

Department of Mathematics and Statistics
Faculty of Science
University of Helsinki
2016

INDEPENDENCE IN MODEL THEORY AND TEAM SEMANTICS

GIANLUCA PAOLINI

ACADEMIC DISSERTATION

Date and place: Kumpula Campus, auditorium CK112 on Tuesday, June 28th, 2016.

Opponent: Wilfrid Hodges.

Custos: Jouko Väänänen.

ISBN 978-951-51-2244-5 (paperpack)

ISBN 978-951-51-2245-2 (PDF)

Unigrafia Oy

Helsinki 2016

A nonno Giovanni, maestro di vita.

ACKNOWLEDGMENTS.

First of all, I would like to thank my supervisors Jouko Väänänen and Tapani Hyttinen. The first for having taught me that logic is a whole and has no borders, and for having given me the opportunity to be where I am now. The second for having taught me the art of mathematical thinking, through hours and hours of intense discussions, for his support, and for his honesty. Secondly, I would like to thank the pre-examiners Professors John Baldwin and Heribert Vollmer, for their careful reading of my thesis. Thirdly, I would like to thank the CIMO and most of all the Väisälä Foundation, for their financial support during my studies. Finally, I thank my family and friends at home, for giving me stability and love.

Gianluca Paolini

INTRODUCTION

GIANLUCA PAOLINI

1. LIST OF PAPERS

This thesis consists of the following seven articles, referred to in the text by their Roman numerals. The papers are reproduced with the permission of their respective copyright holders.

- I Gianluca Paolini and Jouko Väänänen. *Dependence Logic in Pregeometries and ω -Stable Theories*. Journal of Symbolic Logic, 81(01):32-55, 2016.
- II Gianluca Paolini. *Independence Logic and Abstract Independence Relations*. Mathematical Logic Quarterly, 61(03):202-216, 2015.
- III Tapani Hyttinen and Gianluca Paolini. *Reduction of Database Independence to Dividing in Atomless Boolean Algebras*. Archive for Mathematical Logic, 55(03):505-518, 2016.
- IV Tapani Hyttinen and Gianluca Paolini. *Beyond Abstract Elementary Classes: On The Model Theory of Geometric Lattices*. Submitted, December 2015.
- V Gianluca Paolini. *A Finite Axiomatization of G-Dependence*. Submitted, December 2015.
- VI Tapani Hyttinen, Gianluca Paolini and Jouko Väänänen. *Quantum Team Logic and Bell's Inequalities*. Review of Symbolic Logic, 08(04):722-742, 2015.
- VII Tapani Hyttinen, Gianluca Paolini and Jouko Väänänen. *A Logic for Arguing About Probabilities in Measure Teams*. Submitted, September 2015.

Some words about contributions. Paper I is entirely based on my Master Thesis, written under the supervision of Väänänen. The topic and open questions were introduced to me by Väänänen. Most of the work is due to me. The introduction was written by him. Theorem 3.13 of Paper II is due to Hyttinen. Paper III is the answer to a question posed by me to Hyttinen. The main idea (the reduction to dividing) is due to Hyttinen, but all the details have been elaborated by me. Paper IV is the paper with the most intricate history. The study of geometric lattices and of the relative construction of principal extension arises from my attempts to work on a problem in geometric model theory. The idea to frame this work under the perspective of abstract elementary classes is due to Hyttinen. Most of the technical work has been done by me, but the help of Hyttinen was fundamental. Section 6 is mostly due to Hyttinen, a part from the details of some of the technical lemmas, such as 6.2-6.9. Paper V is the solution of an open question posed by Väänänen. The main idea behind paper VI (quantum teams) is due to me, and so are the connections with the work of Abramsky and the choice of syntax and deductive system. The completeness theorem is due to Hyttinen. The details of the proof have been elaborated by me. The final version of the paper is due to Väänänen, who improved it substantially. Also, the introduction was written by him. Finally, paper VII is the result of a long process of formulation of a probabilistic version of dependence logic. The notion of measure team is the outcome of several discussions

on the matter between Hyttinen, Väänänen and me. The choice of syntax and the completeness theorem are due to Hyttinen. The details of the proof have been elaborated by me. Section 4 (the examples) is due to Hyttinen-Väänänen, a part from Example 4.5, which is due to me. The introduction is due to Väänänen. All the papers have been written by me. In all the paper but I, II and V, the guidance of Hyttinen was essential.

