Reduction of database independence to dividing in atomless Boolean algebras

Tapani Hyttinen¹ · Gianluca Paolini¹

Received: 5 September 2014 / Accepted: 22 February 2016 / Published online: 4 March 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract We prove that the form of conditional independence at play in database theory and independence logic is reducible to the first-order dividing calculus in the theory of atomless Boolean algebras. This establishes interesting connections between independence in database theory and stochastic independence. As indeed, in light of the aforementioned reduction and recent work of Ben-Yaacov (Isr. J. Math. 194(2):957–1012, 2013), the former case of independence can be seen as the discrete version of the latter.

Keywords Independence · Database theory · Dividing · Atomless Boolean algebras

Mathematics Subject Classification 03C45 · 06E05 · 68P15 · 03C80

1 Introduction

At the current state of the art, the model-theoretic study of independence is a well-developed subject with numerous examples arising from different areas of mathematics. So-called independence calculi have been elaborated for several classes of structures, e.g. classes elementary in classical first-order logic, classes elementary

¹ The research of the second author was supported by the Finnish Academy of Science and Letters (Vilho, Yrjö and Kalle Väisälä foundation). The authors would like to thank the referee for suggesting a simpler and slightly more general proof of the main theorem. The authors would also like to thank John Baldwin and Jouko Väänänen for useful suggestions and discussions related to this paper.

✉ Gianluca Paolini
  gianluca.paolini@helsinki.fi

¹ Department of Mathematics and Statistics, University of Helsinki,
  Gustaf Hällströminkatu 2b, 00014 Helsinki, Finland
in continuous logic, finitary abstract elementary classes, etc. Furthermore, these calculi are canonical, in the sense that the satisfaction of few crucial axioms imposes their uniqueness (for an overview see [9]). In Table 1 we see some examples. In all these examples the natural internal notion of independence coincides with the model-theoretic one.

Another important case of independence which has been studied from the point of view of model theory is stochastic independence. In [4], Ben-Yaacov shows that the class of atomless probability algebras is elementary in continuous logic and that in this class non-forking corresponds to the familiar notion of independence of probability algebras, which in turn corresponds to independence of random variables, a.k.a. stochastic independence.

In various papers on statistics and database theory, e.g. [17, 19] and [21], stochastic independence has been put in relation with a possibly less familiar notion of independence, known as embedded multivalued dependence or database independence, for simplicity. In these papers it is shown that these two forms of independence agree on a large group of axioms, which resemble very closely some basic axioms satisfied by non-forking in simple first-order theories, e.g. monotonicity, symmetry and transitivity.

We then ask: how does database independence relate to the general theory of independence as developed in model theory? Is there a way to make sense of the noticed similarities between stochastic independence and database independence from the point of view of model theory?

In the following we answer positively both questions, showing that, in analogy with Ben Yaacov’s work, database independence is reducible to first-order dividing calculus in the theory of atomless Boolean algebras. Our reduction shows that both

| Class                        | Example                          | Independence                   | References |
|------------------------------|----------------------------------|--------------------------------|------------|
| Elementary in classical FOL | Vector spaces                    | Linear independence            | [20]       |
| Elementary in continuous logic | Hilbert spaces                   | Orthogonality                  | [5]        |
| Homogeneous metric classes   | $\ell_p$-spaces                  | Variations of orthogonality    | [10]       |
| Quasi-minimal pregeometry structures | Covers of smooth algebraic curves | Independence from Zariski topology | [12] |
| Other abstract elementary classes | $\mathbb{Z}_p$-modules without elem. of infinite height | For closed pure submodules the usual amalgam of modules$^a$ | [11]$^b$ |
| Almost abstract elementary classes$^c$ | Certain classes of geometric lattices of fixed finite rank | Free amalgam of geometric lattices$^a$ | [14] |

$^a$ I.e. the sets are independent if and only if relative to each other the positions of the sets are as in the amalgam

$^b$ This is an AEC version of the metric class studied in [11]. For more see [3]

$^c$ I.e. classes of structures with a strong submodel relation which are not AECs but still behave well. E.g. in [14] all the axioms of an AEC but the Smoothness Axiom are satisfied, but it is still possible to define an independence notion along to the lines of good frames
database independence and stochastic independence are essentially based on the same independence relation, i.e. free amalgamation, considered in the category of Boolean algebras and probability algebras, respectively.

The structure of the paper is as follows. In Sect. 2 we first give some basic definitions, and then introduce an independence relation between Boolean algebras and prove some preparatory lemmas. In Sect. 2.1 we prove our main theorem, which is a partial characterization of dividing in atomless Boolean algebra. In Sect. 2.2 we then use this characterization to establish an abstract reduction of database independence to dividing in atomless Boolean algebras. In Sect. 3.1 we introduce some basic definitions from database theory, formulate database independence, and make concrete the above mentioned reduction. Finally, we spend a few words on dependence logic, explaining why this form of independence can be identified with database independence and how the present work motivates probabilistic versions of independence logic, currently under development by the authors.

2 Independence in atomless Boolean algebras

For model-theoretic notations and conventions we refer to Marker [16]. As usual with contemporary model theory, we work in large homogeneous models with a high degree of saturation, also known as monster models.

**Definition 2.1** Let $T$ be a complete first-order theory with infinite models, and $\mathcal{M}$ its monster model. For $a \in \mathcal{M}^n$, $b \in \mathcal{M}^m$ and $C \subseteq \mathcal{M}$ we say that $a$ is dividing independent from $b$ over $C$, in symbols $a \vdash_{C} b$, if for every $(b_1)_{i<\omega} \in (\mathcal{M}^m)^{\omega}$ indiscernible over $C$ such that $b_0 = b$, there exists $a' \in \mathcal{M}^n$ such that:

(i) $\text{tp}(a'/Cb) = \text{tp}(a/Cb)$;
(ii) $(b_i)_{i<\omega}$ is indiscernible over $Ca'$.

