \|\varphi_1 \wedge \varphi_2\|(\mathbf{A}, (a_1, \ldots, a_{n+m})) =$$
$$\|\varphi_1\|(\mathbf{A}, (a_1, \ldots, a_n)) \cdot \|\varphi_2\|(\mathbf{A}, (a_{n+1}, \ldots, a_{n+m})) =$$
$$\varphi^{\mathbf{A}}(a_1, \ldots, a_n) \cdot \varphi^{\mathbf{A}}(a_{n+1}, \ldots, a_{n+m}) = (\varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}})(a_1, \ldots, a_{n+m}),$$

where the first equation follows from Definition 12, the second follows the semantics of $\mathcal{BRC}(\tau)$-formulas in Definition 11 and from the Relevance Lemma 1, the third follows from Definition 12, and the last follows from the definition of the operations on $\mathbb{K}$-relations in Definition 6.

Second, if $(a_1, \ldots, a_{n+m}) \notin A^{n+m}$, then $(\varphi_1 \wedge \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_{n+m}) = 0$ by Definition 12. Furthermore, we have that $(a_1, \ldots, a_n) \notin A^n$ or $(a_{n+1}, \ldots, a_{n+m}) \notin A^m$. By Definition 12, we have that $\varphi_1^{\mathbf{A}}(a_1, \ldots, a_n) = 0$ or $\varphi_2^{\mathbf{A}}(a_{n+1}, \ldots, a_{n+m}) = 0$. In either case, we have that

$$0 = \varphi^{\mathbf{A}}(a_1, \ldots, a_n) \cdot \varphi^{\mathbf{A}}(a_{n+1}, \ldots, a_{n+m}) = (\varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}})(a_1, \ldots, a_{n+m}),$$

hence $(\varphi_1 \wedge \varphi_2)^{\mathbf{A}}(a_1, \ldots, a_{n+m}) = (\varphi_1^{\mathbf{A}} \times \varphi_2^{\mathbf{A}})(a_1, \ldots, a_{n+m})$.

For the fourth part, assume that $\varphi(y_1, \ldots, y_n)$ is a $\mathcal{BRC}(\tau)$-formula and $\mathbf{A} = (A, R_1 \ldots, R_p)$ is a $\mathbb{K}$-structure. To show that $(\nabla\varphi)^{\mathbf{A}} = \mathrm{supp}(\varphi^{\mathbf{A}})$, we have to show that for every tuple $(a_1, \ldots, a_n) \in D^n$, we have that $(\nabla\varphi)^{\mathbf{A}}(a_1, \ldots, a_n) = \mathrm{supp}(\varphi^{\mathbf{A}})(a_1, \ldots, a_n)$. We distinguish two cases for the tuple $(a_1, \ldots, a_n) \in D^n$. If $(a_1, \ldots, a_n) \in D^n$, then, by Definition 12,

$$(\nabla\varphi)^{\mathbf{A}}(a_1, \ldots, a_n) = \begin{cases} 1 & \text{if } \varphi^{\mathbf{A}}(a_1, \ldots, a_n) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $(\nabla\varphi)^{\mathbf{A}}(a_1, \ldots, a_n) = \mathrm{supp}(\varphi^{\mathbf{A}})(a_1, \ldots, a_n)$. If $(a_1, \ldots, a_n) \notin D^n$, then, by Definition 12, both $(\nabla\varphi)^{\mathbf{A}}(a_1, \ldots, a_n) = 0$ and $\varphi^{\mathbf{A}}(a_1, \ldots, a_n) = 0$, so $\mathrm{supp}(\varphi^{\mathbf{A}})(a_1, \ldots, a_n) = 0 = (\nabla\varphi)^{\mathbf{A}}(a_1, \ldots, a_n)$.

For the fifth part, assume that $\varphi(y_1, \ldots, y_n)$ is a $\mathcal{BRC}(\tau)$-formula, i is an index with $1 \leq i \leq n$, and $v = (i_1, \ldots, i_{n-1})$ is the sequence of distinct indices from $\{1, \ldots, n\} \setminus \{i\}$ ordered by the standard ordering of the natural numbers. Let $\mathbf{A} = (A, R_1 \ldots, R_p)$ is a $\mathbb{K}$-structure. To show that $(\exists y_i \varphi)^{\mathbf{A}} = \pi_v(\varphi^{\mathbf{A}})$, we have to show that for every tuple $(a_1, \ldots, a_{n-1}) \in D^{n-1}$, we have that $(\exists y_i \varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1}) = \pi_v(\varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1})$. We distinguish two cases for the tuple $(a_1, \ldots, a_{n-1}) \in D^{n-1}$, namely, whether or not $(a_1, \ldots, a_{n-1}) \in A^{n-1}$. Suppose that $(a_1, \ldots, a_{n-1}) \in A^{n-1}$. From Definition 6 of the operations on $\mathbb{K}$-relations, we have that

$$\pi_v(\varphi^{\mathbf{A}})(a_1, \ldots, a_{n-1}) = \sum_{\mathbf{b}[v]=(a_1,\ldots,a_{n-1}) \text{ and } \varphi^{\mathbf{A}}(\mathbf{b})\neq 0} \varphi^{\mathbf{A}}(\mathbf{b}).$$

Furthermore, we have that

$$(\exists y_i \varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1}) = \|\exists y_i \varphi\|(\mathbf{A}, (a_1, \ldots a_{n-1})) = \sum_{\substack{\mathbf{b} \in A^n \\ \mathbf{b}[v]=(a_1,\ldots,a_{n-1})}} \|\varphi\|(\mathbf{A}, \mathbf{b}) = \sum_{\substack{\mathbf{b} \in A^n \\ \mathbf{b}[v]=(a_1,\ldots,a_{n-1})}} \varphi^{\mathbf{A}}(\mathbf{b}),$$

where the first and the third equations follow from Definition 12 and the second equation follows from the semantics of $\mathcal{BRC}(\tau)$-formulas in Definition 11. Observe that $\varphi_1^{\mathbf{A}}(\mathbf{b}) \neq 0$ only if $\mathbf{b}$ is a tuple in A^n (since, if $\mathbf{b}$ is not a tuple in A^n, then $\varphi_1^{\mathbf{A}}(\mathbf{b}) = 0$ by Definition 12). Thus,

$$(\exists y_i \varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1}) = \sum_{\mathbf{b}[v]=(a_1,\ldots,a_{n-1}) \text{ and } \varphi^{\mathbf{A}}(\mathbf{b})\neq 0} \varphi^{\mathbf{A}}(\mathbf{b}) = \pi_v(\varphi^{\mathbf{A}})(a_1, \ldots, a_{n-1}).$$

Suppose that $(a_1, \ldots, a_{n-1}) \notin A^{n-1}$. By Definition 11, we have $(\exists y_i \varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1}) = 0$. Also, if $\mathbf{b} \in D^n$ is a tuple such that $\mathbf{b}[v] = (a_1, \ldots, a_{n-1})$, then $\mathbf{b} \notin A^n$, hence, by Definition 11 again, $\varphi^{\mathbf{A}}(\mathbf{b}) = 0$. It follows that $\pi_v(\varphi_1^{\mathbf{A}})(a_1, \ldots, a_{n-1}) = 0$ and so $(\exists y_i \varphi)^{\mathbf{A}}(a_1, \ldots, a_{n-1}) = \pi_v(\varphi^{\mathbf{A}})(a_1, \ldots, a_{n-1})$. □

Remark 3. *Let $x_i \neq x_j$ be the formula $(x_i = x_i) \prec (x_i = x_j)$. It is easy to verify that for every $\mathbb{K}$-structure $\mathbf{A}$ and every two elements a and b in the universe of $\mathbf{A}$, we have that*

$$\|x_i \neq x_j\|(\mathbf{A}, a, b) = \begin{cases} 1 & \text{if } a \neq b \\ 0 & \text{if } a = b. \end{cases}$$

Therefore, $\mathcal{BRC}$ can express negated equalities, a fact that we will use in the sequel.

Remark 4. *In the case of the Boolean semiring $\mathbb{B}$, the connectives $\wedge, \vee, \neg$ are inter-definable with the connectives $\wedge, \vee, \prec$. In particular, $\varphi \prec \psi$ has the same semantics on the Boolean semiring as $\varphi \wedge \neg\psi$, while $\neg\varphi$ has the same semantics as $(x_1 = x_1) \prec \varphi$. This does not hold for arbitrary semirings. Concretely, it is not true that on every complete bounded distributive lattice (see Example 3), the connectives $\wedge, \vee, \neg$ are inter-definable with the connectives $\wedge, \vee, \prec, \top$. Such examples can be extracted from [21, 25] by considering the algebraic* duals *(i.e., the algebras where one reverses the order and swaps meets with joins) of the complete bounded distributive lattices presented there.*

3.3 Basic Relational Algebra vs. Basic Relational Calculus

In this section, $\mathbb{K} = (K, +, \cdot, -, \mathrm{s}, 0, 1)$ is a fixed semiring with monus and support, and $\tau = (\hat{R_1}, \ldots, \hat{R_p})$ is a fixed schema. All $\mathbb{K}$-databases and $\mathbb{K}$-structures considered are over τ.

Definition 13 (Active domain). *Let* $\mathbf{I} = (R_1, \ldots, R_p)$ *be a* $\mathbb{K}$*-database. The* active domain *of* $\mathbf{I}$*, denoted by* $\mathrm{adom}(\mathbf{I})$*, is the set of all elements of* D *that occur in the support of a relation* R_i *of* $\mathbf{I}$*.*

Since we made the blanket assumption that we only consider non-trivial $\mathbb{K}$-databases (see Definition 8), it follows that the active domain $\mathrm{adom}(\mathbf{I})$ is a non-empty set. Furthermore, since every relation R_i, $1 \le i \le p$, of $\mathbf{I}$ has finite support, it follows that $\mathrm{adom}(\mathbf{I})$ is a finite subset of the domain D. Thus, $\mathrm{adom}(\mathbf{I})$ can be identified with a unary $\mathbb{K}$-relation that takes values 0 and 1 only. There is a $\mathcal{BRA}$-relational algebra expression that uniformly defines the active domain on every $\mathbb{K}$-database, as spelled out in the next proposition. This is a basic fact that will be repeatedly used.

Proposition 3. *Let* $\mathbb{K}$ *be a semiring with monus and support, and let* τ *be a schema. Then there is a* $\mathcal{BRA}(\tau)$*-expression* E_{adom} *such that for every* $\mathbb{K}$*-database* $\mathbf{I}$*, we have that* $E^{\mathbf{I}}_{\mathrm{adom}} = \mathrm{adom}(\mathbf{I})$*.*

Proof. For concreteness, let us assume that τ consists of a single binary relation symbol $\hat{R}$. The argument presented below extends easily to arbitrary schemas with only slight modifications.

Let E_{adom} be the $\mathcal{BRA}(\tau)$-relational algebra expression $\mathrm{supp}(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))$. We claim that for every $\mathbb{K}$-database, we have that $E^{\mathbf{I}}_{\mathrm{adom}} = \mathrm{adom}(\mathbf{I})$.

Let $\mathbf{I} = (R)$ be a $\mathbb{K}$-database and let a be an element of the domain D. First, observe that, using the definitions of the semantics of $\mathcal{BRA}$-expressions, we have that

$$(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))^{\mathbf{I}}(a) = ((\pi_1(\hat{R}))^{\mathbf{I}} \cup (\pi_2(\hat{R}))^{\mathbf{I}})(a) = \sum_{R(a,b)\neq 0} R(a,b) \; + \sum_{R(b,a)\neq 0} R(b,a).$$

If a is in the active domain $\mathrm{adom}(\mathbf{I})$ of $\mathbf{I}$, there is at least one element $b \in D$ such that $R(a,b) \neq 0$ or $R(b,a) \neq 0$. Therefore, at least one summand in at least one of the two sums above is different from 0. Since $\mathbb{K}$ is a zero-sum-free semiring, it follows that $\sum_{R(a,b)\neq 0} R(a,b) \; + \sum_{R(b,a)\neq 0} R(b,a) \neq 0$, hence $(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))^{\mathbf{I}}(a) \neq 0$. Consequently, $a \in (\mathrm{supp}(\pi_1(\hat{R}) \cup \pi_2(\hat{R})))^{\mathbf{I}} = E^{\mathbf{I}}_{\mathrm{adom}}$.