2. INTRODUCTION

The subject of this doctoral thesis is the mathematical theory of independence, and its various manifestations in logic and mathematics. The topics covered in this doctoral thesis range from model theory and combinatorial geometry, to database theory, quantum logic and probability logic. This study has two intertwined centres:

- (1) classification theory, independence calculi and combinatorial geometry (papers I-IV);
- (2) new perspectives in team semantics (papers V-VII).

The first topic is a classical topic in model theory, which we approach from different directions (implication problems, abstract elementary classes, unstable first-order theories). The second topic is a relatively new logical framework where to study non-classical logical phenomena (dependence and independence, uncertainty, probabilistic reasoning, quantum foundations). Although these two centres seem to be far apart, we will see that they are linked to each others in various ways, under the guiding thread of *independence*.

3. CLASSIFICATION THEORY, INDEPENDENCE CALCULI AND COMBINATORIAL GEOMETRY

The post-Morley era of model theory is dominated by what is known as *classification theory*. In a few words, classification theory aims at the determination of structural properties of a class of structures, e.g. the association of cardinal invariants, in terms of logical properties of this class, e.g. the number of types or the categoricity in a certain power. The research questions in this area lead to the elaboration of a highly sophisticated theory of independence: *the theory of independence calculi*. This theory (and its associated notions) is now one of core fields of research in model theory, and, despite its generality, it has had applications in many areas of mathematics, such as algebra, combinatorics and geometry.

The theme of model-theoretic independence is the center of the first part of the thesis. In Articles I-IV we treat this theme from several perspectives, stressing its relation with combinatorial geometry (Articles I and IV), and establishing a two-way connection between model-theoretic independence and database dependency theory (Articles I-III). The notion of database and the related field of database dependency theory represents also the link between the first part of the thesis and the second. For this reason we will treat this topic in a certain detail in this introduction.

We now give an overview of the articles forming the first part of the thesis, and then introduce the reader to the articles one by one, stressing the connections between them. In Article I we analyse the notions of independence coming from combinatorial geometry and first-order model theory from the perspective of database dependency theory, specifically the study of *implication problems*. In Article II we generalise the results of Article I to the context of *abstract elementary classes*

and abstract independence relations. In Article III we analyse one of the most important database dependencies, called *embedded multivalued dependence* or simply *database independence*, from the perspective of model theory, showing that this notion of independence is reducible to the first-order dividing calculus in the theory of atomless boolean algebras. Article IV is by far the most technical and elaborated paper of the whole thesis. In this article we develop the model theory of *geometric lattices* (of finite rank) from the perspective of abstract elementary classes. Based on a fundamental combinatorial result due to Crapo, we find various classes of geometric lattices that behave very well from the point of view of stability theory. One of them, (\mathbf{K}^3, \preceq) , is ω -stable, it has a monster model and an independence calculus that satisfies all the usual properties of non-forking. On the other hand, the class (\mathbf{K}^3, \preceq) is not an abstract elementary class, in fact the Smoothness Axiom fails.

3.1. Implication Problems in Model Theory

In this section we will describe the work of Articles I and II. We now introduce the setting of database dependency theory. We first define what is a database.

Definition. Let Var be a countable set of symbols, called *attributes* or *individual variables*. A *relation schema* is a finite set $R = \{x_0, \dots, x_{n-1}\}$ 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 $t : R \rightarrow \bigcup_{i < n} \text{dom}(x_i)$ with $t(x_i) \in \text{dom}(x_i)$, for all $i < n$. A *database* 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 .

Given a database r it is often of crucial importance to know which *constraints* the database r is subject to, in order to deduce further information about it. For this end, one introduces so-called *database dependencies*. To give an example, we introduce what is probably the most well-known database dependency: *functional dependence*.