For notions and notation concerning Boolean algebras we refer to Givant and Halmos [6], for the model-theoretic aspects of the theory of Boolean algebras to Poizat [18]. We denote by BA the theory of Boolean algebras in the signature $\mathcal{L} = \{0, 1, \wedge, \vee, \neg\}$, with ABA the theory of atomless Boolean algebras and with ATBA the theory of atomic Boolean algebras. It is well-known that the theory ABA is complete and admits elimination of quantifiers. This theory is neither NIP nor NSOP, and in particular it is not simple. Analogously, the theory ATBA is complete and admits elimination of quantifiers in the language extended with predicates $(A_i(x))_{i<\omega}$ saying that $x$ dominates at most $i$ atoms.

The class of Boolean algebras with homomorphisms of Boolean algebras as morphisms is a category, we denote it by $\text{BolAlg}$. Given $\mathcal{A} \models \text{BA}$ and $\mathcal{B}$ a $\{0, 1, \wedge, \vee, \neg\}$-structure, we denote by $\mathcal{B} \leqslant \mathcal{A}$ the substructure relation (notice that BA is a $\forall$-theory and so its class of models is closed under taking substructures). For $\mathcal{A}, \mathcal{B} \models \text{BA}$ and $\mathcal{C} \leqslant \mathcal{A}, \mathcal{B}$, we denote by $\mathcal{A} \otimes_{\mathcal{C}} \mathcal{B}$ the push-out of $\mathcal{A}$ and $\mathcal{B}$ over $\mathcal{C}$ in $\text{BolAlg}$. The object $\mathcal{A} \otimes_{\mathcal{C}} \mathcal{B}$ can be shown to be the Boolean algebras corresponding to the tensor product of the Boolean rings corresponding to $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$, respectively.
Definition 2.2 Let $\mathcal{D} \models \mathcal{B}, \mathcal{A}, \mathcal{B} \leq \mathcal{D}$ and $\mathcal{C} \leq \mathcal{A}, \mathcal{B}$. We let

$$\mathcal{A} \downarrow \mathcal{B} \iff (A \cup B) \cong A \otimes \mathcal{C} B.$$ 

This structural definition of independence between Boolean algebras is equivalent to a more intuitive condition of topological nature.

Lemma 2.3 Let $\mathcal{D} \models \mathcal{B}, \mathcal{A}, \mathcal{B} \leq \mathcal{D}$ and $\mathcal{C} \leq \mathcal{A}, \mathcal{B}$. Then

$$\mathcal{A} \downarrow \mathcal{B} \iff \forall a \in A \text{ and } b \in B \left[ a \leq b \Rightarrow \exists c \in C \ a \leq c \leq b \right].$$

Proof See [2].

Given $\mathcal{A} \models \mathcal{B}$, we denote by $\text{At}(\mathcal{A})$ the set of atoms of $\mathcal{A}$. We state few elementary facts about Boolean algebras that will be used freely throughout the paper.

Proposition 2.4 Let $\mathcal{A} \models \mathcal{B}$ and $a \in A$. The following are equivalent.

(i) $a$ is an atom.
(ii) For every $b \in A$ either $a \leq b$ or $a \wedge b = 0$, but not both.

Proposition 2.5 Let $\mathcal{A} \models \mathcal{B}$ and $a \in A$. The following are equivalent.

(i) $\mathcal{A}$ is atomic;
(ii) Every $a \in A$ is the supremum of the atoms that it dominates.
(iii) The unit is the supremum of the set of all atoms.

Proposition 2.6 Let $\mathcal{A} \models \mathcal{B}$, $a \in A$ and $(b_i)_{i \in I} \in A^I$. If $\bigvee_{i \in I} b_i$ exists, then

$$a \wedge \bigvee_{i \in I} b_i = \bigvee_{i \in I} (a \wedge b_i).$$

In particular, the supremum on the right-hand side of the equation exists.

Notation 2.7 In Proposition 2.6, we considered sups of families of points of a Boolean algebras without specifying the algebra with respect to which they were considered. This lack of notation may lead to some problems, in fact given $\mathcal{C} \leq \mathcal{A}$ and $(c_j)_{j \in J} \in C^J$ there may be a point $c \in C$ such that $c$ is the sup of the $c_j$ with respect to the algebra $\mathcal{C}$ but not with respect to the algebra $\mathcal{A}$. When we want to make clear with respect to which algebra we are considering a sup (resp. an inf) we use the notation $(\bigvee_{j \in J} c_j)^\mathcal{A}$ (resp. $(\bigwedge_{j \in J} c_j)^\mathcal{A}$), where obviously the superscript indicates the algebra with respect to which the sup (resp. the inf) is taken. This notation is heavier and it will be avoided whenever possible, it is, though, of crucial importance in the following definitions.

Definition 2.8 (Regular Subalgebra [6]) Let $\mathcal{A} \models \mathcal{B}$ and $\mathcal{C} \leq \mathcal{A}$. We say that $\mathcal{C}$ is a regular subalgebra of $\mathcal{A}$ if whenever $(c_i)_{i \in I} \in C^I$ and $(\bigvee_{i \in I} c_i)^\mathcal{C}$ exists, then $(\bigvee_{i \in I} c_i)^\mathcal{C} = (\bigvee_{i \in I} c_i)^\mathcal{A}$. 

$\copyright$ Springer
Definition 2.9 Let $\mathcal{A} \models \text{BA}$ and $\mathcal{C} \leq \mathcal{A}$. We say that $\mathcal{C}$ sits nicely in $\mathcal{A}$, in symbols $\mathcal{C} \preceq^* \mathcal{A}$, if the following conditions are satisfied:

(i) $\mathcal{C}$ is a regular subalgebra of $\mathcal{A}$;
(ii) for every $a \in \mathcal{A}$ there exists least $c \in \mathcal{C}$ such that $a \leq c$.