If a is not in the active domain of $\mathbf{I}$, then each summand in both of the two sums above is equal to 0, hence $(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))^{\mathbf{I}}(a) = 0$ and so $a \notin (\mathrm{supp}(\pi_1(\hat{R}) \cup \pi_2(\hat{R})))^{\mathbf{I}} = E^{\mathbf{I}}_{\mathrm{adom}}$. Note that in this step we did not use the hypothesis that $\mathbb{K}$ is a zero-sum-free semiring. □

Remark 5. *As seen earlier, if* $\mathbb{K}$ *is a semiring with monus, then it is zero-sum free. If* $\mathbb{K}$ *is not zero-sum-free, then the* $\mathcal{BRA}$*-expression* $\mathrm{supp}(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))$ *does not define the active domain. Indeed, suppose that* k_1 *and* k_2 *are two elements in the universe* K *of* $\mathbb{K}$ *such that* $k_1 + k_2 = 0$*. Let* a, b *be two distinct elements of the domain* D *and consider the* $\mathbb{K}$*-database* $\mathbf{I} = (R)$*, where* $R(a,b) = k_1$*,* $R(b,a) = k_2$*, and* $R(c,d) = 0$*, for all other pairs* (c,d) *from* D*. Then* $\mathrm{adom}(\mathbf{I}) = \{a, b\}$*, but* $(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))^{\mathbf{I}}(a) = k_1 + k_2 = 0$ *and* $(\pi_1(\hat{R}) \cup \pi_2(\hat{R}))^{\mathbf{I}}(b) = k_2 + k_1 = 0$*, hence* $E^{\mathbf{I}}_{\mathrm{adom}} = \emptyset \neq \mathrm{adom}(\mathbf{I})$*.*

Definition 14 (Query). *Let* $n \ge 0$*. An* n*-ary* query q *is a mapping defined on* $\mathbb{K}$*-databases and such that on every* $\mathbb{K}$*-database* $\mathbf{I}$*,* q *returns an* n*-ary* $\mathbb{K}$*-relation* $q^{\mathbf{I}}$ *on* $\mathrm{adom}(\mathbf{I})$*.*

Clearly, every $\mathcal{BRA}$-expression E of arity $n \ge 0$ gives rise to an n-ary query associating with every $\mathbb{K}$-database $\mathbf{I}$ the n-ary $\mathbb{K}$-relation $E^{\mathbf{I}}$ as defined in Definition 8.

Definition 15. *Given a* $\mathbb{K}$*-database* $\mathbf{I} = (R_1, \ldots, R_p)$*, we write* $\mathbf{A}(\mathbf{I})$ *to denote the* $\mathbb{K}$*-structure with* $\mathrm{adom}(\mathbf{I})$ *as its universe and with relations those of* $\mathbf{I}$*. In symbols,* $\mathbf{A}(\mathbf{I}) = (\mathrm{adom}(\mathbf{I}), R_1, \ldots, R_p)$*.*

Clearly, every $\mathcal{BRC}$-formula φ with $n \ge 0$ free variables, gives to an n-ary query associating with every $\mathbb{K}$-database $\mathbf{I}$ the n-ary $\mathbb{K}$-relation $\varphi^{\mathbf{A}(\mathbf{I})}$ as defined in Definition 12.

Definition 16. *A $\mathcal{BRC}$-formula φ is* domain independent *if for every $\mathbb{K}$-database $\mathbf{I} = (R_1, \ldots, R_p)$ and every $\mathbb{K}$-structure $\mathbf{B} = (B, R_1, \ldots, R_p)$ with* $\mathrm{adom}(\mathbf{I}) \subseteq B \subseteq D$*, we have that $\varphi^{\mathbf{B}} = \varphi^{\mathbf{A}(\mathbf{I})}$.*

The next proposition will be used in the proofs of the main results; it can be proved in a straightforward way using Proposition 2.

Proposition 4. *Let $\tau = (\hat{R}_1, \ldots \hat{R}_p)$ be a schema. The following statements are true.*

(1) *Let $\varphi_1(y_1, \ldots, y_n)$ and $\varphi_2(y_1, \ldots, y_n)$ be two $\mathcal{BRC}(\tau)$-formulas with the same free variables. If φ_1 and φ_2 are domain independent $\mathcal{BRC}$-formulas, then so are $\varphi_1 \vee \varphi_2$ and $\varphi_1 \prec \varphi_2$.*
(2) *If $\varphi_1(y_1, \ldots, y_n)$ and $\varphi_2(y_{n+1}, \ldots, y_{n+m})$ are two domain independent $\mathcal{BRC}(\tau)$-formulas, where $y_1, \ldots, y_{n+m}$ are distinct variables, then so are $\varphi_1 \wedge \varphi_2$.*
(3) *If $\varphi_1(y_1, \ldots, y_n)$ is a domain independent $\mathcal{BRC}(\tau)$-formula, then so is $\nabla\varphi$.*
(4) *If $\varphi(y_1, \ldots, y_n)$ is a is a domain independent $\mathcal{BRC}(\tau)$-formula, then so is $\exists y_i \varphi$.*

We now present the first main theorem about basic relational algebra and basic relational calculus.

Theorem 1. *Let $\mathbb{K}$ be a zero-sum-free semiring with monus and support, let $\tau = (\hat{R}_1, \ldots \hat{R}_p)$ be a schema, and let q be an n-ary query. The following statements are equivalent:*

(1) *There is a $\mathcal{BRA}(\tau)$-expression E of arity n such that $q^{\mathbf{I}} = E^{\mathbf{I}}$, for every $\mathbb{K}$-database $\mathbf{I}$.*
(2) *There is a domain independent $\mathcal{BRC}(\tau)$-formula $\varphi(y_1, \ldots, y_n)$ such that $q^{\mathbf{I}} = \varphi^{\mathbf{A}(\mathbf{I})}$, for every $\mathbb{K}$-database $\mathbf{I}$.*
(3) *There is a $\mathcal{BRC}(\tau)$-formula $\varphi(y_1, \ldots, y_n)$ such that $q^{\mathbf{I}} = \varphi^{\mathbf{A}(\mathbf{I})}$, for every $\mathbb{K}$-database $\mathbf{I}$.*

Proof. The implication (2) $\Longrightarrow$ (3) is obvious. The implications (1) $\Longrightarrow$ (2) and (3) $\Longrightarrow$ (1) are established by induction on the construction of $\mathcal{BRA}(\tau)$-expressions and of $\mathcal{BRC}(\tau)$-formulas, respectively. In some steps in the proof of (3) $\Longrightarrow$ (1), we will make some simplifying assumptions about the arity of the relational algebra expressions and the pattern of free variables of the relational calculus formulas. The proof extends to the general case with minor modifications only.
(1) $\Longrightarrow$ (2) : We proceed by induction on the construction of $\mathcal{BRA}(\tau)$-expressions. For every such expression E, we produce a $\mathcal{BRC}(\tau)$-formula φ_E such that $E^{\mathbf{I}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, for every $\mathbb{K}$-database $\mathbf{I}$. Moreover, in each case, we show that the $\mathcal{BRC}(\tau)$-formula φ_E is domain independent.
Case 1: Assume that E is one of the relation symbols $\hat{R}_i$, $1 \leq i \leq p$. In this case, we take φ_E to be the atomic formula $\hat{R}_i(x_1, \ldots, x_n)$, where n is the arity of $\hat{R}_i$. To see why this works, let $\mathbf{I} = (R_1, \ldots, R_p)$ be a $\mathbb{K}$-database. From Definition 8 of the semantics of $\mathcal{BRA}(\tau)$-expressions and the first part of Proposition 2, we have $E^{\mathbf{I}} = R_i = (\hat{R}_i(x_1, \ldots, x_n))^{\mathbf{A}(\mathbf{I})} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, which establishes the correctness of the formula φ_E. Furthermore, the first part of Proposition 2 immediately implies the domain independence of the formula φ_E.
Case 2: Assume that E is a $\mathcal{BRA}(\tau)$-expression of the form $E_1 \cup E_2$ or of the form $E_1 \setminus E_2$, where E_1 and E_2 are $\mathcal{BRA}(\tau)$-expressions of the same arity, say $n \geq 0$. By induction hypothesis, there are domain independent $\mathcal{BRC}(\tau)$-formulas φ_1 and φ_2 each with n free variables such that for every $\mathbb{K}$-database $\mathbf{I}$, we have that $E_i^{\mathbf{I}} = \varphi_i^{\mathbf{A}(\mathbf{I})}$, for $i = 1, 2$. If E is $E_1 \cup E_2$, we take φ_E to be the $\mathcal{BRC}(\tau)$-formula $\varphi_1 \vee \varphi_2$; if E is $E_1 \setminus E_2$, we take φ_E to be the $\mathcal{BRC}(\tau)$-formula $\varphi_1 \prec \varphi_2$. The correctness of the formula φ_E follows from the semantics of $\mathcal{BRA}(\tau)$-expressions in Definition 8, the inductive hypothesis, and Proposition 2. Specifically, if $\mathbf{I}$ is a $\mathbb{K}$ database, then $E^{\mathbf{I}} = (E_1 \cup E_2)^{\mathbf{I}} = E_1^{\mathbf{I}} \cup E_2^{\mathbf{I}} = \varphi_1^{\mathbf{A}(\mathbf{I})} \cup \varphi_2^{\mathbf{A}(\mathbf{I})} = (\varphi_1 \vee \varphi_2)^{\mathbf{A}(\mathbf{I})} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, as desired. The proof for the case of the difference operation $\setminus$ is entirely analogous. The domain independence of the formula φ_E follows from the inductive hypothesis and the first part of Proposition 4.
Case 3: Assume that E is a $\mathcal{BRA}(\tau)$-expression of the form $E_1 \times E_2$, where E_1 is a $\mathcal{BRA}(\tau)$-expression of arity n and E_2 is a $\mathcal{BRA}(\tau)$-expression of arity m. By induction hypothesis, there are

domain independent formulas $\mathcal{BRC}(\tau)$-formulas φ_1 and φ_2 such that φ_1 has n free variables, φ_2 has m free variables, and for every for every $\mathbb{K}$-database $\mathbf{I}$, we have that $E_i^{\mathbf{I}} = \varphi_i^{\mathbf{A}(\mathbf{I})}$, for $i = 1, 2$. By renaming variables if necessary, we may assume that φ_1 and φ_2 have no free variables in common, so they are of the form $\varphi_1(y_1, \dots, y_n)$ and $\varphi_2(y_{n+1}, \dots, y_{n+m})$, where $y_1, \dots, y_{n+m}$ are distinct variables. We take φ_E to be the $\mathcal{BRC}(\tau)$-formula $\varphi_1(y_1, \dots, y_n) \wedge \varphi_2(y_{n+1}, \dots, y_{n+m})$ with $n + m$ free variables.

To show the correctness of the formula φ_E, we have to show that if $\mathbf{I}$ is a $\mathbb{K}$-database, then $E^{\mathbf{I}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$. This holds because $E^{\mathbf{I}} = E_1^{\mathbf{I}} \times E_2^{\mathbf{I}} = \varphi_1^{\mathbf{A}(\mathbf{I})} \times \varphi_2^{\mathbf{A}(\mathbf{I})} = (\varphi_1 \wedge \varphi_2)^{\mathbf{A}(\mathbf{I})} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, where the first equation follows from the semantics of $\mathcal{BRA}(\tau)$-expressions in Definition 8, the second follows from the inductive hypothesis, and third follows from the third part of Proposition 2, and the fourth follows from the definition of φ_E. The domain independence of the formula φ_E follows by the inductive hypothesis and the second part of Proposition 4.

Case 4: Assume that E is a $\mathcal{BRA}(\tau)$-expression of the form $\pi_v(E_1)$, where E_1 is an n-ary $\mathcal{BRA}(\tau)$-expression and $v = (i_1, \dots, i_k)$ is a sequence of distinct indices from $\{1, \dots, n\}$. By induction hypothesis, there is a domain independent $\mathcal{BRC}(\tau)$-formula $\varphi_1(y_1, \dots, y_n)$ such that for every $\mathbb{K}$-database $\mathbf{I}$, we have $E_1^{\mathbf{I}} = \varphi_1^{\mathbf{A}(\mathbf{I})}$. In this case, we take φ_E to be the $\mathcal{BRC}(\tau)$-formula $\exists y_{j_1} \dots \exists y_{j_{n-k}} \varphi_1$, where $j_1, \dots, j_{n-k}$ are the indices in $\{1, \dots, n\}$ that do not appear in the sequence $v = (i_1, \dots, i_k)$. Note that the free variables of φ_E are $y_{i_1}, \dots, y_{i_k}$. For example, if E is of the form $\pi_{1,3}(E_1)$, where E_1 has arity 4, then φ_E is $\exists y_2 \exists y_4 \varphi_1$ and its free variables are y_1, y_3.

To show the correctness of the formula φ_E, we have to show that if $\mathbf{I}$ is a $\mathbb{K}$-database, then $E^{\mathbf{I}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$. This holds because $E^{\mathbf{I}} = \pi_v(E_1^{\mathbf{I}}) = \pi_v(\varphi_1^{\mathbf{A}(\mathbf{I})}) = (\exists y_{j_1} \dots \exists y_{j_{n-k}} \varphi_1)^{\mathbf{A}(\mathbf{I})} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, where the first equation follows from the semantics of $\mathcal{BRA}(\tau)$-expressions in Definition 8, the second follows from the inductive hypothesis, the third follows by iterating the fifth part of Proposition 2 to handle the string of quantifiers $\exists y_{j_1} \dots \exists y_{j_{n-k}}$, and the fourth follows from the definition of φ_E. The domain independence of φ_E follows from the inductive hypothesis and Proposition 4.