Definition (Functional dependence [2]). Let R be a relation schema, x and y tuples of attributes from R , and r a database over R . We define

$$r \text{ satisfies } x \rightarrow y \Leftrightarrow \forall t_0, t_1 \in r (t_0(x) = t_1(x) \Rightarrow t_0(y) = t_1(y)).$$

If r satisfies $x \rightarrow y$ we say that r manifests the *functional dependency* $x \rightarrow y$.

One of the main goals of database dependency theory is to axiomatise given forms of dependence D , i.e. finding a set of rules that govern D completely. These kind of problems are also called *implication problems*, since the question is to find a set of rules that determines completely when a given dependency σ is implied by a set of dependencies Σ . The canonical example of this line of research is Armstrong's axiomatization of functional dependence [2].

This way of approaching the study of a given form of dependence can be exported from database dependency theory to any well-defined notion of dependence. The object of Article I is the study of implication problems with respect to the notions of dependence arising from *combinatorial geometry* (cfr. e.g. [7]) and the *first-order forking calculus* (cfr. e.g. [22]).

When we refer to a notion of dependence D we are intentionally vague, in fact we want to subsume as many cases as possible. This may lead to certain confusions, as among the accepted forms of dependence D there are notions that may more

meaningfully be referred to as *independencies*. This is the case for the well-known *embedded multivalued dependence* (cfr. e.g. [21]), which we will refer to simply as *database independence*. As already mentioned, this form of independence is also the topic of Article III.

Definition (Database independence [21]). Let R be a relation schema, x, y and z tuples of attributes from R , and r a database over R . We define

$$r \text{ satisfies } z \twoheadrightarrow x | y$$

$$\Leftrightarrow \forall t_0, t_1 \in r(t_0(z) = t_1(z) \Rightarrow \exists t_2 \in r(t_2(z) = t_0(z) \& t_2(x) = t_0(x) \& t_2(y) = t_1(y))).$$

If r satisfies $z \twoheadrightarrow x | y$ we say that r manifests the *database independency* $z \twoheadrightarrow x | y$.

Notice that database independence generalizes functional dependence, as r satisfies $x \rightarrow y$ iff r satisfies $x \twoheadrightarrow y | y$. We refer to expressions of the form $z \twoheadrightarrow x | y$ as *conditional independence atoms*, while to expressions of the form $\emptyset \twoheadrightarrow x | y$ simply as *independence atoms*. Parker and Parsaye-Ghomi [18] proved that it is *not possible* to find a *finite* complete axiomatization for the conditional independence atoms. Furthermore, in [15] and [16] Hermann proved that the implication problem for these atoms is *undecidable*. On the other hand, the independence atom admits a *very natural finite axiomatization* (see [12] and [11]). The main question of Article I and II is the Completeness Question for independence atoms under semantics based on combinatorial geometry and the first-order forking calculus. That is, when is the deductive system that axiomatizes database independence complete with respect to these different semantics? The way the new semantics are given is straightforward. Given a pregeometry (M, cl) and an assignment s from the set of variables Var into M we say that $(M, \text{cl}) \models \emptyset \twoheadrightarrow x | y$ if $s(x)$ is independent from $s(y)$ in the sense of combinatorial geometry. Similarly, given an ω -stable first-order theory T , a model \mathcal{M} of T and an assignment s from Var into M we say that $\mathcal{M} \models \emptyset \twoheadrightarrow x | y$ if $s(x)$ is independent from $s(y)$ in the sense of the first-order forking calculus. The main results of Article I are the following.

Theorem. Suppose \mathbf{C} is a class of pregeometries including a pregeometry (M, cl) satisfying (P1)-(P3)¹. Then the Completeness Question for \mathbf{C} has a positive answer, i.e. if Σ is a set of independence atoms and ϕ is an independence atom, then

$$\Sigma \vdash_{\mathcal{I}} \phi \iff \Sigma \models_{\mathbf{C}} \phi.$$

Theorem. Let T be an ω -stable first-order theory T such that there exists $\mathcal{M} \models T$ satisfying the following requirement: there exists a strongly minimal set $F = \phi(\mathcal{M}, \emptyset)$ such that (F, cl) has properties (P1), (P2) and (P3). Then the Completeness Question for T has a positive answer, i.e. if Σ is a set of independence atoms and ϕ is an independence atom, then

$$\Sigma \vdash_{\mathcal{I}} \phi \iff \Sigma \models_T \phi.$$

Article I had left several questions open. On the one hand, there was the Completeness Question for first-order theories beyond ω -stability and for classes of structures non-axiomatizable in first-order logic. On the other hand, there was the question of when *exactly* the Completeness Question had a positive answer. In

¹Conditions (P1)-(P3) are three natural conditions. They are satisfied in the case of vector spaces and algebraically closed fields, see Article I for details.