Under the assumptions of atomicity of $\mathcal{C}$ and that $\mathcal{C}$ sits nicely in $\mathcal{A}$ and $\mathcal{B}$ we can give a characterization of the relation $\mathcal{A} \perp_{\mathcal{C}} \mathcal{B}$ which will be very useful in the proof of the main theorem. The proof of this characterization is fairly standard, we include it for completeness of exposition.

Lemma 2.10 Let $\mathcal{D} \models \text{BA}$, $\mathcal{A}, \mathcal{B} \leq \mathcal{D}$ and $\mathcal{C} \preceq^* \mathcal{A}, \mathcal{B}$, with $\mathcal{C}$ atomic. Then

$\mathcal{A} \perp_{\mathcal{C}} \mathcal{B} \iff \forall c \in \text{At}(\mathcal{C}), a \in \mathcal{A}$ and $b \in \mathcal{B}$

\[
\left[ a, b \leq c \text{ and } a, b \not\in \{0, c\} \Rightarrow a \text{ incomparable to } b \right].
\]

Proof For the direction ($\Rightarrow$), suppose there are $c \in \text{At}(\mathcal{C})$, $a \in \mathcal{A}$ and $b \in \mathcal{B}$, such that $a, b \leq c$, $a, b \not\in \{0, c\}$, but $a \leq b$ or $b \leq a$. Suppose $a \leq b$, then it is impossible to find $c' \in \mathcal{C}$ such that $0 < a \leq c' \leq b < c$, because $c$ is an atom. Suppose $b \leq a$, then $\neg a \leq \neg b$, and so again it is impossible to find $c' \in \mathcal{C}$ such that $\neg c < \neg a \leq c' \leq \neg b < 1$, because $c$ is an atom.

For the direction ($\Leftarrow$), assume the right hand side and let $a \in \mathcal{A}$ and $b \in \mathcal{B}$, with $a \leq b$. Given that $\mathcal{C} \preceq^* \mathcal{A}$, there is least $c \in \mathcal{C}$ such that $a \leq c$. We show that $c \leq b$. By atomicity of $\mathcal{C}$, we can find distinct $(c_i)_{i \in I} \in \text{At}(\mathcal{C})^I$ such that $c = (\bigvee_{i \in I} c_i)^{\mathcal{C}}$, and, by the regularity assumption, $(\bigvee_{i \in I} c_i)^{\mathcal{C}} = (\bigvee_{i \in I} c_i)^{\mathcal{A}} = (\bigvee_{i \in I} c_i)^{\mathcal{B}}$. It suffices to show that for every $j \in I$ we have that $c_j \leq b$, because then $c = (\bigvee_{i \in I} c_i)^{\mathcal{C}} = (\bigvee_{i \in I} c_i)^{\mathcal{B}} \leq b$. Let then $j \in I$. By assumption $a \leq b$, so we have that $a \wedge c_j \leq b \wedge c_j \leq c_j$ and then, by the right hand side, $a \wedge c_j \in \{0, c_j\}$ or $b \wedge c_j \in \{0, c_j\}$. Now, if $a \wedge c_j \in \{0, c_j\}$, then $a \leq (\bigvee_{i \in I \setminus \{j\}} c_i)^{\mathcal{C}} = (\bigvee_{i \in I \setminus \{j\}} c_i)^{\mathcal{A}} < c$, contradicting the fact that $c$ is least. On the other hand, if $b \wedge c_j = 0$, then $a \wedge b \wedge c_j = a \wedge c_j = 0$, which as already been shown impossible. Then, we must have that $b \wedge c_j = c_j$, i.e. $c_j \leq b$, as wanted. 

If on top of the assumptions of the previous lemma we further assume that $\mathcal{A}$ and $\mathcal{B}$ are atomic we can improve our characterization of independence of Boolean algebras. This characterization will not be needed in the proof of the main theorem, but it will simplify the arguments in Sect. 2.2. We include it here for coherence of exposition. We first need a couple of lemmas, whose proofs are also standard.

Lemma 2.11 Let $\mathcal{A} \models \text{BA}$ and $\mathcal{C} \leq \mathcal{A}$, with $\mathcal{C}$ atomic and regular in $\mathcal{A}$. Then, for every $a \in \text{At}(\mathcal{A})$ there exists exactly one $c \in \text{At}(\mathcal{C})$ such that $a \leq c$.

Proof Let $a \in \text{At}(\mathcal{A})$ and suppose that there are $c_0, c_1 \in \text{At}(\mathcal{C})$ such that $a \leq c_0, c_1$ and $c_0 \neq c_1$. Then $a \leq c_0 \land c_1 = 0$, a contradiction. Suppose now that for every
c ∈ At(\mathcal{C}) we have that \( a \not\leq c \). Then \( a \land c = 0 \) for every \( c ∈ At(\mathcal{C}) \), because \( a \) is an atom of \( \mathcal{A} \). Thus,

\[
a = a \land 1 = a \land \left( \bigvee_{c ∈ At(\mathcal{C})} c \right)^\mathcal{C} = a \land \left( \bigvee_{c ∈ At(\mathcal{C})} c \right)^\mathcal{A} = \left( \bigvee_{c ∈ At(\mathcal{C})} (a \land c) \right)^\mathcal{A} = 0,
\]
a contradiction.

By Lemma 2.11 whenever we have \( \mathcal{A} \models BA \) and \( \mathcal{C} \leq \mathcal{A} \), with \( \mathcal{C} \) atomic, we have a function \( f: At(\mathcal{A}) → At(\mathcal{C}) \) which sends an atom \( a \) to the unique atom \( c \) that dominates \( a \). We denote the image of this map at a point \( a \) by \( C^{*}a \).