Case 5: Assume that E is a $\mathcal{BRA}(\tau)$-expression of the form $\sigma_\theta(E_1)$, where θ is a selection condition and E_1 is an n-ary $\mathcal{BRA}(\tau)$-expression. By induction hypothesis, there is a domain independent $\mathcal{BRC}(\tau)$-formula $\varphi_1(y_1, \dots, y_n)$ such that for every $\mathbb{K}$-database $\mathbf{I}$, we have $E_1^{\mathbf{I}} = \varphi_1^{\mathbf{A}(\mathbf{I})}$. We take φ_E to be the formula $\varphi_1 \wedge \nabla(\varphi_\theta)$, where $\varphi_1(y_1, \dots, y_n)$ is the formula given by the inductive hypothesis for E_1, and φ_θ is a $\mathcal{BRC}(\tau)$-formula expressing the selection condition θ (recall that $\mathcal{BRC}(\tau)$ can express negated equalities, as spelled out in Remark 3).

To show the correctness of the formula φ_E, we must show that if $\mathbf{I}$ is a $\mathbb{K}$-database, then $E^{\mathbf{I}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$. Let $(a_1, \dots, a_n) \in \mathrm{adom}(\mathbf{I})^n$. It is not difficult to see that $P_\theta(a_1, \dots, a_n) = (\nabla\varphi_\theta)^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n)$. The reason is that, if $p, q \in \{0, 1\}$, then $\mathrm{s}(p+q) = \begin{cases} 1 & \text{if } p = 1 \text{ or } q = 1 \\ 0 & \text{otherwise} \end{cases}$ and $\mathrm{s}(p \cdot q) = \begin{cases} 1 & \text{if } p = q = 1 \\ 0 & \text{otherwise} \end{cases}$, hence disjunction and conjunction in $(\nabla\varphi_\theta)$ are expressed by means of addition and product prefixed by s. Now, by induction hypothesis, we have that $E_1^{\mathbf{I}}(a_1, \dots, a_n) = \varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n)$, hence $\varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) \cdot \varphi_\theta^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) = E_1^{\mathbf{I}}(a_1, \dots, a_n) \cdot P_\theta(a_1, \dots, a_n) = \sigma_\theta(E_1)^{\mathbf{I}}(a_1, \dots, a_n)$. If $(a_1, \dots, a_n) \notin \mathrm{adom}(\mathbf{I})^n$, by induction hypothesis, $E_1^{\mathbf{I}}(a_1, \dots, a_n) = \varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) = 0$, and thus $\varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) \cdot \varphi_\theta^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) = 0 = E_1^{\mathbf{I}}(a_1, \dots, a_n) \cdot P_\theta(a_1, \dots, a_n) = \sigma_\theta(E_1)^{\mathbf{I}}(a_1, \dots, a_n)$.

Finally, we show that the formula φ_E is domain independent. Let $\mathbf{I} = (R_1, \dots, R_p)$ be a $\mathbb{K}$-database and let $\mathbf{B} = (B, R_1, \dots, R_p)$ be a $\mathbb{K}$-structure with $\mathrm{adom}(\mathbf{I}) \subseteq B \subseteq D$. First, observe that we have $\varphi_\theta^{\mathbf{B}}(a_1, \dots, a_n) = \varphi_\theta^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n)$. Furthermore, by induction hypothesis, we have that $\varphi_1^{\mathbf{B}}(a_1, \dots, a_n) = \varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n)$. Thus,

$$\varphi_1^{\mathbf{B}}(a_1, \dots, a_n) \cdot \varphi_\theta^{\mathbf{B}}(a_1, \dots, a_n) = \varphi_1^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n) \cdot \varphi_\theta^{\mathbf{A}(\mathbf{I})}(a_1, \dots, a_n),$$

which entails that $\varphi_E^{\mathbf{B}}(a_1,\ldots,a_n) = \varphi_E^{\mathbf{A}(\mathbf{I})}(a_1,\ldots,a_n)$, as desired. If $(a_1,\ldots,a_n)$ is not a tuple from $\mathrm{adom}(\mathbf{I})$, since by induction hypothesis $\varphi_1^{\mathbf{B}}(a_1,\ldots,a_n) = \varphi_1^{\mathbf{A}(\mathbf{I})}(a_1,\ldots,a_n)$ and $\varphi_1^{\mathbf{A}(\mathbf{I})}(a_1,\ldots,a_n) = 0$, we must have that $\varphi_E^{\mathbf{B}}(a_1,\ldots,a_n) = 0 = \varphi_E^{\mathbf{A}(\mathbf{I})}(a_1,\ldots,a_n)$, as desired.

Case 6: Assume that E is a $\mathcal{BRA}(\tau)$-expression of the form $\mathrm{supp}(E_1)$, where E_1 is an n-ary $\mathcal{BRA}(\tau)$-expression. By induction hypothesis, there is a domain independent $\mathcal{BRC}(\tau)$-formula $\varphi_1(y_1,\ldots,y_n)$ such that for every $\mathbb{K}$-database $\mathbf{I}$, we have $E_1^{\mathbf{I}} = \varphi_1^{\mathbf{A}(\mathbf{I})}$. We take φ_E to be the $\mathcal{BRC}(\tau)$-formula $\nabla\varphi_1$. To show the correctness of the formula φ_E, we have to show that if $\mathbf{I}$ is a $\mathbb{K}$-database, then $E^{\mathbf{I}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$. This holds because $E^{\mathbf{I}} = \mathrm{supp}(E_1^{\mathbf{I}}) = \mathrm{supp}(\varphi_1^{\mathbf{A}(\mathbf{I})}) = (\nabla\varphi_1)^{\mathbf{A}(\mathbf{I})} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, where the first equation follows from the semantics of $\mathcal{BRA}(\tau)$-expressions in Definition 8, the second follows from the inductive hypothesis, the third follows from the fourth part of Proposition 2, and the fourth follows from the definition of φ_E. The domain independence of the formula φ_E follows from the inductive hypothesis and the fourth part of Proposition 4.

(3) $\implies$ (1) : We proceed by induction on the size of the $\mathcal{BRC}(\tau)$-formula φ.

Case 1: Assume that $\varphi(y_1,y_2)$ is the $\mathcal{BRC}(\tau)$-formula $y_1 = y_2$. In this case, we take E_φ to be $\mathrm{supp}(\sigma_\theta(E_{\mathrm{adom}} \times E_{\mathrm{adom}}))$, where θ is $x_1 = x_2$, thus $P_\theta(a,b) = 1$ if $a = b$, and $P_\theta(a,b) = 0$ if $a \neq b$.

Observe that for a,b in $\mathrm{adom}(\mathbf{I})$, if $a = b$, we have that $E_{\mathrm{adom}}^{\mathbf{I}}(a) \neq 0$, $E_{\mathrm{adom}}^{\mathbf{I}}(b) \neq 0$, and thus $E_{\mathrm{adom}}^{\mathbf{I}}(a) \cdot E_{\mathrm{adom}}^{\mathbf{I}}(b) \neq 0$. Furthermore, since $P_\theta(a,b) = 1$, we have that

$$(\sigma_\theta(E_{\mathrm{adom}} \times E_{\mathrm{adom}}))^{\mathbf{I}}(a,b) = (E_{\mathrm{adom}} \times E_{\mathrm{adom}})^{\mathbf{I}}(a,b),$$

which means that $(\sigma_\theta(E_{\mathrm{adom}} \times E_{\mathrm{adom}}))^{\mathbf{I}}(a,b) \neq 0$, hence $E_\varphi^{\mathbf{I}}(a,b) = 1$. Similarly, if $a \neq b$, we have that $E_\varphi^{\mathbf{I}}(a,b) = 0$. Thus, if $a,b \in \mathrm{adom}(\mathbf{I})$, we have that $E_\varphi^{\mathbf{I}}(a,b) = (y_1 = y_2)^{\mathbf{A}(\mathbf{I})}(a,b)$. Now, if $a \notin \mathrm{adom}(\mathbf{I})$, then $(y_1 = y_2)^{\mathbf{A}(\mathbf{I})}(a,b) = 0$ by definition and $E_{\mathrm{adom}}^{\mathbf{I}}(a) = 0$, which entails that $E_\varphi^{\mathbf{I}}(a,b) = 0$. The case in which $b \notin \mathrm{adom}(\mathbf{I})$ is identical.

Case 2: Assume that $\varphi(y_1,\ldots,y_n)$ is a $\mathcal{BRC}(\tau)$-formula of the form $\hat{R}_i(y_1,\ldots,y_n)$, where $\hat{R}_i$ is one of the relation symbols of τ. In this case, we take E_φ to be simply the relation symbol $\hat{R}_i$. The correctness of E_φ follows from Definitions 8, 11, and 12.

Case 3: Assume that $\varphi(y_1,y_2,y_3,y_4)$ is a $\mathcal{BRC}(\tau)$-formula of the form is $\psi(y_1,y_2,y_4) \prec \chi(y_2,y_3)$. In this case, we take E_φ to be $\pi_{1,2,4,3}(E_\psi \times E_{\mathrm{adom}}) \setminus \pi_{1,3,4,2}(E_{\mathrm{adom}} \times E_{\mathrm{adom}} \times E_\chi)$.

For all elements $a_1,a_2,a_3,a_4 \in D$, we have that
$E_\varphi^{\mathbf{I}}(a_1,a_2,a_3,a_4) = (\pi_{1,2,4,3}(E_\psi \times E_{\mathrm{adom}}))^{\mathbf{I}}(a_1,a_2,a_3,a_4) - (\pi_{1,3,4,2}(E_{\mathrm{adom}} \times E_{\mathrm{adom}} \times E_\chi))^{\mathbf{I}}(a_1,a_2,a_3,a_4)$.

If $a_1,a_2,a_3,a_4 \in \mathrm{adom}(\mathbf{I})$, then

$(\pi_{1,2,4,3}(E_\psi \times E_{\mathrm{adom}}))^{\mathbf{I}}(a_1,a_2,a_3,a_4) = \sum_{\mathbf{w}[1,2,4,3]=a_1,a_2,a_3,a_4 \text{ and } (E_\psi \times E_{\mathrm{adom}})^{\mathbf{I}}(\mathbf{w}) \neq 0} (E_\psi \times E_{\mathrm{adom}})^{\mathbf{I}}(\mathbf{w})$.
Observe that when $\mathbf{w}[1,2,4,3] = a_1,a_2,a_3,a_4$ (so $\mathbf{w}$ must be (a_1,a_2,a_4,a_3)), we have that

$$(\pi_{1,2,4,3}(E_\psi \times E_{\mathrm{adom}}))^{\mathbf{I}}(a_1,a_2,a_3,a_4) = (E_\psi \times E_{\mathrm{adom}})^{\mathbf{I}}(\mathbf{w}) = E_\psi^{\mathbf{I}}(a_1,a_2,a_4) \cdot E_{\mathrm{adom}}^{\mathbf{I}}(a_3).$$

Thus, $E_\psi^{\mathbf{I}}(a_1,a_2,a_4) \cdot E_{\mathrm{adom}}(a_3) = E_\psi^{\mathbf{I}}(a_1,a_2,a_4) \cdot 1 = \psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)$, where the last equality follows from the induction hypothesis, so $(\pi_{1,2,4,3}(E_\psi \times E_{\mathrm{adom}}))^{\mathbf{I}}(a_1,a_2,a_3,a_4) = \psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)$. Similarly, $(\pi_{1,3,4,2}(E_{\mathrm{adom}} \times E_{\mathrm{adom}} \times E_\chi))^{\mathbf{I}}(a_1,a_2,a_3,a_4) = \chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3)$. Thus, when $a_1,a_2,a_3,a_4 \in \mathrm{adom}(\mathbf{I})$, we have that $E_\varphi^{\mathbf{I}}(a_1,a_2,a_3,a_4) = \psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4) - \chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3) = \varphi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_3,a_4)$, as desired. If at least one of a_1,a_2,a_3,a_4 is not in $\mathrm{adom}(\mathbf{I})$, then, by Definition 12, we have that $\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4) = 0$ or $\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3) = 0$; hence, by induction hypothesis, $E_\psi^{\mathbf{I}}(a_1,a_2,a_4) = 0$ or $0 = E_\chi^{\mathbf{I}}(a_2,a_3)$. Suppose then that $a_1 \notin \mathrm{adom}(\mathbf{I})$, but $a_2,a_3 \in \mathrm{adom}(\mathbf{I})$ and $\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3) \neq 0$ (the other cases are similar). Thus, $\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4) = 0 = E_\psi^{\mathbf{I}}(a_1,a_2,a_4)$, but $E_\chi^{\mathbf{I}}(a_2,a_3) = \chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3) \neq 0$. Now,

$$E_{\mathrm{adom}}^{\mathbf{I}}(a_1) \cdot E_{\mathrm{adom}}^{\mathbf{I}}(a_4) \cdot E_\chi^{\mathbf{I}}(a_2,a_3) = 0 \cdot E_{\mathrm{adom}}^{\mathbf{I}}(a_4) \cdot E_\chi^{\mathbf{I}}(a_2,a_3) = 0,$$

and $E^{\mathbf{I}}_{\psi}(a_1,a_2,a_4)\cdot E^{\mathbf{I}}_{\mathrm{adom}}(a_3)=0\cdot E^{\mathbf{I}}_{\mathrm{adom}}(a_3)=0$. Thus,

$$(\pi_{1,2,4,3}(E_{\psi}\times E_{\mathrm{adom}}))^{\mathbf{I}}(a_1,a_2,a_3,a_4)=0=(\pi_{1,3,4,2}(E_{\mathrm{adom}}\times E_{\mathrm{adom}}\times E_{\chi}))^{\mathbf{I}}(a_1,a_2,a_3,a_4).$$

Then, $E_{\varphi}(a_1,a_2,a_3,a_4)=0$ since $0-0=0$, and by Definition 12, $\varphi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_3,a_4)=0$. Thus, $E^{\mathbf{I}}_{\varphi}(a_1,a_2,a_3,a_4)=\varphi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_3,a_4)$, as desired.