Article II we answer both questions. Before introducing the statement of the main theorem of Article II, we introduce the setting of abstract elementary classes [23] and abstract independence relations, which will be relevant also in Article IV.

The framework of abstract elementary classes is a very general platform where to study model-theoretic phenomena that go beyond first-order logic. An abstract elementary class is a pair (\mathbf{K}, \preceq) such that \mathbf{K} is a class of structures of the same similarity type and \preceq is a binary relation satisfying a certain set of axioms, which generalise some of the properties satisfied by the relation of elementary submodel of first-order logic. In its full generality the theory is very abstract and, usually, in order to develop the classification theory for abstract elementary classes further assumptions are made on (\mathbf{K}, \preceq) , e.g. finitary, type-short, tame, etc. Under further assumptions, concrete independence calculi have been defined for an abstract elementary class (\mathbf{K}, \preceq) , on the style of the forking calculus of first-order logic, but there is at the moment no general theory of independence for an arbitrary abstract elementary class.

In the context of Article II we work with an arbitrary abstract elementary class (\mathbf{K}, \preceq) admitting a monster model, and we define an abstract notion of (pre)-independence on (\mathbf{K}, \preceq) , denoted as \perp . The resulting setting is consequently very general. Interestingly, it is exactly this high level of abstraction that allowed us to realise which were the conditions sufficient and necessary to prove a completeness theorem with respect to the deductive system arising from database theory, i.e. what we call federation (following [3]) and admissibility of an algebraic point.

Theorem. Let (\mathbf{K}, \preceq) be an abstract elementary class with amalgamation property, joint-embedding property and arbitrary large models. Let also \perp be a pre-independence relation for (\mathbf{K}, \preceq) . Then

$$\Sigma \vdash_{\mathcal{I}} \phi \iff \Sigma \models_{(\mathbf{K}, \perp)} \phi$$

if and only if \perp is federated and admits an algebraic point.

3.2. Database Independence and Boolean Algebras

The theory of independence calculi is the result of a far-reaching process of generalisation of well-known forms of independence such as linear independence (vector spaces) and algebraic independence (algebraically closed fields). The development of so-called applied model theory has further shown that many other naturally arising forms of independence can be subsumed by the general theory of independence as formulated in model theory. Among these forms of independence there is a notion of independence which is crucial in probability theory: *stochastic independence*. In [6] Itai 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, as well-known, corresponds to independence of random variables, i.e. stochastic independence.

In various papers on statistics and database theory, e.g. [19], [21] and [24], stochastic independence has been put in tight connection with database independence. 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.

The motivating question of Article III is then the following: how does database independence relate to the theory of independence calculi? 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?

The answers to these questions go through the identification of database independence with an independence notion between boolean algebras, and the further reduction of this notion to the dividing calculus in the theory of atomless boolean algebras (ABA). This reduction shows that both 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.

To this extent, in Article III we first prove a partial characterization of dividing in ABA, based on the above-mentioned notion of free amalgamation of boolean algebras. Secondly, given a database r over the relation schema R , and a tuple of attributes x from R , we associate a boolean algebra $\pi(\dot{x})$, and show that independence between tuples of attributes corresponds to free amalgamation between the respective boolean algebras (see Definition 2.9 of Article III for $\mathcal{C} \leq^* \mathcal{A}$).