**Lemma 2.12** Let \( \mathcal{D} \models BA \), \( \mathcal{A}, \mathcal{B} \leq \mathcal{D} \) and \( \mathcal{C} \leq \mathcal{A} \), with \( \mathcal{A}, \mathcal{B} \) and \( \mathcal{C} \) atomic. Then

\[
\mathcal{A} \downarrow \mathcal{C} \mathcal{B} ⇔ \forall a ∈ At(\mathcal{A}) \text{ and } b ∈ At(\mathcal{B})
\[
\left[ a \leq \neg b \Rightarrow \exists c ∈ At(\mathcal{C})\left[ a \leq \neg c \text{ and } b \leq c \right] \right].
\]

**Proof** For the direction \((⇒)\), let \( a ∈ At(\mathcal{A}) \) and \( b ∈ At(\mathcal{B}) \), and suppose that \( a \land b = 0 \). By hypothesis there is \( c ∈ C \) such that \( a \land c = 0 \) and \( b \land \neg c = 0 \). Let \( C^{*}b \downarrow = c^{*} \). Notice that \( c \land c^{*} \neq 0 \) because otherwise \( b = 0 \), given that \( b \leq c, c^{*} \). But then \( c^{*} \leq c \) because \( c^{*} ∈ At(\mathcal{C}) \) and \( c ∈ C \). Hence, \( a \leq \neg c^{*} \) and \( b \leq c^{*} \) because \( a \leq \neg c \leq \neg c^{*} \).

For the direction \((⇐)\), let \( a ∈ A \) and \( b ∈ B \), and suppose that \( a \land b = 0 \). If \( a = 0 \) or \( b = 0 \), we can find the wanted \( c \), in the first case \( c = 1 \) and in the second \( c = 0 \). Suppose that \( a, b \neq 0 \), and let \( (a_{i})_{i ∈ I} ∈ At(\mathcal{A})^{I} \) and \( (b_{j})_{j ∈ J} ∈ At(\mathcal{B})^{J} \) such that

\[
a = \bigvee_{i ∈ I} a_{i} \text{ and } b = \bigvee_{j ∈ J} b_{j}.
\]

Notice that \( a_{i} \land b_{j} = 0 \) for every \( i ∈ I \) and \( j ∈ J \), because \( a \land b = 0 \). Thus, by assumption, for every \( i ∈ I \) and \( j ∈ J \) there is \( c_{i,j} ∈ At(\mathcal{C}) \) such that

\[
a_{i} \land c_{i,j} = 0 \text{ and } b_{j} \land \neg c_{i,j} = 0.
\]

For \( j ∈ J \), let \( c_{j}^{*} = C^{*}b_{j} \downarrow \). Then for every \( i ∈ I \) and \( j ∈ J \) we have that \( c_{i,j} = c_{j}^{*} \). And so for every \( i ∈ I \) and \( j ∈ J \) we have that

\[
\neg c_{j}^{*} = \neg c_{i,j} \geq a_{i} \text{ and } c_{j}^{*} = c_{i,j} \geq b_{j}.
\]

Define \( c = (\bigvee_{j ∈ J} c_{j}^{*})^{\mathcal{C}} \). Notice that \( c \) exists, and also

\[
\left( \bigvee_{j ∈ J} c_{j}^{*} \right)^{\mathcal{B}} = c = \left( \bigvee_{j ∈ J} c_{j}^{*} \right)^{\mathcal{A}}
\]

\( \text{Springer} \)
because by hypothesis $C \subseteq^* A, B$. We claim that i) $c \geq b$ and ii) $\neg c \geq a$. To see i), notice that for every $j \in J$ we have that $c_j^* \geq b_j$ and so $c = \bigvee_{j \in J} c_j^* \geq \bigvee_{j \in J} b_j = b$. To see ii), notice that $\neg c = \bigwedge_{j \in J} \neg c_j^*$ and that $\neg c_j^* \geq a$ for every $j \in J$, because for every $i \in I$ we have that $\neg c_j^* \geq a_i$ and $a = \bigvee_{i \in I} a_i$. \hfill \Box

Finally, we can state our improved characterization of $A \downarrow_C B$.

**Corollary 2.13** Let $\mathcal{D} \models BA$, $A, B \subseteq \mathcal{D}$ and $C \subseteq^* A, B$, with $A, B$ and $C$ atomic. Then

$$A \downarrow_C B \iff \forall a \in \text{At}(A) \text{ and } b \in \text{At}(B) \left[ a \leq C^\geq b \downarrow \Rightarrow a \land b \neq 0 \right].$$

**Proof** Immediate from Lemma 2.12. \hfill \Box

**Remark 2.14** Notice that in Lemma 2.12 the assumption $C \subseteq^* A, B$ is necessary. To see this, consider the following subalgebras of $\mathcal{P}(\omega)$

$$A = \langle [\{2n, 2n + 2\} \mid n \in 2\omega] \cup \langle [\{2n + 1\} \mid n \in \omega]\rangle,$$

$$C = \langle [\{2n + 1, 2n + 3\} \mid n \in 2\omega] \cup \langle [\{2n, 2n + 2\} \mid n \in 2\omega]\rangle,$$

$$B = \langle [\{2n + 1, 2n + 3\} \mid n \in 2\omega] \cup \langle [\{2n\} \mid n \in \omega]\rangle.$$ 

where for $\mathcal{D} \models BA$ with $\mathcal{D}$ complete and $\mathcal{E} \subseteq \mathcal{D}$ we denote by $\overline{\mathcal{E}}$ the completion of $\mathcal{E}$ in $\mathcal{D}$, and, as usual with fields of sets, we identify structures with their domains.

Then $C \subseteq A, B$ and for all $a \in \text{At}(A)$ and $b \in \text{At}(B)$ whenever $a \cap b = \emptyset$ there is $c \in \text{At}(C)$ such that $a \cap c = \emptyset$ and $b \cap \neg c = \emptyset$. But $A \downarrow_C B$ because for

$$a = \{2n \mid n \in \omega\} \text{ and } b = \{2n + 1 \mid n \in \omega\}$$

we have that $a \cap b = \emptyset$, but for every $c \in C$ if $b \subseteq c$ then $a \cap c \neq \emptyset$.

We conclude this section with a basic lemma about atomless Boolean algebras which will be relevant in the proof of the main theorem.