Note that this inductive step is the first place where the formula φ may have no free variables. We deal with this case by producing an expression E_{φ} that defines a 0-ary query. We argue along the same lines we just did above, but this time all we have to show is that $\varphi^{\mathbf{A}(\mathbf{I})}(<>)=E^{\mathbf{I}}_{\varphi}(<>)$. The details are left to the reader. This same comment applies to the remaining inductive steps when the formula φ has no free variables.

Case 4: Assume that $\varphi(y_1,y_2,y_3,y_4)$ is a $\mathcal{BRC}(\tau)$-formula of the form $\psi(y_1,y_2,y_4)\wedge\chi(y_2,y_3)$. Without loss of generality, assume that the variable y_5 does not occur in the formulas ψ and χ. Let χ' be the formula obtained from χ by renaming every occurrence of the variable y_2 in χ y_5, thus the free variables of χ' are y_5 and y_3. We take E_{φ} to be $\pi_{1,2,5,3}(\sigma_{\theta}(E_{\psi}\times E_{\chi'}))$, where θ is the selection condition $y_2=y_5$. For all elements $a_1,a_2,a_3,a_4\in D$, we have that

$$E^{\mathbf{I}}_{\varphi}(a_1,a_2,a_3,a_4)=\sum_{\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4\text{ and }\sigma_{\theta}(E_{\psi}\times E_{\chi'})^{\mathbf{I}}(\mathbf{w})\neq 0}(\sigma_{\theta}(E_{\psi}\times E_{\chi'}))^{\mathbf{I}}(\mathbf{w}),$$

and that

$$\sum_{\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4\text{ and }\sigma_{\theta}(E_{\psi}\times E_{\chi'})^{\mathbf{I}}(\mathbf{w})\neq 0}(\sigma_{\theta}(E_{\psi}\times E_{\chi'}))^{\mathbf{I}}(\mathbf{w})$$

is identical to

$$\sum_{\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4\text{ and }\sigma_{\theta}(E_{\psi}\times E_{\chi'})^{\mathbf{I}}(\mathbf{w})\neq 0}(E^{\mathbf{I}}_{\psi}(\mathbf{w}[v_1])\cdot E^{\mathbf{I}}_{\chi'}(\mathbf{w}[v_2])\cdot P_{\theta}(\mathbf{w})),$$

where $v_1=(1,2,4)$ and $v_2=(5,3)$.

If $a_1,a_2,a_3,a_4\in\mathrm{adom}(\mathbf{I})$, then, by induction hypothesis, we have that

$$\sum_{\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4\text{ and }\sigma_{\theta}(E_{\psi}\times E_{\chi'})^{\mathbf{I}}(\mathbf{w})\neq 0}(E^{\mathbf{I}}_{\psi}(\mathbf{w}[v_1])\cdot E^{\mathbf{I}}_{\chi'}(\mathbf{w}[v_2])\cdot P_{\theta}(\mathbf{w}))$$

is identical to

$$\sum_{\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4\text{ and }\sigma_{\theta}(E_{\psi}\times E_{\chi'})^{\mathbf{I}}(\mathbf{w})\neq 0}(\psi^{\mathbf{A}(\mathbf{I})}(\mathbf{w}[v_1])\cdot \chi'^{\mathbf{A}(\mathbf{I})}(\mathbf{w}[v_2])\cdot P_{\theta}(\mathbf{w})),$$

which in turn is

$$\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)\cdot\chi'^{\mathbf{A}(\mathbf{I})}(a_2,a_3)\cdot 1=\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)\cdot\chi'^{\mathbf{A}(\mathbf{I})}(a_2,a_3),$$

since, when $\mathbf{w}[1,2,5,3]=a_1,a_2,a_3,a_4$, we have that $P_{\theta}(\mathbf{w})=1$ only when $\mathbf{w}$ is (a_1,a_2,a_3,a_4,a_2) and $P_{\theta}(\mathbf{w})=0$ for all other possible tuples. Thus, as desired,

$$E^{\mathbf{I}}_{\varphi}(a_1,a_2,a_3,a_4)=\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)\wedge\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3).$$

If at least one of a_1,a_2,a_3,a_4 is not in $\mathrm{adom}(\mathbf{I})$, then, by Definition 12, $\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)=0$ or $\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3)=0$; hence, by induction hypothesis, we have that $E^{\mathbf{I}}_{\psi}(a_1,a_2,a_4)=0$ or $E^{\mathbf{I}}_{\chi}(a_2,a_3)=0$. Suppose then that $a_1\notin\mathrm{adom}(\mathbf{I})$, but $a_2,a_3\in\mathrm{adom}(\mathbf{I})$ and $\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3)\neq 0$ (the other cases are similar). Thus, $\psi^{\mathbf{A}(\mathbf{I})}(a_1,a_2,a_4)=0=E^{\mathbf{I}}_{\psi}(a_1,a_2,a_4)$, but $E^{\mathbf{I}}_{\chi}(a_2,a_3)=\chi^{\mathbf{A}(\mathbf{I})}(a_2,a_3)\neq 0$. Then $E^{\mathbf{I}}_{\psi}(a_1,a_2,a_4)\cdot E^{\mathbf{I}}_{\chi'}(a_2,a_3)\cdot P_{\theta}(a_1,a_2,a_3,a_4,a_2)=0$. Thus, $(\pi_{1,2,5,3}(\sigma_{\theta}(E_{\psi}\times E_{\chi'})))^{\mathbf{I}}(a_1,a_2,a_3,a_4)=0$.

Then, $E^{\mathbf{I}}_{\varphi}(a_1, a_2, a_3, a_4) = 0$, and by Definition 12, we have that $\varphi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_3, a_4) = 0$. Thus, $E_{\varphi}(a_1, a_2, a_3, a_4) = \varphi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_3, a_4)$, as desired.

Case 5: Assume that $\varphi(y_1, y_2, y_3, y_4)$ is a $\mathcal{BRC}(\tau)$-formula of the form $\psi(y_1, y_2, y_4) \vee \chi(y_2, y_3)$. In this case, we take E_{φ} to be $\pi_{1,2,4,3}(E_{\psi} \times E_{\text{adom}}) \cup \pi_{1,3,4,2}(E_{\text{adom}} \times E_{\text{adom}} \times E_{\chi})$.

For all elements $a_1, a_2, a_3, a_4 \in D$, we have that

$$E^{\mathbf{I}}_{\varphi}(a_1, a_2, a_3, a_4) = (\pi_{1,2,4,3}(E_{\psi} \times E_{\text{adom}}))^{\mathbf{I}}(a_1, a_2, a_3, a_4) + (\pi_{1,3,4,2}(E_{\text{adom}} \times E_{\text{adom}} \times E_{\chi}))^{\mathbf{I}}(a_1, a_2, a_3, a_4).$$

If $a_1, a_2, a_3, a_4 \in \text{adom}(\mathbf{I})$, then

$$(\pi_{1,2,4,3}(E_{\psi} \times E_{\text{adom}}))^{\mathbf{I}}(a_1, a_2, a_3, a_4) = \sum_{\mathbf{w}[1,2,4,3]=(a_1,a_2,a_3,a_4) \text{ and } (E_{\psi} \times E_{\text{adom}})^{\mathbf{I}}(\mathbf{w}) \neq 0} (E_{\psi} \times E_{\text{adom}})^{\mathbf{I}}(\mathbf{w}).$$

There is only one $\mathbf{w}$ such that $\mathbf{w}[1, 2, 4, 3] = (a_1, a_2, a_3, a_4)$, namely $\mathbf{w} = (a_1, a_2, a_4, a_3)$). Thus,

$$(\pi_{1,2,4,3}(E_{\psi} \times E_{\text{adom}}))^{\mathbf{I}}(a_1, a_2, a_3, a_4) = (E_{\psi} \times E_{\text{adom}})(a_1, a_2, a_4, a_3) = E^{\mathbf{I}}_{\psi}(a_1, a_2, a_4) \cdot E^{\mathbf{I}}_{\text{adom}}(a_3)$$

and $E^{\mathbf{I}}_{\psi}(a_1, a_2, a_4) \cdot E^{\mathbf{I}}_{\text{adom}}(a_3) = E_{\psi}(a_1, a_2, a_4) \cdot 1 = \psi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_4)$, where the last equality follows by induction hypothesis, so $(\pi_{1,2,4,3}(E_{\psi} \times E_{\text{adom}}))^{\mathbf{I}}(a_1, a_2, a_3, a_4) = \psi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_4)$. Similarly,

$$(\pi_{1,3,4,2}(E_{\text{adom}} \times E_{\text{adom}} \times E_{\chi}))^{\mathbf{I}}(a_1, a_2, a_3, a_4) = \chi^{\mathbf{A}(\mathbf{I})}(a_2, a_3).$$

Thus, when $a_1, a_2, a_3, a_4 \in \text{adom}(\mathbf{I})$, we have

$$E^{\mathbf{I}}_{\varphi}(a_1, a_2, a_3, a_4) = \psi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_4) + \chi^{\mathbf{A}(\mathbf{I})}(a_2, a_3) = \varphi^{\mathbf{A}(\mathbf{I})}(a_1, a_2, a_3, a_4).$$

The argument for the case in which at least one of a_1, a_2, a_3, a_4 is not in $\text{adom}(\mathbf{I})$ is similar to the argument for the analogous case in the previous two inductive steps.

Case 6: Assume that φ is a $\mathcal{BRC}(\tau)$-formula of the form $\nabla\psi$. In this case, we take E_{φ} to be $\text{supp}(E_{\psi})$. By the fourth part of Proposition 2, we have that $(\nabla\psi)^{\mathbf{A}(\mathbf{I})} = \text{supp}(\psi^{\mathbf{A}(\mathbf{I})})$. By induction hypothesis, we have that $\psi^{\mathbf{A}(\mathbf{I})} = E^{\mathbf{I}}_{\psi}$, hence $(\nabla\psi)^{\mathbf{A}(\mathbf{I})} = \text{supp}(E^{\mathbf{I}}_{\psi}) = E^{\mathbf{I}}_{\varphi}$, as desired.

Case 7: Assume that φ is a $\mathcal{BRC}(\tau)$-formula of the form $\exists y_i \psi$, where ψ is a $\mathcal{BRC}(\tau)$-formula with $y_1, \ldots, y_n$ as its free variables and i is an index such that $1 \leq i \leq n$. In this case, we take E_{φ} to be $\pi_v(E_{\psi})$, where $v = (i_1, \ldots, i_{n-1})$ is the sequence of distinct indices from $\{1, \ldots, n\} \setminus \{i\}$ ordered by the standard ordering of the natural numbers.

By the fifth part of Proposition 2, we have that $(\exists y_i \psi)^{\mathbf{A}(\mathbf{I})} = \pi_v(\psi^{\mathbf{A}(\mathbf{I})})$. By induction hypothesis, we have that $\psi^{\mathbf{A}(\mathbf{I})} = E^{\mathbf{I}}_{\psi}$, hence $(\exists y_i \psi)^{\mathbf{A}(\mathbf{I})} = \pi_v(E^{\mathbf{I}}_{\psi}) = (\pi_v(E_{\psi}))^{\mathbf{I}} = E^{\mathbf{I}}_{\varphi}$, as desired. □

4 Codd's Theorem for Relational Algebra with Division

4.1 Relational algebra with division

We now introduce an additional relational algebra operation that is definable from the basic ones in the Boolean case but not in the general semiring framework, as we will soon see.