Theorem. Let \mathfrak{M} be the monster model of ABA, and $\mathcal{A}, \mathcal{B}, \mathcal{C} \leq \mathfrak{M}$, with \mathcal{C} atomic and $\mathcal{C} \leq^* \mathcal{A}, \mathcal{B}$. Then

$$\mathcal{A} \downarrow_{\mathcal{C}} \mathcal{B} \Leftrightarrow \forall a \in \mathcal{A} \text{ and } b \in \mathcal{B} \left[a \downarrow_{\mathcal{C}}^d b \right].$$

Theorem. Let R be a relation schema, x, y and z tuples of attributes from R , and r a database over R . Then

$$\begin{aligned} r \text{ satisfies } z \twoheadrightarrow x \mid y &\Leftrightarrow \dot{x} \downarrow_{\dot{z}} \dot{y} \Leftrightarrow \pi(\dot{x}\dot{z}) \downarrow_{\pi(\dot{z})} \pi(\dot{y}\dot{z}) \\ &\Leftrightarrow \forall a \in \text{At}(\pi(\dot{x}\dot{z})) \text{ and } b \in \text{At}(\pi(\dot{y}\dot{z})) \left[a \downarrow_{\pi(\dot{z})}^d b \right]. \end{aligned}$$

3.3. Geometric Lattices

The interaction between combinatorial geometry and model theory dates back to the birth of classification theory, and it is now an established field of research known as *geometric model theory* (see e.g. [20]). In this line of research results from combinatorial geometry are used to establish structural results on certain well-behaved classes of structures, the canonical example being the class of models of a strongly minimal theory, as well-known. On the other hand, the analysis of combinatorial geometries using model-theoretic tools (as it is done for example in the study of algebraically closed fields) is an underdeveloped field of study. This fact, paired with the familiarity with combinatorial geometry coming from the work of Article I, lead us to the attempt of delineating a model theory of combinatorial geometries from the point of view of abstract elementary classes (cfr. Section 3.1). This attempt has proven to be a fruitful one, in fact we were able to find various classes of geometries that, although rather unusual, are very well-behaved from the point of view of classification theory.

Combinatorial geometries are usually defined as sets with a closure operator subject to certain specific requirements (most notably the satisfaction of the Exchange Axiom). Under the assumption of finite dimensionality, there are several equivalent characterizations of these mathematical objects, e.g. via independence sets, closed sets, circuits, etc. Among these, there is the perspective of *geometric*

lattices (cfr. [7]). Geometric lattices are classical ordered algebraic structures, and thus, unlike sets with a closure operator, they fit perfectly with the model-theoretic perspective. Consequently, we develop our analysis of combinatorial geometries from the algebraic perspective of geometric lattices, under the assumption of finite dimensionality.

The motivating question of Article IV is the search for *good* classes of geometric lattices with respect to the framework of abstract elementary classes. For a class to be good our minimum requirements are that the class has a monster model and that it is stable. The most obvious classes fail to satisfy these requirements. Specifically, the class \mathbf{K}_0^n of geometric lattices of a fixed rank $n \geq 3$ with the submodel relation \preceq_L in the vocabulary $L = \{0, 1, \vee, \wedge\}$ as the strong submodel relation, yields an abstract elementary class with a monster model, but $(\mathbf{K}_0^n, \preceq_L)$ has the independence property, and so it is unstable.

Theorem. $(\mathbf{K}_0^3, \preceq_L)$ is an abstract elementary class with amalgamation property, joint embedding property and arbitrary large models². Furthermore, $(\mathbf{K}_0^3, \preceq_L)$ has the independence property.

Corollary. $(\mathbf{K}_0^3, \preceq_L)$ is unstable.

So one needs to look elsewhere to find good classes of geometric lattices. This is achieved by a certain use of a fundamental result on geometric lattices, known as Crapo's theorem on *one-point extensions of combinatorial geometries* [8]. Based on this result, we were able to find a class of geometric lattices, denoted as (\mathbf{K}^3, \preceq) , which, although stable and admitting an independence calculus, fails one of the crucial axioms of abstract elementary classes, namely the Smoothness Axiom, making this class a *unique phenomenon* in the model-theoretic universe.

Theorem. (1) (\mathbf{K}^3, \preceq) satisfies all the axioms of abstract elementary classes but the Smoothness Axiom.

(2) (\mathbf{K}^3, \preceq) has amalgamation property, joint embedding property and arbitrary large models.

(3) (\mathbf{K}^3, \preceq) is stable in every infinite cardinality.