**Lemma 2.15** Let $A \models ABA$. Then for every $a \in A$ with $a \neq 0$ there exists $(p_n(a))_{n<\omega} \in A^\omega$ such that for every $k < \omega$ the following holds:

(i) $p_k(a) \neq 0$;

(ii) $p_k(a) \land \bigvee_{i<k} p_i(a) = 0$;

(iii) $\bigvee_{i\leq k} p_i(a) \leq a$.

### 2.1 Main theorem

The various lemmas proved in the previous section allowed us to characterize the relation $A \downarrow_C B$ under the assumption of atomicity of $C$ and that it sits nicely in $A$ and $B$. We now use this characterization to establish a reduction of the relation $A \downarrow_C B$ to dividing in ABA. This is the content of the following theorem.
Theorem 2.16 (Main theorem) Let $\mathcal{M}$ be the monster model of ABA, and $\mathcal{A}, \mathcal{B}, \mathcal{C} \subseteq \mathcal{M}$, with $\mathcal{C}$ atomic and $\mathcal{C} \preceq^* \mathcal{A}, \mathcal{B}$. Then

$$A \downarrow \mathcal{C} B \iff \forall a \in A \text{ and } b \in B \left[ a \downarrow \mathcal{C} b \right].$$

Proof For the direction ($\Leftarrow$), suppose that $a, b, c$ are a witness for $A \downarrow \mathcal{C} B$ as in Lemma 2.10. Possibly replacing $a$ with $c \land \neg a$ and $b$ with $c \land \neg b$, we may assume that $a \leq b$. Since $b < c$, using Lemma 2.15, we can find $(b_i)_{i < \omega}$ indiscernible over $C$ such that $b_0 = b$ and $b_i \land b_j = 0$ for every $i < j < \omega$, witnessing that $\text{tp}(a/Cb)$ divides over $C$.

For the direction ($\Rightarrow$), suppose that $A \downarrow \mathcal{C} B$ and let $a \in A$, $b \in B$ and $(b_i)_{i < \omega}$ indiscernible over $C$ such that $b_0 = b$. Let $c \in \text{At}(\mathcal{C})$. If $a \land c \in \{0, c\}$ or $b \land c \in \{0, c\}$, let $a_c = a \land c$. Otherwise, $a \land c$ and $b \land c$ are incomparable. Choose $a_c \leq c$ so that $a_c \land b_i \neq 0$ but $a_c \leq b$ for every $i < \omega$, this is possible because of Lemma 2.15 and the saturation of the monster model with respect to small sets. Thus, we have that $\text{tp}(a_c/Cb_i) = \text{tp}(a \land c/Cb)$ for every $i < \omega$. Again, by saturation, let $a'$ be such that $a' \land c = a_c$ for every $c \in \text{At}(\mathcal{C})$. By quantifier elimination in ABA we have that $\text{tp}(a'/Cb_i) = \text{tp}(a/Cb)$ and $(b_i)_{i < \omega}$ is indiscernible over $C a'$.

We then saw that under the assumptions $\mathcal{C} \preceq^* \mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ atomic, the relation $A \downarrow \mathcal{C} B$ is reducible to dividing in ABA. At this point one may wonder: why ABA? Can we reduce this relation to dividing in another completion of BA? The first theory that comes to mind is the theory of infinite atomic Boolean algebras (ATBA), which is also complete. The following remark shows that the reduction fails when dividing is considered in ATBA.

Remark 2.17 Let $\mathcal{M}$ be the monster model of ATBA, $(c, d, e, f) \in \text{At}(\mathcal{M})^4$ injective, $a = c \lor d$, $b = c \lor e$, $\mathcal{A} = \{a\}$, $\mathcal{B} = \{b\}$ and $2 = \emptyset$. Then $A \downarrow_2 B$ because $a \land b = c$, $a \land \neg b = d$, $\neg a \land b = e$ and $f \leq \neg a \land \neg b$. On the other hand, $a \not\downarrow_2 b$. Let indeed $(p_i)_{i < \omega} \in (\text{At}(\mathcal{M}) - \{c, e\})^\omega$ injective and define $b_0 = b$ and $b_i = p_{2i} \lor p_{2i+1}$ for $0 < i < \omega$. Then $(b_i)_{i < \omega}$ is indiscernible over $\emptyset$. Let now $a' \in \mathcal{M}$ with $\text{tp}(a'/b) = \text{tp}(a/b)$, then $a' \land b \neq 0$ and $a' = q_0 \lor q_1$ for $q_0, q_1 \in \text{At}(\mathcal{M})$. Let now $0 < i < \omega$ be such that $p_{2i} \neq q_0, q_1$ and $p_{2i+1} \neq q_0, q_1$, then $a' \land b_i = 0$. Thus $(b_i)_{i < \omega}$ is not indiscernible over $a'$.

2.2 Reduction

In this section we first introduce the notion of algebra generated by a tuple of functions and study some of its elementary properties. We then define an independence relation between tuples of functions, which is just an abstraction of the way independence is defined in database theory. Finally, we show that this form of independence is reducible to the independence relation between Boolean algebras that we dealt with in the previous sections.

As usual, by field of sets on $M$ we mean a Boolean algebra of subsets of $M$.扩容
**Definition 2.18** Let $M$ be a set and $\mathcal{F}$ a field of sets on $M$. We define the $n$-th product of $\mathcal{F}$, in symbols $(\mathcal{F})^n$, to be the field of sets on $M^n$ generated by the family of sets $\{\prod_{i<n} A_i \mid (A_i)_{i<n} \in \mathcal{F}^n\}$.