Definition 17 (Division). *Let $R_1 \colon D^{n_1} \longrightarrow K$ and $R_2 \colon D^{n_2} \longrightarrow K$ be two $\mathbb{K}$-relations with $n_1 > n_2$. The* division *of R_1 and R_2 is the $\mathbb{K}$-relation $R_1 \div R_2 \colon D^{n_1 - n_2} \longrightarrow K$ defined by*

$$R_1 \div R_2(\mathbf{a}) = \mathrm{s}\Big(\sum_{\mathbf{b} \in D^{n_2}} R_1(\mathbf{a}, \mathbf{b})\Big) \cdot \prod_{\mathbf{b} \,:\, R_2(\mathbf{b}) \neq 0} R_1(\mathbf{a}, \mathbf{b}).$$

We stipulate that the empty product is equal to 1. *Observe that $R_1 \div R_2 = R_1 \div supp(R_2)$.*

The quotient $R_1 \div R_2$ is well defined. Indeed, for every $\mathbf{a} \in D^{n_1 - n_2}$, the product $\prod_{R_2(\mathbf{b}) \neq 0} R_1(\mathbf{a}, \mathbf{b})$ is finite because R_2 is a $\mathbb{K}$-relation, thus there are finitely many tuples $\mathbf{b}$ such that $R_2(\mathbf{b}) \neq 0$. Furthermore, $R_1 \div R_2$ has finite support because so does R_1, hence we have that $\mathrm{s}(\sum_{\mathbf{b} \in D^{n_2}} R_1(\mathbf{a}, \mathbf{b})) \neq 0$ for only finitely many tuples $\mathbf{a} \in D^{n_1 - n_2}$. Thus, $R_1 \div R_2$ is indeed a $\mathbb{K}$-relation. Note that if R_2 is the

n_2-ary $\mathbb{K}$-relation with empty support, then $\prod_{R_2(\mathbf{b})\neq 0} R_1(\mathbf{a},\mathbf{b}) = 1$, for every tuple $\mathbf{a} \in D^{n_1-n_2}$, since this product is empty. This is why the expression $\mathrm{s}(\sum_{\mathbf{b}\in D^{n_2}} R_1(\mathbf{a},\mathbf{b}))$ is needed in the definition of $R_1 \div R_2$; without it we would not have a $\mathbb{K}$-relation, i.e., a function with finite support.

We extend the Basic Relational Algebra $\mathcal{BRA}$ to obtain the *Relational Algebra* $\mathcal{RA}$ by adding the following clause to Definition 7:

- If E_1, E_2 are $\mathcal{RA}(\tau)$-expressions of arities n_1, n_2, respectively, such that $n_1 > n_2$, then $E_1 \div E_2$ is an $\mathcal{RA}(\tau)$-expression of arity $n_1 - n_2$.

We also extend the semantics by adding the following clause to Definition 8:

- $(E_1 \div E_2)^{\mathbf{I}} = E_1^{\mathbf{I}} \div E_2^{\mathbf{I}}$.

Example 7. *If $\mathbb{B} = (\{0,1\}, \wedge, \vee, 0, 1)$ is the Boolean semiring, then the division operation introduced here coincides with the division (also known as* quotient*) operation of relational algebra, i.e., if R_1 and R_2 are ordinary relations with arity(R_2) < arity(R_1), then $R_1 \div R_2(\mathbf{a}) = 1$ if and only if for all $\mathbf{b}$ with $R_2(\mathbf{b}) = 1$, we have that $R_1(\mathbf{a},\mathbf{b}) = 1$.*

Example 8. *We illustrate the meaning of the division operation for other semirings of interest.*

Let $\mathbb{K}$ be a semiring with monus and support. Assume that R_1 and R_2 are two $\mathbb{K}$-relations with arity(R_2) < arity(R_1), and let $\mathbf{a}$ be a tuple in $D^{n_1-n_2}$.

1. If $\mathbf{a} \notin \mathrm{supp}(\pi_{1\ldots n_1-n_2}(R_1))$, then $R_1 \div R_2(\mathbf{a}) = 0_{\mathbb{K}}$.

2. If $\mathbf{a} \in \mathrm{supp}(\pi_{1\ldots n_1-n_2}(R_1))$ and there is a tuple $\mathbf{b} \in \mathrm{supp}(R_2)$ such that $(\mathbf{a},\mathbf{b}) \notin \mathrm{supp}(R_1)$, then $R_1 \div R_2(\mathbf{a}) = 0_{\mathbb{K}}$.

3. If $\mathbf{a} \in \mathrm{supp}(\pi_{1\ldots n_1-n_2}(R_1))$ and for every $\mathbf{b} \in \mathrm{supp}(R_2)$ we have that $(\mathbf{a},\mathbf{b}) \in \mathrm{supp}(R_1)$, then:

– If $\mathbb{K}$ is the bag semiring $\mathbb{N} = (N, +, \cdot, 0, 1)$, then $R_1 \div R_2(\mathbf{a}) = \prod_{R_2(\mathbf{b})\neq 0} R_1(\mathbf{a},\mathbf{b})$.

– If $\mathbb{K}$ is the tropical semiring $\mathbb{T} = (\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$, then $R_1 \div R_2(\mathbf{a}) = \sum_{R_2(\mathbf{b})\neq 0} R_1(\mathbf{a},\mathbf{b})$.

– If $\mathbb{K}$ is the fuzzy semiring $\mathbb{F} = ([0,1], \max, \min, 0, 1)$, then $R_1 \div R_2(\mathbf{a}) = \min_{R_2(\mathbf{b})\neq 0} R_1(\mathbf{a},\mathbf{b})$.

In the case of the Boolean semiring $\mathbb{B}$, it is well known that the division operation is expressible in terms of the other relational algebra operations; in fact, it is expressible using projection, Cartesian product, and difference (e.g., see [27, pages 153-154]). In what follows, we show that the state of affairs is quite different for the bag semiring $\mathbb{N}$ and the tropical semiring $\mathbb{T}$.

Theorem 2. *Let $\mathbb{K}$ be either the bag semiring $\mathbb{N} = (N, +, \cdot, \dot{-}, \mathrm{s}, 0, 1)$ expanded with the truncated subtraction $\dot{-}$ as monus and with support* s*, or the tropical semiring $\mathbb{T} = (\mathbb{R} \cup \{\infty\}, \min, +, -, \mathrm{s}, \infty, 0)$ expanded with its corresponding monus $-$ and with support* s*. Then there is no expression of basic relational algebra $\mathcal{BRA}$ that defines the division operation $\div$ on $\mathbb{K}$-databases*

Proof. Let $\tau = (R, S)$ be a schema consisting of a binary relation symbol R and a unary relation symbol S. We start with the case where $\mathbb{K}$ is $\mathbb{N}$ expanded with monus and support. For every $n \geq 1$, let $a, b_1, \ldots, b_n$ be distinct elements from the domain D and let $\mathbf{I}(n) = (R_n, S_n)$ be the bag database ($\mathbb{N}$-database), where R_n and S_n are the following bags ($\mathbb{N}$-relations):

- $R_n(a, b_i) = 2$, for $1 \leq i \leq n$, while $R_n(c,d) = 0$, for all other pairs of elements of D.
- $S_n(b_i) = 1$, for $1 \leq i \leq n$, while $S_n(c) = 0$, for every other element of D.

For every $n \geq 1$, the division $R_n \div S_n$ is the bag such that $(R_n \div S_n)(a) = 2^n$, while $(R_n \div S_n)(c) = 0$, for all other elements of D.

We will show that no $\mathcal{BRC}$-expression defines $R \div S$ by showing that, for sufficiently large n, no such expression involving R_n and S_n can have an element of multiplicity 2^n.

Let E be an expression of the basic relational calculus $\mathcal{BRC}$.

- The *length* $l(E)$ of E is defined by induction as follows: if E is R or S, then $l(E) = 1$; if E is E_1 op E_2, where op is union, difference, or Cartesian product, then $l(E) = l(E_1) + l(E_2) + 1$; if E is op(F), where op is projection, selection, or support, then $l(E) = l(F) + 1$.

- We write $|\mathrm{supp}(E^{\mathbf{I}(n)})|$ for the cardinality of the support of the bag $E^{\mathbf{I}(n)}$.
- We write $\mathrm{hm}(E^{\mathbf{I}(n)})$ to denote the highest multiplicity of a tuple in the bag $E^{\mathbf{I}(n)}$.

Claim: The following statements are true:

(1) $|\mathrm{supp}(E^{\mathbf{I}(n)})| \leq n^{l(E)}$, for all $n \geq 2$.
(2) There is a polynomial $p_E(n)$ such that $\mathrm{hm}(E^{\mathbf{I}(n)}) \leq p_E(n)2^{l(E)}$, for all $n \geq 2$.

Both statements are proved by induction on the construction of the $\mathcal{BRC}$-expression E; the proof of the second statement uses also the first statement.

For the first statement, if E is R or E is S, then $|\mathrm{supp}(E^{\mathbf{I}(n)})| = n = n^{l(E)}$, since in this case $l(E) = 1$. For the induction steps, the more interesting cases are those of union and Cartesian product. If $E = E_1 \cup E_2$, then $|\mathrm{supp}(E^{\mathbf{I}(n)})| \leq |\mathrm{supp}(E_1^{\mathbf{I}(n)})| + |\mathrm{supp}(E_2^{\mathbf{I}(n)})| \leq n^{l(E_1)} + n^{l(E_2)} \leq 2n^{l(E_1)+l(E_2)} \leq n^{l(E_1)+l(E_2)+1} = n^{l(E)}$, where the second inequality follows from the induction hypothesis, while the one before the last one follows from the assumption that $n \geq 2$. If $E = E_1 \times E_2$, then

$$|\mathrm{supp}(E^{\mathbf{I}(n)})| = |\mathrm{supp}(E_1^{\mathbf{I}(n)})| \cdot |\mathrm{supp}(E_2^{\mathbf{I}(n)})| \leq n^{l(E_1)} \cdot n^{l(E_2)} = n^{l(E_1)+l(E_2)} \leq n^{l(E_1)+l(E_2)+1} = n^{l(E)},$$

where the first inequality follows from the induction hypothesis.

For the second statement, if E is R or E is S, then $\mathrm{hm}(E^{\mathbf{I}(n)}) \leq 2$, so we can put $p_E(n) = 1$. For the induction steps, the more interesting cases are those of Cartesian product and projection. If $E = E_1 \times E_2$, then

$$\mathrm{hm}(E^{\mathbf{I}(n)}) = \mathrm{hm}(E_1^{\mathbf{I}(n)})\mathrm{hm}(E_2^{\mathbf{I}(n)}) \leq p_{E_1}(n)2^{l(E_1)}p_{E_2}(n)2^{l(E_2)} =$$

$$p_{E_1}(n)p_{E_2}(n)2^{l(E_1)+l(E_2)} \leq p_{E_1}(n)p_{E_2}(n)2^{l(E_1)+l(E_2)+1} = p_{E_1}(n)p_{E_2}(n)2^{l(E)},$$

where the first inequality follows from the induction hypothesis. So, in this case, we can put $p_E(n) = p_{E_1}(n)p_{E_2}(n)$. If E is $\pi_v(F)$, where π is the projection operation and v is a (possibly empty) sequence of indices, then

$$\mathrm{hm}(E^{\mathbf{I}(n)}) = \mathrm{hm}((\pi_v(F))^{\mathbf{I}(n)}) \leq |\mathrm{supp}(F^{\mathbf{I}(n)})|\mathrm{hm}(F^{\mathbf{I}(n)}) \leq n^{l(F)}p_F(n)2^{l(F)} \leq$$

$$n^{l(F)}p_F(n)2^{l(F)+1} = n^{l(F)}p_F(n)2^{l(E)},$$

where the first inequality follows from the first part of the claim and the induction hypothesis. Thus, in this case, we can put $p_E(n) = n^{l(F)}p_F(n)$.

The second part of the preceding claim implies easily the theorem we are after. Indeed, if the division $R \div S$ could be expressed by some $\mathcal{BRC}$-expression E, then, by the second statement in the above claim, there is a polynomial $p_E(n)$ such that for every $n \geq 2$, it is the case that $\mathrm{hm}(E^{\mathbf{I}(n)}) \leq p_E(n)2^{l(E)}$. However, we also have that $\mathrm{hm}(E^{\mathbf{I}(n)}) = 2^n$, which, for sufficiently large n, is greater than $p_E(n)2^{l(E)}$, since $p_E(n)$ is some fixed polynomial and $2^{l(E)}$ is a constant.

Finally, for the case where $\mathbb{K}$ is $\mathbb{T}$ expanded with monus and support, we reason similarly. This time $\mathbf{I}(n) = (R_n, S_n)$ is the $\mathbb{T}(R)$-database, where R_n and S_n are the following $\mathbb{T}$-relations:

- $R_n(a, b_i) = i$, for $1 \leq i \leq n$, while $R_n(c, d) = 0$, for all other pairs of elements of D.
- $S_n(b_i) = 1$, for $1 \leq i \leq n$, while $S_n(c) = 0$, for every other element of D.

Claim: $h(E^{\mathbf{I}(n)}) \leq l(E) \cdot n$ holds, for all $n \geq 1$, where $h(E^{\mathbf{I}(n)})$ is the biggest (in the standard order) number in $\mathbb{R}$ that is an annotation of a tuple in the support of $\mathbf{I}(n)$ (in particular $h(E^{\mathbf{I}(n)}) \neq \infty$).

The proof of the claim is by induction on the length $l(E)$ of E. For the basis of the induction we have two cases: (1) $E = R$ or (2) $E = S$. For (1), we have that $h(E^{\mathbf{I}(n)}) = n = l(E) \cdot n$ since $l(E) = 1$. For (2), we have that $h(E^{\mathbf{I}(n)}) = 1 \leq l(E) \cdot n$ since $l(E) = 1$. Next we focus on the inductive steps.