(4) (\mathbf{K}^3, \preceq) admits an independence calculus \downarrow .

The reason behind the failure of smoothness is that the relation \preceq in (\mathbf{K}^3, \preceq) is based on a well-foundedness notion of construction, which resembles the notion of constructible sets for (abstract) isolation notions of classification theory (see e.g. [9]). In the last part of Article IV, we show that, with some non-trivial work, lifting the well-foundedness requirement from \preceq , and further loosening up this relation so as to impose coherence, we *do get* an abstract elementary class (denoted as $(\mathbf{K}_+^3, \preceq^+)$), but $(\mathbf{K}_+^3, \preceq^+)$ may not be ω -stable anymore. However, $(\mathbf{K}_+^3, \preceq^+)$ remains at least stable.

Theorem. $(\mathbf{K}_+^3, \preceq^+)$ is an abstract elementary class with amalgamation property, joint embedding property and arbitrary large models. Furthermore, $(\mathbf{K}_+^3, \preceq^+)$ is stable in every cardinal κ such that $\kappa^\omega = \kappa$.

4. NEW PERSPECTIVES IN TEAM SEMANTICS

We now come to the second part of the thesis, which is centred around the notion of *team* and generalisations thereof. The key link between this part of the thesis

²These three properties ensure the existence of a monster model for $(\mathbf{K}_0^3, \preceq_L)$.

and the first is represented by Articles I and III, in particular the latter. We will see in fact that the perspective of Article III leads team semantics from the world of database dependency theory to the world of *probability and quantum logic*. We will make this point clear soon, we first introduce the setting of team semantics, and explain the motivations that lie behind it.

By team semantics we mean the study of logical systems whose semantics are not based on single assignments to the propositional or individual variables (as in propositional and first-order logic), but with respect to *sets of assignments*. To avoid ambiguities we state the exact definition of a team, where with Var we denote the set of propositional or individual variables.

Definition (Team). A *team* X with values in A and domain $\text{dom}(X) \subseteq \text{Var}$ is a set of assignments from $\text{dom}(X)$ into A .

Team semantics was originally defined by Väänänen (following pioneering work of Hodges [17]) in order to give a purely logical treatment of the notions of dependence and independence. In [25] Väänänen builds an extension of first-order logic with an extended set of atomic formulas, known as *dependence atoms*, denoted as $=(x, y)$. He calls this logic *dependence logic*. The key idea behind the choice of team semantics for the semantics of this logic is that the concept of dependence can not be given meaning with respect to a specific state of affairs, but only with respect to a *multiplicity* of states of affairs, i.e. a team. The semantics of the dependence atom is defined as follows.

Definition. Let \mathcal{M} be an L -structure, X a team with values in M , and $x, y \subseteq \text{dom}(X)$. We let

$$\mathcal{M} \models_X =(x, y) \Leftrightarrow \forall s, s' \in X (s(x) = s'(x) \Rightarrow s(y) = s'(y)).$$

Evidently, given $R = \{x_0, \dots, x_{n-1}\} \subseteq \text{Var}$, each database r over R can be seen as a team X_r with values in $\bigcup_{i < n} \text{dom}(x_i)$ and domain R . On the other hand, each team X with values in $\bigcup_{i < n} \text{dom}(x_{k_i})$ and domain R can be seen as a database r_X over R . That is, a team is nothing but a database and a database is nothing but a team. Using this identification, we immediately see that the way we give meaning to the dependence atom is exactly as in the case of functional dependence (cfr. Section 3.1).

Remark. Let \mathcal{M} be a first-order structure, X a team with values in M , and $x, y \subseteq \text{dom}(X) \subseteq \text{Var}$. Then

$$\mathcal{M} \models_X =(x, y) \Leftrightarrow r_X \text{ satisfies } x \rightarrow y.$$

In [13] Väänänen and Grädel undertake an analysis of the notion of independence in line with the analysis of the notion of dependence undertaken in [25]. In this paper, they introduce a new atom, which they call *independence atom*, denoted as $x \perp_z y$. The semantics for this atom is given as follows.