**Proposition 2.19** Let $M$ be a set and $\mathcal{F}$ a field of sets on $M$. Then

$$(\mathcal{F})^n = \left\{ \bigcup_{i<k} \left( \prod_{j<n} A_{i,j} \right) \mid (A_{i,j})_{i<k,j<n} \in \mathcal{F}^{nk}, k < \omega \right\}.$$ 

As known, given two sets $I$ and $M$, a field of sets on $M$ and a function $f : I \to M$, the family of sets $\{f^{-1}(A) \mid A \in \mathcal{F}\}$ is a field of sets on $I$. In the following, given a tuple of functions $(f_i)_{i<n}$ from $I$ to $M$, we will identify the tuple $(f_i)_{i<n}$ with the function $f : I \to M^n$ such that $f(j) = (f_0(j), \ldots, f_{n-1}(j))$ for every $j \in I$.

**Definition 2.20** Let $f = (f_i)_{i<n}$ be a tuple of functions from $I$ to $M$, and $\mathcal{F}$ the finite-cofinite field of sets on $M$. We let the Boolean algebra generated by $f$, in symbols $\pi(f)$, to be the field of sets $\{f^{-1}(A) \mid A \in (\mathcal{F})^n\}$.

**Proposition 2.21** Let $f = (f_i)_{i<n}$ be a tuple of functions from $I$ to $M$. Then

(i) $\pi(f) = (\pi(f_0), \ldots, \pi(f_{n-1})) = \langle f_i^{-1}(a) \mid i < n, a \in \text{ran}(f_i) \rangle$;

(ii) $\text{At}(f) = \{f^{-1}(a) \mid a \in \text{ran}(f)\}$;

(iii) $\pi(f)$ is atomic.

**Lemma 2.22** Let $f = (f_i)_{j<n}$ be a tuple of functions from $I$ to $M$. Then $\pi(f)$ is a regular subalgebra of $\mathcal{P}(I)$.

**Proof** Suppose otherwise, then we can find $(C_t)_{t \in T} \in \pi(f)^T$ such that

$$(\bigvee_{t \in T} C_t)_{\pi(f)} = I \neq \left( \bigvee_{t \in T} C_t \right)^{\mathcal{P}(I)} = \bigcup_{t \in T} C_t,$$

and so there exists $i \in I$ such that $i \notin C_t$ for every $t \in T$. Let $C = f^{(-1)}(f(i))$. If there were $t \in T$ such that $C \subseteq C_t$, then $i \in f^{(-1)}(f(i)) = C \subseteq \bigcup_{t \in T} C_t$, a contradiction. Thus, $C \cap C_t = \emptyset$ for every $t \in T$, because $C$ is an atom. But then $C_t \subseteq C^0$ for every $t \in T$ and $C^0 \neq I$ because $C \neq \emptyset$, a contradiction.

**Corollary 2.23** Let $f = (f_j)_{j<n}$ and $h = (h_j)_{j<k}$ be tuples of functions from $I$ to $M$. Then

(i) $\pi(h)$ is a regular subalgebra of $\pi(fh)$;

(ii) In both $\pi(h)$ and $\pi(fh)$ sups are unions.

**Proof** Immediate from the coherence of the regular subalgebra relations, i.e. if $C \subseteq B$, $B$ is regular in $A$ and $C$ is regular in $A$, then $C$ is regular in $B$.

**Lemma 2.24** Let $f = (f_j)_{j<n}$ and $h = (h_j)_{j<k}$ be tuples of functions from $I$ to $M$. Then $\pi(h) \leq^* \pi(fh)$.

© Springer
Proof Let $(A_i)_{i \in I} \in \text{At}(\pi(fh))^I$ and suppose that $\bigcup_{i \in I} A_i = A$ exists in $\pi(fh)$. Let $C_i = \pi(h) \supseteq A_i \uparrow$, we want to show that $\bigcup_{i \in I} C_i \in \pi(h)$. Given that $A \in \pi(fh)$, there are $(D_{i,j})_{i \prec t \ j} \in \mathcal{F}^{mt}$ and $(E_{i,j})_{i \prec t \ j} \in \mathcal{F}^{kl}$ such that

$$A = \bigcup_{i \prec t} (fh)^{(-1)} \left( \prod_{j \prec n} D_{i,j} \times \prod_{j \prec k} E_{i,j} \right) = (fh)^{(-1)} \left( \bigcup_{i \prec t} \left( \prod_{j \prec n} D_{i,j} \times \prod_{j \prec k} E_{i,j} \right) \right).$$

But then

$$\bigcup_{i \in I} C_i = \bigcup_{i \prec t} (h^{(-1)} \left( \prod_{j \prec k} E_{i,j} \right)) \in \pi(h).$$

\[\square\]

Definition 2.25 Let $f = (f_i)_{i < n}$, $g = (g_i)_{i < m}$ and $h = (h_i)_{i < k}$ be tuples of functions from $I$ to $M$. We define

$$\begin{array}{c}
f \downarrow^h g \\
\forall p, q \in I \ (h(p) = h(q) \Rightarrow \exists t \in I \ (h(t) = h(p) \& f(t) = f(p) \& g(t) = g(q))).
\end{array}$$

As we will see in Sect. 3, the following theorem establishes a reduction of the form of independence at play in database theory and independence logic to dividing in ABA.

Theorem 2.26 Let $f = (f_i)_{i < n}$, $g = (g_i)_{i < m}$ and $h = (h_i)_{i < k}$ be tuples of functions from $I$ to $M$. Then

$$\begin{array}{c}
f \downarrow^h g \\
\pi(fh) \downarrow^\pi(h) \pi(gh).
\end{array}$$

Proof By Corollary 2.13 this is straightforward.

3 Independence in database theory and team semantics

3.1 Database independence

Let Var be a countable set of symbols, called attributes or individual variables. A relation schema is a finite set $R = \{x_0, \ldots, x_{n-1}\}$ of attributes from Var. Each attribute $x_i$ of a relation schema is associated with a domain $\text{dom}(x_i)$ which represents the set of possible values that can occur as values of $x_i$. A tuple over $R$ is a function $t : R \to \bigcup_{i < n} \text{dom}(x_i)$ with $t(x_i) \in \text{dom}(x_i)$, for all $i < n$. A database\footnote{In the context of database theory a database is usually taken to be a finite set of tuples over a relation schema $R$. In our framework the assumption of finiteness does not play any role and so we drop it.} $r$ over $R$
is a set of tuples over \( R \). For \( x \subseteq R \) and \( t \in r \) we let \( t(x) \) to be the restriction of the function \( t \) to \( x \).