Assume that $E = E_1 \cup E_2$. For every tuple $\overline{a} \in \mathrm{supp}(E^{\mathbf{I}(n)})$, we have that

$$E^{\mathbf{I}(n)}(\overline{a}) = \min\{E_1^{\mathbf{I}(n)}(\overline{a}), E_2^{\mathbf{I}(n)}(\overline{a})\}.$$

Hence, $h(E^{\mathbf{I}(n)}) \leq h(E_1^{\mathbf{I}(n)}) \leq l(E_1) \cdot n \leq (l(E_1) + l(E_2) + 1) \cdot n = l(E) \cdot n$ where the second inequality follows by inductive hypothesis.

Assume that $E = E_1 \setminus E_2$. For every tuple $\overline{a}$ in $\mathrm{supp}(E^{\mathbf{I}(n)})$, we have that $E^{\mathbf{I}(n)}(\overline{a}) = E_1^{\mathbf{I}(n)}(\overline{a}) - E_2^{\mathbf{I}(n)}(\overline{a}) = E_1^{\mathbf{I}(n)}(\overline{a})$ (otherwise we would have that $E^{\mathbf{I}(n)}(\overline{a}) = \infty$, hence $\overline{a} \notin \mathrm{supp}(E^{\mathbf{I}(n)})$). Thus, $h(E^{\mathbf{I}(n)}) \leq h(E_1^{\mathbf{I}(n)}) \leq l(E_1) \cdot n \leq (l(E_1) + l(E_2) + 1) \cdot n = l(E) \cdot n$ where the second inequality follows by inductive hypothesis.

Assume that $E = E_1 \times E_2$. Let $(\overline{a}, \overline{b})$ be in $\mathrm{supp}(E^{\mathbf{I}(n)})$. Then $\overline{a} \in \mathrm{supp}(E_1^{\mathbf{I}(n)})$ and $\overline{b} \in \mathrm{supp}(E_2^{\mathbf{I}(n)})$ and $E^{\mathbf{I}(n)}(\overline{a}, \overline{b}) = E_1^{\mathbf{I}(n)}(\overline{a}) + E_2^{\mathbf{I}(n)}(\overline{a})$. Thus, $h(E^{\mathbf{I}(n)}) \leq h(E_1^{\mathbf{I}(n)}) + h(E_2^{\mathbf{I}(n)}) \leq l(E_1) \cdot n + l(E_2) \cdot n \leq (l(E_1) + l(E_2) + 1) \cdot n = l(E) \cdot n$, where the second inequality follows by inductive hypothesis.

Assume that E is $\pi_v(E_1)$ or $\sigma_\theta(E_1)$. If $\overline{a} \in \mathrm{supp}(E^{\mathbf{I}(n)})$, then clearly $h(E^{\mathbf{I}(n)}) \leq h(E_1^{\mathbf{I}(n)}) \leq l(E_1) \cdot n \leq (l(E_1) + 1) \cdot n = l(E) \cdot n$, where the second inequality follows by inductive hypothesis.

Using the claim, we can now prove that division is not expressible in $\mathcal{BRA}$ over the tropical semiring $\mathbb{T}(R)$. Consider the $\mathbb{T}$-databases $\mathbf{I}(n)$ for $n \geq 1$ as above. Take the division $(R \div S)^{\mathbf{I}(n)}$. From the definition of $\mathbf{I}(n)$, $n \geq 1$, we have that $(R \div S)^{\mathbf{I}(n)}$ is the $\mathbb{T}(R)$-relation such that $(R \div S)^{\mathbf{I}(n)}(a) = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ and $(R \div S)^{\mathbf{I}(n)}(c) = \infty$ for all other values c. Suppose that there were a $\mathcal{BRA}$-expression E such that $(R \div S)^{\mathbf{I}(n)} = E^{\mathbf{I}(n)}$. Then $h(E^{\mathbf{I}(n)}) \leq l(E) \cdot n$ for all $n \geq 1$. in particular, we must have that $E^{\mathbf{I}(n)}(a) \leq l(E) \cdot n$. However, $E^{\mathbf{I}(n)}(a) = (R \div S)^{\mathbf{I}(n)}(a) = \frac{n(n+1)}{2}$. This gives rise to a contradiction since if n is large enough, then $\frac{n(n+1)}{2} > l(E) \cdot n$. ◻

4.2 Relational calculus with universal quantifier

For the remainder of this section, we assume that $\mathbb{K}$ is a positive semiring with monus and support (equivalently, $\mathbb{K}$ is a semiring with monus and support, and no zero divisors). We extend the basic relational calculus $\mathcal{BRC}$ with a universal quantifier $\forall$ and call the resulting system *relational calculus* $\mathcal{RC}$. For this, we add the following clause to the syntax in Definition 9:

- If φ is an $\mathcal{RC}(\tau)$-formula with $y_1, \ldots, y_n$ as free variables, then the expression $\forall y_i \varphi$, where $1 \leq i \leq n$, is an $\mathcal{RC}(\tau)$-formula with the variables $y_1, \ldots, y_n$ other than y_i as free variables.

Furthermore, we add the following clause to the semantics in Definition 11:

- $\|\forall y_i \varphi\|(\mathbf{A}, \alpha) = \prod_{b \in A} \|\varphi\|(\mathbf{A}, \alpha[y_i/b])$.

As in the case of the existential quantifier, the semantics of the universal quantifier is well defined because of our blanket assumption that the structure $\mathbf{A}$ has a finite universe.

Connections to non-classical logics. The syntax and the semantics of relational calculus $\mathcal{RC}$ that we defined here have also been considered in the context of algebraic semantics of non-classical logics. Specifically, there are connections to dual intuitionistic logic that we now discuss.

First-order intuitionistic logic has $\wedge, \vee, \rightarrow, 0$ as primitive connectives and $\forall$ and $\exists$ as quantifiers. The connective $\rightarrow$ is *intuitionistic implication*; it is used to define the negation of a formula as $\neg\varphi := \varphi \rightarrow 0$. Algebraic semantics is one of several different approaches to defining rigorous semantics of intuitionistic logic. The main idea, which originated with Heyting [18], is to consider the class $\mathcal{H}$ of all *Heyting algebras*, where a Heyting algebra is a complete bounded distributive lattices expanded with a binary operation $\Rightarrow$ that has the following property: $(a \wedge b) \preceq c$ if and only if $a \preceq (b \Rightarrow c)$. If $\mathbb{H}$ is a Heyting algebra and φ is a formula of first-order intuitionistic logic, then the semantics $\|\varphi\|$ of φ on $\mathbb{H}$ is defined using $\Rightarrow$ to interpret the intuitionistic implication $\rightarrow$. A sentence ψ of intuitionistic logic is *valid* if for every Heyting algebra $\mathbb{K}$ and every (finite or infinite) $\mathbb{K}$-structure $\mathbf{A}$, we have that $\psi^{\mathbf{A}} = 1$. Soundness and completeness theorems are then established about proof systems that are capable to derive precisely the valid sentences of intuitionistic logic.

Dual first-order intuitionistic logic has the same syntax as relational calculus $\mathcal{RC}$, except that the support ∇ is not included as a connective. Informally, dual intuitionistic logic is obtained from intuitionistic logic by replacing the $\Rightarrow$ connective by the $\prec$ connective. Goodman [12] gave algebraic semantics to dual first-order intuitionistic logic by considering the class $\mathcal{B}$ of all Brouwerian algebras discussed in Example 3. Goodman's semantics $\|\varphi\|$ of a formula φ of dual first-order intuitionistic logic on a Brouwerian algebra $\mathbb{K}$ is precisely the semantics of φ as an $\mathcal{RC}$-formula (without ∇) on $\mathbb{K}$ that we gave here. In particular, the connective $\prec$ is interpreted using the monus of $\mathbb{K}$.

4.3 Codd's Theorem for Relational Algebra and Relational Calculus

We establish a version of Codd's Theorem for relational algebra $\mathcal{RA}$ and relational calculus $\mathcal{RC}$.

Theorem 3. *Let $\mathbb{K}$ be a positive semiring with monus and support, let $\tau = (\hat{R}_1, \dots \hat{R}_p)$ be a schema, and let q be an n-ary query. The following statements are equivalent:*

1. *There is an $\mathcal{RA}(\tau)$-expression E of arity n such that $q^{\mathbf{I}} = E^{\mathbf{I}}$, for every $\mathbb{K}$-database $\mathbf{I}$.*
2. *There is a domain independent $\mathcal{RC}(\tau)$-formula $\varphi(y_1, \dots, y_n)$ such that $q^{\mathbf{I}} = \varphi^{\mathbf{A(I)}}$, for every $\mathbb{K}$-database $\mathbf{I}$.*
3. *There is an $\mathcal{RC}(\tau)$-formula $\varphi(y_1, \dots, y_n)$ such that $q^{\mathbf{I}} = \varphi^{\mathbf{A(I)}}$, for every $\mathbb{K}$-database $\mathbf{I}$.*

Proof. The proof expands that of Theorem 1. The direction (2) $\implies$ (3) is obvious, so we only prove the other two directions. We focus on the case of the division operation in direction (1) $\implies$ (2), and on the case of universal quantification in direction (3) $\implies$ (1). All other cases are identical to the ones in the proof of Theorem 1.

(1) $\implies$ (2) : Let E be an $\mathcal{RA}(\tau)$-expression of the form $E_1 \div E_2$, where E_1 and E_2 are $\mathcal{RA}(\tau)$-expressions of arities n_1 and n_2 with $n_1 > n_2$. By induction hypothesis and by renaming variables if necessary, there are domain independent $\mathcal{RC}(\tau)$-formulas $\varphi_1(y_1, \dots, y_{n_1})$ and $\varphi_2(y_{n_1-n_2+1}, \dots, y_{n_1})$, such that for every $\mathbb{K}$-database $\mathbf{I}$, we have that $E_1^{\mathbf{I}} = \varphi_1^{\mathbf{A(I)}}$ and $E_2^{\mathbf{I}} = \varphi_2^{\mathbf{A(I)}}$. We will construct a domain independent $\mathcal{RC}(\tau)$-formula φ_E such that $E^{\mathbf{I}} = \varphi_E^{\mathbf{A(I)}}$, for every $\mathbb{K}$-database $\mathbf{I}$.

Let η be the $\mathcal{RC}(\tau)$-sentence $\nabla\exists y_1(y_1 = y_1)$. It is easy to see that η has the following property:

- For every $\mathbb{K}$-database $\mathbf{I} = (R_1, \dots, R_p)$ and for every $\mathbb{K}$-structure $\mathbf{B} = (B, R_1, \dots, R_p)$ with the same relations as $\mathbf{I}$ and with $\mathrm{adom}(\mathbf{I}) \subseteq B \subseteq D$, we have that $\eta^{\mathbf{B}}(<>) = 1$.

This property holds because $\eta^{\mathbf{B}}(<>) = \|\nabla\exists y_1(y_1 = y_1)\|(\mathbf{B}, <>) = \mathrm{s}(\sum_{b \in B}(\|y_1 = y_1\|(\mathbf{B}, b))) = 1$, where the first two equalities follow from Definitions 11 and 12, while the third equality holds for the following reason: from the blanket assumption that $\mathbf{I}$ is non-trivial, we have that $\mathrm{adom}(\mathbf{I}) \neq \emptyset$, hence $B \neq \emptyset$, and so $\sum_{b \in B}(\|y_1 = y_1\|(\mathbf{B}, b)) = \sum_{b \in B} 1 \neq 0$, since $\mathbb{K}$ is a zero-sum free semiring. [1]

With the $\mathcal{RC}(\tau)$-sentence η at hand, we take φ_E to be the $\mathcal{RC}(\tau)$-formula

$$(\nabla\exists y_{n_1-n_2+1} \dots \exists y_{n_1}\varphi_1) \land (\forall y_{n_1-n_2+1} \dots \forall y_{n_1}((\eta \prec \nabla\varphi_2) \lor (\nabla\varphi_2 \land \varphi_1))),$$

whose free variables are $y_1, \dots, y_{n_1-n_2}$. We will show that φ_E is domain independent and that $E^{\mathbf{I}} = \varphi_E^{\mathbf{A(I)}}$ holds for every $\mathbb{K}$-database $\mathbf{I}$. We begin with the latter task.

We must show that $E^{\mathbf{I}}(a_1, \dots, a_{n_1-n_2}) = \varphi_E^{\mathbf{A(I)}}(a_1, \dots, a_{n_1-n_2})$ holds for all $a_1, \dots, a_{n_1-n_2} \in D$. Assume first that $a_1, \dots, a_{n_1-n_2} \in \mathrm{adom}(\mathbf{I})$. Put $\mathbf{a} = (a_1, \dots, a_{n_1-n_2})$ and $\mathbf{y} = (y_{n_1-n_2+1} \dots y_{n_1})$. Then

$$\varphi_E^{\mathbf{A(I)}}(\mathbf{a}) = \|(\nabla\exists\mathbf{y}\varphi_1) \land (\forall\mathbf{y}((\eta \prec \nabla\varphi_2) \lor (\nabla\varphi_2 \land \varphi_1)))\|(\mathbf{A(I)}, \mathbf{a}) =$$

$$\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A(I)}, \mathbf{a}) \cdot \prod_{\mathbf{b} \in \mathrm{adom}(\mathbf{I})^{n_2}} \|(\eta \prec \nabla\varphi_2) \lor (\nabla\varphi_2 \land \varphi_1)\|(\mathbf{A(I)}, \mathbf{a}, \mathbf{b}),$$

[1] Clearly, the $\mathcal{RC}$-sentence η uses equality =. There are equality-free $\mathcal{RC}$-sentences that have the properties of η. For example, it is easy to verify that $\nabla(\bigvee_{i=1}^{p} \exists y_1 \dots \exists y_{r_i} \hat{R}_i(y_1, \dots, y_{r_i}))$ is such a sentence.

where the first equality follows from the definition of the formula φ_E and the second equality follows from Definitions 11 and 12 (and their extensions for formulas with universal quantification). We now distinguish two cases, namely, whether or not there is a tuple $\mathbf{b}$ such that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) \neq 0$.

If for all $\mathbf{b}$, we have that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) = 0$, then $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) = 0$, hence $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = 0$. By induction hypothesis about E_1 and φ_1, we have that $E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b}) = 0$, for all $\mathbf{b}$. Hence, $\mathrm{s}(\sum_{\mathbf{b}\in D^{n_2}} E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b})) = 0$, hence, by Definition 17, $E_1^{\mathbf{I}} \div E_2^{\mathbf{I}}(\mathbf{a}) = 0$ and so $E^{\mathbf{I}}(\mathbf{a}) = 0$. Thus, in this case $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = E^{\mathbf{I}}(\mathbf{a})$.

If there is a tuple $\mathbf{b}$ such that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) \neq 0$, then $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) = 1$. Therefore, we have that

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \prod_{\mathbf{b}\in \mathrm{adom}(\mathbf{I})^{n_2}} (\|\eta \prec \nabla\varphi_2\|(\mathbf{A}(\mathbf{I}),\mathbf{b}) + \|(\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b})) =$$

$$\prod_{\mathbf{b}\in \mathrm{adom}(\mathbf{I})^{n_2}} (\eta^{\mathbf{A}(\mathbf{I})}(<>) - (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) + (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) \cdot \varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b})) =$$

$$\prod_{\mathbf{b}\in \mathrm{adom}(\mathbf{I})^{n_2}} (1 - (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) + (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) \cdot \varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b})),$$

where the second equality follows from Definitions 11 and 12, and the third equality follows from the properties of the sentence η. Fix a tuple $\mathbf{b} \in \mathrm{adom}(\mathbf{I})^{n_2}$ and consider the expression

$$1 - (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) + (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) \cdot \varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}).$$

If $\varphi_2^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 0$, then $(\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 0$, hence the above expression evaluates to 1, since $1 - 0 = 1$ and $0 \cdot x = 0$, for every x in the universe of the semiring $\mathbb{K}$. If $\varphi_2^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) \neq 0$, then $(\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 1$, hence the above expression evaluates to $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b})$, since $1 - 1 = 0$. It follows that

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \prod_{\mathbf{b}\,:\, \varphi_2^{\mathbf{A}(\mathbf{I})}(\mathbf{b})\neq 0} \varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}).$$

By induction hypothesis, we have that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) = E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b})$ and $\varphi_2^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = E_2^{\mathbf{I}}(\mathbf{b})$. Furthermore, $\mathrm{s}(\sum_{\mathbf{b}\in D^{n_2}} E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b})) = 1$, because $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) = 1$. Therefore, we have that

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \mathrm{s}(\sum_{\mathbf{b}\in D^{n_2}} E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b})) \cdot \prod_{\mathbf{b}\,:\, E_2^{\mathbf{I}}(\mathbf{b})\neq 0} E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b}) = (E_1^{\mathbf{I}} \div E_2^{\mathbf{I}})(\mathbf{a}) = E^{\mathbf{I}}(\mathbf{a}).$$

Suppose now that $\mathbf{a} = (a_1,\ldots,a_{n_1-n_2}) \notin \mathrm{adom}(\mathbf{I})^{n_1-n_2}$. In this case, we have that $\varphi_E^{\mathbf{A}(\mathbf{I})} = 0$. Furthermore, for every tuple $\mathbf{b}$, we have that $(\mathbf{a},\mathbf{b}) \notin \mathrm{adom}(\mathbf{I})^{n_1}$, hence $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) = 0$. By induction hypothesis about φ_1 and E_1, we have that $E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b}) = 0$ for every tuple $\mathbf{b}$, hence $\mathrm{s}(\sum_{\mathbf{b}\in D^{n_2}} E_1^{\mathbf{I}}(\mathbf{a},\mathbf{b})) = 0$ and so. by Definition 17, $E^{\mathbf{I}}(\mathbf{a}) = (E_1^{\mathbf{I}} \div E_2^{\mathbf{I}})(\mathbf{a}) = 0$. Thus, once again, we have that $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = E^{\mathbf{I}}(\mathbf{a})$. This completes the proof of the correctness of the formula φ_E.

Next, we show that φ_E is domain independent. We must show that for every $\mathbb{K}$-database $\mathbf{I} = (R_1,\ldots,R_p)$ and for every $\mathbb{K}$-structure $\mathbf{B} = (B, R_1,\ldots,R_p)$ with the same relations as $\mathbf{I}$ and with $\mathrm{adom}(\mathbf{I}) \subseteq B \subseteq D$, we have that $\varphi_E^{\mathbf{B}} = \varphi_E^{\mathbf{A}(\mathbf{I})}$, which means that $\varphi_E^{\mathbf{B}}(\mathbf{a}) = \varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$ holds, for every $\mathbf{a} \in D^{n_1-n_2}$. We distinguish three cases for a tuple $\mathbf{a}$ in $D^{n_1-n_2}$.

Case 1: $\mathbf{a} \notin B^{n_1-n_2}$. Since $\mathrm{adom}(\mathbf{I}) \subseteq B$, we have that $\mathbf{a} \notin \mathrm{adom}(\mathbf{I})^{n_1-n_2}$, as well. Hence, by Definition 12, we have that $\varphi_E^{\mathbf{B}}(\mathbf{a}) = 0 = \varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$.

Case 2: $\mathbf{a} \in B^{n_1-n_2}$ and for every $\mathbf{b} \in D^{n_2}$, we have that $\varphi_1^{\mathbf{B}}(\mathbf{a},\mathbf{b}) = 0$. Since $\mathbf{a} \in B^{n_1-n_2}$, we have that

$$\varphi_E^{\mathbf{B}}(\mathbf{a}) = \|(\nabla\exists\mathbf{y}\varphi_1) \wedge (\forall\mathbf{y}((\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1)))\|(\mathbf{B},\mathbf{a}) =$$

$$\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{B},\mathbf{a}) \cdot \prod_{\mathbf{b}\in B^{n_2}} (\|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{B},\mathbf{a},\mathbf{b}).$$

Since $\varphi_1^{\mathbf{B}}(\mathbf{a},\mathbf{b}) = 0$ for every tuple $\mathbf{b} \in B^{n_2}$, it follows that $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{B},\mathbf{a}) = 0$, hence $\varphi_E^{\mathbf{B}}(\mathbf{a}) = 0$. Let us now consider $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$. If $\mathbf{a} \notin \mathrm{adom}(\mathbf{I})^{n_1-n_2}$, then, by Definition 12, we have that $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = 0$. If $\mathbf{a} \in \mathrm{adom}(\mathbf{I})^{n_1-n_2}$, then by the same calculations as above we have that

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) \cdot \prod_{\mathbf{b}\in\mathrm{adom}(\mathbf{I})^{n_2}} (\|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b}).$$

Since φ_1 is domain independent and since $\varphi_1^{\mathbf{B}}(\mathbf{a},\mathbf{b}) = 0$ for every tuple $\mathbf{b} \in B^{n_2}$, we have that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) = 0$ for every $\mathbf{b} \in \mathrm{adom}(\mathbf{I})^{n_2}$, hence $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) = 0$ and so $\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = 0 = \varphi_E^{\mathbf{B}}(\mathbf{a})$.
Case 3: $\mathbf{a} \in B^{n_1-n_2}$ and there is a tuple $\mathbf{b} \in D^{n_2}$ such that $\varphi_1^{\mathbf{B}}(\mathbf{a},\mathbf{b}) \neq 0$. By the domain independence of φ_1, we have that $\varphi_1^{\mathbf{A}(\mathbf{I})}(\mathbf{a},\mathbf{b}) \neq 0$, hence the tuple $(\mathbf{a},\mathbf{b})$ must belong to $\mathrm{adom}(\mathbf{I})^{n_1}$ and so the tuple $\mathbf{a}$ must belong to $\mathrm{adom}(\mathbf{I})^{n_1-n_2}$. Furthermore, we have that $\|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{B},\mathbf{a}) = 1 = \|\nabla\exists\mathbf{y}\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a}) = 1$, by the domain independence of $\nabla\varphi_1$ (which follows from the domain independence of φ_1 and the third part of Proposition 4). Using the same calculations as above, we have that

$$\varphi_E^{\mathbf{B}}(\mathbf{a}) = \prod_{\mathbf{b}\in\mathbf{B}^{n_2}} \|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{B},\mathbf{a},\mathbf{b})$$

and

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \prod_{\mathbf{b}\in\mathrm{adom}(\mathbf{I})^{n_2}} \|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b}).$$

Thus, to show that $\varphi_E^{\mathbf{B}}(\mathbf{a}) = \varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$, it remains to show that the expressions in the right-hand sides of the previous two equations have the same value. Fix a $\mathbf{b} \in B^{n_2}$. If $\|\nabla\varphi_2\|(\mathbf{B},\mathbf{b}) = 0$, then

$$\|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{B},\mathbf{a},\mathbf{b}) = 1,$$

hence such tuples $\mathbf{b}$ do not contribute to the product and can be ignored. Suppose $\|\nabla\varphi_2\|(\mathbf{B},\mathbf{b}) = 1$, which also means that $(\nabla\varphi_2)^{\mathbf{B}}(\mathbf{b}) = 1$. Since φ_2 is domain independent, the third part of Proposition 4 implies that $\nabla\varphi_2$ is domain independent, hence $(\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 1$, which, in turn, implies that $\mathbf{b}$ belongs to $\mathrm{adom}(\mathbf{I})^{n_2}$ (otherwise, $(\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 0$). This analysis shows that the only tuples that contribute to the right-hand sides of the two equations above are the tuples $\mathbf{b} \in \mathrm{adom}(\mathbf{I})^{n_2}$ such that $(\nabla\varphi_2)^{\mathbf{B}}(\mathbf{b}) = (\nabla\varphi_2)^{\mathbf{A}(\mathbf{I})}(\mathbf{b}) = 1$, which also means that $\|\nabla\varphi_2\|(\mathbf{B},\mathbf{b}) = \|\nabla\varphi_2\|(\mathbf{A}(\mathbf{I}),\mathbf{b}) = 1$. For such tuples $\mathbf{b}$ and since $\mathbf{a} \in \mathrm{adom}(\mathbf{I})^{n_1-n_2}$, we have that

$$\|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{B},\mathbf{a},\mathbf{b}) = \|\eta \prec \nabla\varphi_2\|(\mathbf{B},\mathbf{b}) + \|\nabla\varphi_2 \wedge \varphi_1\|(\mathbf{B},\mathbf{a},\mathbf{b}) = \|\varphi_1\|(\mathbf{B},\mathbf{a},\mathbf{b})$$

and

$$\|(\eta \prec \nabla\varphi_2) \vee (\nabla\varphi_2 \wedge \varphi_1))\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b}) = \|\eta \prec \nabla\varphi_2\|(\mathbf{A}(\mathbf{I}),\mathbf{b}) + \|\nabla\varphi_2 \wedge \varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b}) = \|\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b}).$$

By the domain independence of φ_1, we have that $\|\varphi_1\|(\mathbf{B},\mathbf{a},\mathbf{b}) = \|\varphi_1\|(\mathbf{A}(\mathbf{I}),\mathbf{a},\mathbf{b})$, which, in view of the preceding calculations, implies that $\varphi_E^{\mathbf{B}}(\mathbf{a}) = \varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$ holds in Case 3, as well. This completes the proof of the direction (1) $\implies$ (2). Note that, since the building blocks of the formula φ_E do not have the same free variables, we could not have used Proposition 4 to show that φ_E is domain independent in a modular way.