**Definition 3.1 (Functional dependence [1])** Let \( R \) be a relation schema, \( x \) and \( y \) tuples of attributes from \( R \), and \( r \) a database over \( R \). We define

\[
\text{\( r \) satisfies } x \rightarrow y \iff \forall t_0, t_1 \in r (t_0(x) = t_1(x) \Rightarrow t_0(y) = t_1(y)).
\]

If \( r \) satisfies \( x \rightarrow y \) we say that \( r \) manifests the functional dependency \( x \rightarrow y \).

**Definition 3.2 (Database independence [19])** Let \( R \) be a relation schema, \( x, y \) and \( z \) tuples of attributes from \( R \), and \( r \) a database over \( R \). We define

\[
\text{\( r \) satisfies } z \rightarrow x | y \iff \forall t_0, t_1 \in r (t_0(z) = t_1(z) \Rightarrow \exists t_2 \in r (t_2(z) = t_0(z) \land t_2(x) = t_0(x) \land t_2(y) = t_1(y))).
\]

If \( r \) satisfies \( z \rightarrow x | y \) we say that \( r \) manifests the database independence \( z \rightarrow x | y \).

In the database theory literature the term embedded multivalued dependence is preferred to the simpler database independence. The reason for this choice of terminology is that embedded multivalued dependence is a generalization of functional dependence, as the following proposition shows.

**Proposition 3.3** Let \( R \) be a relation schema, \( x \) and \( y \) tuples of attributes from \( R \), and \( r \) a database over \( R \). Then

\[
\text{\( r \) satisfies } x \rightarrow y \iff \text{\( r \) satisfies } x \rightarrow y | y
\]

**Proof** Immediate.

Let \( R \) be a relation schema, \( x \) a tuple of attributes from \( R \) and \( r \) a database over \( R \). Let also \( r = (t_i)_{i \in I} \) and \( M = \bigcup_{i < \omega} \text{dom}(x_i) \). We can then define a function \( \hat{x} : I \rightarrow M^n \) by letting \( \hat{x}(i) = t_i(x) \) for every \( i \in I \). Clearly, for \( x = (x_{j_0}, \ldots, x_{j_{m-1}}) \) and \( i \in I \), we have that \( \hat{x}(i) = (\hat{x}_{j_0}(i), \ldots, \hat{x}_{j_{m-1}}(i)) \), and so we can identify the objects \( \hat{x} \) and \( (\hat{x}_{j_0}, \ldots, \hat{x}_{j_{m-1}}) \). Following the notation of Sect. 2.2, we can then consider the Boolean algebra \( \pi(\hat{x}) \) (as a subalgebra of \( \mathcal{P}(I) \)). Notice now that being the theory ABA the model completion of the theory BA, there is an embedding \( i : \pi(\hat{x}) \rightarrow \mathfrak{M} \), where as usual we denote by \( \mathfrak{M} \) the monster model of ABA. Thus \( \pi(\hat{x}) \cong i(\pi(\hat{x})) \), and so, modulo isomorphism, the Boolean algebra \( \pi(\hat{x}) \) can be thought as living in \( \mathfrak{M} \). This little argument allows us to formulate in exact terms a reduction of database independence to dividing in ABA.

---

1 In standard references in database theory (among which [19]) it is usually assumed that \( x \cap y \subseteq z \). We relax this assumption because it comes at no conceptual cost and simplifies the treatment.
**Theorem 3.4** Let $R$ be a relation schema, $x$, $y$ and $z$ tuples of attributes from $R$, and $r$ a database over $R$. Then

\[ r \text{ satisfies } z \rightarrow x \iff \pi(x) \mid \pi(y) \iff \forall a \in \text{At} \left( \pi(x) \right) \text{ and } b \in \text{At} \left( \pi(y) \right) \begin{bmatrix} a & d & b \end{bmatrix}. \]

**Proof** An immediate consequence of Theorems 2.26 and 2.16.

As we saw above, in the case of ABA the choice of embeddings with respect to which we think of the algebras arising from a database as living in the monster model for ABA does not matter, any choice of embeddings would do. In the case of ATBA this is not the case. Remark 2.17 shows in fact that in the case of ATBA the choice of embeddings does matter. For this reason, ABA is a more natural context than ATBA for our reduction to take place.

### 3.2 Independence in team semantics

We now introduce what is known as independence logic [7]. The semantics of this logic is formulated using sets of assignments, also called *teams*, instead of single assignments. This new way of defining semantics for logical languages was introduced by Hodges in [8] and then developed by Väänänen in [22]. The salient characteristic of independence logic is that it contains syntactic expressions modelling dependence and independence, called *dependence and independence atoms*, respectively. On top of this, the language of independence logic allows relational atomic formulas, Boolean connectives and quantifiers, as in classical first-order logic. We focus here only on (in)dependence atoms and consequently give the semantics only for this fragment of the language of this logic. The truth definition can then be extended from independence atoms to any formula in the language of independence logic in a canonical way. The resulting logic has a non-classical flavour and possesses the expressive power of existential second-order logic. The exact definitions are beyond the scope of this paper and will not be presented here, for details see Väänänen [22].

Let $M$ be a first-order structure and $V \subseteq \text{Var}$ a finite set of variables. An assignment $s$ on $M$ with domain $\text{dom}(s) = V$ is a mapping from $V$ to $M$. A team $X$ on $M$ with domain $\text{dom}(X) = V$ is a set of assignments with domain $V$. Let $L$ be a fixed but arbitrary signature. The dependence and independence atoms in the signature $L$ are syntactic expressions of the form $=(u, v)$ and $u \downarrow \downarrow w v$, respectively, where $u, v, w \in (\text{Term}_L)^{<\omega}$. Given $u \in (\text{Term}_L)^{<\omega}$, we denote by $\text{Var}(u)$ the set of variables occurring in at least one of the terms in the tuple $u$.