(3) $\implies$ (1) : Assume that φ is an $\mathcal{RC}(\tau)$-formula of the form $\forall y_i\psi$, where ψ is an $\mathcal{RC}(\tau)$-formula with $y_1,\ldots,y_n$ as its free variables and i is an index such that $1 \leq i \leq n$. For notational simplicity, we assume that $i = n$, so that φ is of the form $\forall y_n\psi$, where the free variables of ψ are $y_1,\ldots,y_n$. By induction hypothesis, there is an $\mathcal{RA}(\tau)$-expression E_ψ of arity n such that $E_\psi^{\mathbf{I}} = \psi^{\mathbf{A}(\mathbf{I})}$, for every $\mathbb{K}$-database $\mathbf{I}$. We take E_φ to be the $\mathcal{RA}$-expression $E_\psi \div E_{\mathrm{adom}}$, which has arity $n-1$. We must show that $E_\varphi^{\mathbf{I}} = \varphi^{\mathbf{A}(\mathbf{I})}$, for every $\mathbf{K}$-database $\mathbf{I}$, which means that $E_\varphi^{\mathbf{I}}(\mathbf{a}) = \varphi^{\mathbf{A}(\mathbf{I})}(\mathbf{a})$ holds, for

every $\mathbf{a} = (a_1, \ldots, a_{n-1}) \in D^{n-1}$. For every $\mathbf{a} \in D^{n-1}$ and by the semantics of the division operation,

$$E_\varphi^{\mathbf{I}}(\mathbf{a}) = (E_\psi \div E_{\mathrm{adom}})^{\mathbf{I}}(\mathbf{a}) = (E_\psi^{\mathbf{I}} \div E_{\mathrm{adom}}^{\mathbf{I}})(\mathbf{a}) = \mathrm{s}(\sum_{b \in D} E_\psi^{\mathbf{I}}(\mathbf{a}, b)) \ \cdot \prod_{E_{\mathrm{adom}}^{\mathbf{I}}(b) \neq 0} E_\psi^{\mathbf{I}}(\mathbf{a}, b).$$

By Proposition 3, $E_{\mathrm{adom}}^{\mathbf{I}} = \mathrm{adom}(\mathbf{I})$. Hence, for every $a \in D$, we have that $E_{\mathrm{adom}}^{\mathbf{I}}(a) \neq 0$ if and only if $a \in \mathrm{adom}(\mathbf{I})$. Consequently, for every tuple $\mathbf{a} \in D^{n-1}$, we have that

$$E_\varphi^{\mathbf{I}}(\mathbf{a}) = \mathrm{s}(\sum_{b \in D} E_\psi^{\mathbf{I}}(\mathbf{a}, b)) \ \cdot \prod_{b \in \mathrm{adom}(\mathbf{I})} E_\psi^{\mathbf{I}}(\mathbf{a}, b).$$

We now distinguish two cases for a tuple $\mathbf{a} \in D^{n-1}$.

Case 1: For every element $b \in D$, we have that $\psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) = 0$. By induction hypothesis about ψ and E_ψ, we have that $E_\psi^{\mathbf{I}}(\mathbf{a}, b) = \psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) = 0$ holds, for every $b \in D$. It follows that $\sum_{b \in D} E_\psi^{\mathbf{I}}(\mathbf{a}, b) = 0$, hence, by the last displayed equation, we have that $E_\varphi^{\mathbf{I}}(\mathbf{a}) = 0$. We claim that also $\varphi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = 0$ holds in this case. Indeed, if $\mathbf{a} \notin \mathrm{adom}(\mathbf{I})$, then $\varphi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = 0$, by Definition 12. If $\mathbf{a} \in \mathrm{adom}(\mathbf{I})$, then, by Definitions 11 and 12, we have that

$$\varphi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \|\forall y_i \psi\|(\mathbf{A}(\mathbf{I}), \mathbf{a}) = \prod_{b \in \mathrm{adom}(\mathbf{I})} \|\psi\|(\mathbf{A}(\mathbf{I}), \mathbf{a}) = \prod_{b \in \mathrm{adom}(\mathbf{I})} \psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) = 0,$$

as desired. Note that the product $\prod_{b \in \mathrm{adom}(\mathbf{I})} \psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b)$ is non-empty because $\mathrm{adom}(\mathbf{I})$ is non-empty.

Case 2: There is an element $b \in D$ such that $\psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) \neq 0$. By induction hypothesis about ψ and E_ψ, we have that $E_\psi^{\mathbf{I}}(\mathbf{a}, b) = \psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b)$, hence $E_\psi^{\mathbf{I}}(\mathbf{a}, b) \neq 0$ holds for at least one element $b \in D$. It follows that $\sum_{b \in D} E_\psi^{\mathbf{I}}(\mathbf{a}, b) = 1$, hence, we have that $E_\varphi^{\mathbf{I}}(\mathbf{a}) = \prod_{b \in \mathrm{adom}(\mathbf{I})} E_\psi^{\mathbf{I}}(\mathbf{a}, b)$. Since $\psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) \neq 0$ for some $b \in D$, it follows that $(\mathbf{a}, b) \in \mathrm{adom}(\mathbf{I})^n$ (otherwise, $\psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) = 0$, by Definition 12). Thus, $\mathbf{a} \in \mathrm{adom}(\mathbf{I})^{n-1}$, hence, by Definitions 11 and 12, and the induction hypothesis,

$$\varphi_E^{\mathbf{A}(\mathbf{I})}(\mathbf{a}) = \prod_{b \in \mathrm{adom}(\mathbf{I})} \psi^{\mathbf{A}(\mathbf{I})}(\mathbf{a}, b) = \prod_{b \in \mathrm{adom}(\mathbf{I})} E_\psi^{\mathbf{I}}(\mathbf{a}, b) = E_\varphi^{\mathbf{I}}(\mathbf{a}). \quad \square$$

Immediate corollaries of Theorem 3 include versions of Codd's Theorem for the bag semiring, the fuzzy semiring, the Łukasiewicz semiring, the tropical semiring, and various provenance semirings.

5 Concluding Remarks

The work presented here delineates the connections between relational algebra and relational calculus on databases over semirings by establishing two versions of Codd's Theorem in this setting. Along the way, new differences between the Boolean semiring and other semirings of importance to databases were unveiled. In particular, the five basic operations of relational algebra cannot express the division operation on the bag semiring. We single out two directions for future research:

1. For which semirings is it the case that basic relational algebra $\mathcal{BRA}$ can express the division operation? Is there a characterization of these semirings in terms of their structural properties?

2. In this paper, we had to make certain choices concerning the difference operation in relational algebra and the form of negation in relational calculus; to this effect, we chose the monus operation and the $\prec$ ("but not") connective, respectively. As described in [17, 26], there are other ways to define the difference operation in relational algebra over semirings. Are there versions of Codd's Theorem when alternative choices for the difference operation and for negation are made?

Finally, our results establish connections between database theory and non-classical logics, since the connective $\prec$ ("but not") has featured in such logics. For example, relational calculus $\mathcal{RC}$, as defined here, has connections to the dual intuitionistic logic. This paves the way for interaction between database theory and non-classical logics, two areas that had no interaction in the past.

References

[1] Serge Abiteboul, Richard Hull, and Viktor Vianu. *Foundations of Databases*, volume 1. Addison-Wesley, 1995.

[2] Samson Abramsky. Relational databases and Bell's theorem. In Val Tannen, Limsoon Wong, Leonid Libkin, Wenfei Fan, Wang-Chiew Tan, and Michael P. Fourman, editors, *In Search of Elegance in the Theory and Practice of Computation - Essays Dedicated to Peter Buneman*, volume 8000 of *Lecture Notes in Computer Science*, pages 13–35. Springer, 2013. doi: 10.1007/978-3-642-41660-6_2. URL https://doi.org/10.1007/978-3-642-41660-6_2.

[3] K. Amer. Equationally complete classes of commutative monoids with monus. *Algebra Universalis*, 18:129–131, 1984.

[4] Yael Amsterdamer, Daniel Deutch, and Val Tannen. On the limitations of provenance for queries with difference. In Peter Buneman and Juliana Freire, editors, *3rd Workshop on the Theory and Practice of Provenance, TaPP'11, Heraklion, Crete, Greece, June 20-21, 2011*. USENIX Association, 2011. URL https://www.usenix.org/conference/tapp11/limitations-provenance-queries-difference.

[5] Albert Atserias and Phokion G. Kolaitis. Acyclicity, consistency, and positive semirings. In Alessandra Palmigiano and Mehrnoosh Sadrzadeh, editors, *Samson Abramsky on Logic and Structure in Computer Science and Beyond*, volume 25 of *Outstanding Contributions to Logic*, pages 623–668. Springer, 2023.

[6] Albert Atserias and Phokion G. Kolaitis. Consistency of relations over monoids. *J. ACM*, 72(3):18:1–18:47, 2025. doi: 10.1145/3721855. URL https://doi.org/10.1145/3721855.

[7] Edgar F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970. doi: 10.1145/362384.362685. URL https://doi.org/10.1145/362384.362685.

[8] Edgar F. Codd. Relational completeness of data base sublanguages. In R. Rustin, editor, *Proceedings of 6th Courant Computer Science Symposium (May 24–25, 1971: New York, N.Y.)*, pages 65–98. Prentice-Hall, 1972.

[9] Katrin M. Dannert, Erich Grädel, Matthias Naaf, and Val Tannen. Semiring provenance for fixed-point logic. In Christel Baier and Jean Goubault-Larrecq, editors, *29th EACSL Annual Conference on Computer Science Logic, CSL 2021, January 25-28, 2021, Ljubljana, Slovenia (Virtual Conference)*, volume 183 of *LIPIcs*, pages 17:1–17:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi: 10.4230/LIPICS.CSL.2021.17. URL https://doi.org/10.4230/LIPIcs.CSL.2021.17.

[10] Floris Geerts and Antonella Poggi. On database query languages for K-relations. *Journal of Applied Logic*, 8(2):173–185, 2010.

[11] Allen Van Gelder and Rodney W. Topor. Safety and translation of relational calculus queries. *ACM Transactions on Database Systems*, 16(2):235–278, 1991. doi: 10.1145/114325.103712. URL https://doi.org/10.1145/114325.103712.

[12] Nicolas D. Goodman. The logic of contradiction. *Mathematical Logic Quarterly*, 27(8–10):119–126, 1981.

[13] Todd J. Green. Containment of conjunctive queries on annotated relations. *Theory of Computing Systems*, 49:429–459, 2011.

[14] Todd J. Green, Grigoris Karvounarakis, and Val Tannen. Provenance semirings. In *PODS 2007 Proceedings of the Twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of Database Systems*, pages 31–40. ACM Press, 2007.

[15] Erich Grädel and Val Tannen. Semiring provenance for first-order model checking. Available at https://arxiv.org/abs/1712.01980, 2017.

[16] Erich Grädel and Val Tannen. Provenance analysis and semiring semantics for first-order logic. In *Model Theory, Computer Science, and Graph Polynomials*, pages 351–401. Birkhäuser, 2025. Festschrift in Honor of János Makowsky.

[17] M. Hajdinjak and A. Bauer. K-relations and beyond. In C. R. Gutiérrez, editor, *Advances in Knowledge Representation*, pages 19–40. InTech, 2012.

[18] Arend Heyting. Die formalen regeln der intuitionistischen logik. *Sitzungsbericht PreuBische Akademie der Wissenschaften Berlin, physikalisch-mathematische Klasse II*, pages 42–56, 1930.

[19] Mahmoud Abo Khamis, Hung Q. Ngo, Reinhard Pichler, Dan Suciu, and Yisu Remy Wang. Convergence of Datalog over (pre-) semirings. *Journal of the ACM*, 71(2):8:1–8:55, 2024. doi: 10.1145/3643027. URL https://doi.org/10.1145/3643027.

[20] Egor V. Kostylev, Juan L. Reutter, and András Z. Salamon. Classification of annotation semirings over containment of conjunctive queries. *ACM Transactions of Database Systems*, 39(1):1:1–1:39, 2014. doi: 10.1145/2556524. URL https://doi.org/10.1145/2556524.

[21] J.C.C. McKinsey. Proof of the independence of the primitive symbols of Heyting's calculus of propositions. *Journal of Symbolic Logic*, 4(4):155–158, 1939.

[22] J.C.C. McKinsey and Alfred Tarski. On closed elements in closure algebras. *Annals of Mathematics*, 47(2):122–162, 1946.

[23] Mikaël Monet. Example of a naturally ordered semiring which is not an m-semiring. https://math.stackexchange.com/questions/1966858/example-of-a-naturally-ordered-semiring-which-is-not-an-m-semiring, 2016. StackExchange.

[24] Cecylia Rauszer. Semi-Boolean algebras and their applications to intuitionistic logic with dual operations. *Fundamenta Mathematicae*, 83(3):219–249, 1974. URL http://eudml.org/doc/214696.

[25] Timothy Smiley. The independence of connectives. *Journal of Symbolic Logic*, 40(2):250–251, 1975. doi: 10.2307/2271931.

[26] Dan Suciu. Different differences in semirings. In Antoine Amarilli and Alin Deutsch, editors, *The Provenance of Elegance in Computation - Essays Dedicated to Val Tannen, Tannen's Festschrift, May 24-25, 2024, University of Pennsylvania, Philadelphia, PA, USA*, volume 119 of *OASIcs*, pages 10:1–10:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024. doi: 10.4230/OASICS.TANNEN.10. URL https://doi.org/10.4230/OASIcs.Tannen.10.
[27] Jeffrey D. Ullman. *Principles of Database Systems, 1st Edition.* Computer Science Press, 1980.
[28] Hangdong Zhao, Shaleen Deep, Paraschos Koutris, Sudeepa Roy, and Val Tannen. Evaluating Datalog over semirings: A grounding-based approach. *Proceedings of the ACM on Management of Data*, 2(2):90, 2024. doi: 10.1145/3651591. URL https://doi.org/10.1145/3651591.

Received June 2025; accepted August 2025