**Definition 3.5** Let $\mathcal{M}$ be an $L$-structure, $X$ a team on $\mathcal{M}$ and $\text{Var}(u), \text{Var}(v) \subseteq \text{dom}(X)$. We let

\[ \mathcal{M} \models_X (u, v) \iff \forall s, s' \in X (s(u) = s'(u) \Rightarrow s(v) = s'(v)). \]
Definition 3.6 Let $\mathcal{M}$ be an $L$-structure, $X$ a team on $\mathcal{M}$ and $\text{Var}(u), \text{Var}(v)$ and $\text{Var}(w) \subseteq \text{dom}(X)$. We let

$$\mathcal{M} \models_X u \perp_w v$$

if

$$\forall s, s' \in X (s(w) = s'(w) \Rightarrow \exists s'' \in X (s''(w) = s(w) \& s''(u) = s(u) \& s''(v) = s'(v))).$$

In analogy with Proposition 3.3, we have the following. It shows that our syntax could have been chosen without dependence atoms.

Proposition 3.7 Let $\mathcal{M}$ be a first-order structure, $X$ a team on $\mathcal{M}$ and $\text{Var}(u), \text{Var}(v) \subseteq \text{dom}(X)$. Then

$$\mathcal{M} \models_X (u, v) \iff \mathcal{M} \models_X v \perp_u u.$$ 

Given a non-empty set $M$, we denote by $(M) = M$ the structure in the empty signature with domain $M$. Clearly, given $R = \{x_0, \ldots, x_{n-1}\} \subseteq \text{Var}$, each database $r$ over $R$ can be seen as a team $X_r$ on $\bigcup_{i<n} \text{dom}(x_i)$ with domain $R$, and each team $X$ on a substructure of $\bigcup_{i<n} \text{dom}(x_i)$ with domain $R$ can be seen as a database $r_X$ over $R$.

Remark 3.8 Let $M$ be a non-empty set, $X$ a team on $M$, $x, y$ and $z \subseteq \text{dom}(X) \subseteq \text{Var}$. Then

$$M \models_X x \perp_z y \iff r_X \text{ satisfies } z \rightarrow x | y.$$ 

As Remark 3.8 shows and any researcher in dependence logic knows, the form of independence at play in team semantics is exactly database independence. Thus, as the latter case of independence is reducible to diving in ABA, so is the former.

As made clear by our exposition of the subject, independence logic is an extension of first-order logic which is based on a database-oriented notion of team, and a particular form of independence, i.e. database independence. In light of the analogies between this form of independence and stochastic independence noted above, it seems plausible to formulate a version of independence logic which admits probabilities and it is able to deal with stochastic independence. This is done by the authors in [15] and [13], where applications of these ideas to quantum logic are also developed.

References

1. Armstrong, W.W.: Dependency structures of data base relationships. In: IFIP Congress, pp. 580–583 (1974)
2. Avils, A., Brech, C.: A Boolean algebra and a banach space obtained by push-out iteration. Topol. Appl. 158(13), 1534–1550 (2011)
3. Baldwin, J.T., Eklof, P.C., Trlifaj, J.: $\perp_N$ as an abstract elementary class. Ann. Pure Appl. Log. 149(1), 25–39 (2007)
4. Ben Yaacov, I.: On theories of random variables. Isr. J. Math. 194(2), 957–1012 (2013)
5. Ben Yaacov, I., Usvyatsov, A.: Continuous first order logic and local stability. Trans. Am. Math. Soc. 362(10), 5213–5259 (2010)
6. Givant, S., Halmos, P.: Introduction to Boolean Algebras. Springer, Berlin (2009)
7. Grädel, E., Väänänen, J.: Dependence, independence, and incomplete information. Stud. Log. 101(2), 399–410 (2013)
8. Hodges, W.: Compositional semantics for a logic of imperfect information. Log. J. IGPL 5, 539–563 (1997)
9. Hirvonen, Å.: Independence in model theory. In: Abramsky, S., Kontinen, J., Vollmer, H., Väänänen, J. (eds.) Dependence Logic: Theory and Applications, Springer, New York (2016)
10. Hirvonen, Å., Hyttinen, T.: Categoricity in homogeneous complete metric spaces. Arch. Math. Log. 48, 269–322 (2009)
11. Hirvonen, Å., Hyttinen, T.: Measuring dependence in metric abstract elementary classes with perturbations (submitted)
12. Hyttinen, T., Kangas, K.: On model theory of covers of algebraically closed fields. Ann. Acad. Sci. Fenn. Math. 40(2), 507–533 (2015)
13. Hyttinen, T., Paolini, G., Väänänen, J.: A logic for arguing about probabilities in measure teams (submitted)
14. Hyttinen, T., Paolini, G.: Beyond abstract elementary classes: on the model theory of geometric lattices (submitted)
15. Hyttinen, T., Paolini, G., Väänänen, J.: Quantum team logic and Bell’s inequalities. Rev. Symb. Log. 08(04), 722–742 (2015)
16. Marker, D.: Introduction to Model Theory. Springer, New York (2002)
17. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference Morgan Kaufman, San Mateo CA (1988)
18. Poizat, B.: A Course in Model Theory. Springer, Berlin (2000)
19. Sagiv, Y., Walecka, S.F.: Subset dependencies and a completeness result for a subclass of embedded multivalued dependencies. J. ACM 29(1), 103–117 (1982)
20. Shelah, S.: Classification Theory: and the Number of Non-isomorphic Models. Elsevier, Amsterdam (1990)
21. Studeny, M.: Conditional independence relations have no finite complete characterization. In: Transactions of the 11th Prague Conference on Information Theory. Kluwer, pp. 377-396 (1992)
22. Väänänen, J.: Dependence Logic, Volume 70 of London Mathematical Society Student Texts. Cambridge University Press, Cambridge (2007)