Definition. Let \mathcal{M} be an L -structure, X a team with values in M , and $x, y, z \subseteq \text{dom}(X)$. We let

$$\begin{aligned} \mathcal{M} \models_X x \perp_z y \\ \Leftrightarrow \\ \forall s, s' \in X (s(z) = s'(z) \Rightarrow \exists s'' \in X (s''(z) = s(z) \& s''(x) = s(x) \& s''(y) = s'(y))). \end{aligned}$$

Also in the case of the independence atom we have a correspondence with a form of independence coming from database dependency theory, i.e. the *database independence* of Article III (cfr. Section 3.1).

Remark. Let \mathcal{M} be a first-order structure, X a team with values in M , and $x, y, z \subseteq \text{dom}(X) \subseteq \text{Var}$. Then

$$\mathcal{M} \models_X x \perp_z y \Leftrightarrow r_X \text{ satisfies } z \rightarrow x \mid y.$$

It is then clear that (in)dependence logic is based on a database-theoretic notion of (in)dependence, i.e. functional dependence and database independence. The crucial point in understanding the developments in team semantics explored in this thesis is how the work of Article III relates to this observation.

In Article III we saw that database independence can be reduced to the general theory of independence as developed in model theory, and specifically that this reduction establishes a parallelism between database independence and stochastic independence, since these two forms of independence can be seen as essentially based on the same independence relation.

The crucial technical detail in this reduction is to treat a database or a team not as a set of tuples, but as a *vector of functions*. Specifically, given a team X with values in M , and $x \subseteq \text{dom}(X)$ we let \dot{x} to be the function with domain X and range M such that $\dot{x}(s) = s(x)$, for every $s \in X$. Clearly, for $x = (x_{k_j})_{j \in J}$ and $s \in X$, we have that $\dot{x}(s) = (\dot{x}_{k_j}(s))_{j \in J}$, and so we can identify the objects \dot{x} and $(\dot{x}_{k_j})_{j \in J}$. That is, a team X becomes nothing but the “random” vector \dot{x} for $x = \text{dom}(X)$. Thought in these terms it is now natural to associate a boolean algebra to a finite tuple of variables $x = (x_0, \dots, x_{k-1})$, in the following sense (see Section 2.2 of Article III for details).

Definition. Let $f = (f_0, \dots, f_{k-1})$ 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 $\{f^{(-1)}(A) : A \in (\mathcal{F})^n\}$.

The point of Article III is then to show that the relation $x \perp_z y$ can be reduced to an algebraic relation between the respective boolean algebras. What is important here is not these algebraic details, but the insights that this reduction brings into the play. One crucial insight is the following. Because of the very same definition of \dot{x} , the function \dot{x} arising from a team X will necessarily be an *injective function*. What happens if we drop the injectivity requirement? The dropping of this requirement leads to an interesting new notion: the notion of *multi-team*³.

Definition (Multi-team). A *multi-team* X with values in A and domain $\text{dom}(X) \subseteq \text{Var}$ is a pair (I, τ) such that I is a set and $\tau : I \rightarrow A^{\text{dom}(X)}$ is a function.

The main additional feature of a multi-team with respect to a simple team is that given a finite multi-team X there is a canonical way to associate probabilities to X , in fact we can talk about the frequency of the occurrence of a given assignment s in X , i.e. the number of rows of X in which the assignment s occurs, divided by the total number of rows of X . I.e. we can compute probabilities using the *normalized counting measure*. This brings team semantics towards the field of *information theory*, where these kind of probabilities are of crucial interest. In Article VI we

³Although this notion is not explicitly introduced in Article III, the idea of a multi-team (and the perspective it brings) originates there.

will see how the notion of multi-team can be generalized to a new notion, the notion of *quantum team*, so as to take into account the non-classical phenomena arising from *quantum information theory*, where one deals with correlations between probabilities which violate the basic axioms of probability theory.

But the insights of Article III are not limited to this. As it is clear to anybody familiar with the basics of probability theory, the idea behind Article III is simply to bring ideas from the theory of random variables into the world of database theory, or team semantics. I.e. a multi-team is nothing but a random vector after dropping the measurability requirements. And, as we associate to random vectors probability algebras, we associate to multi-teams boolean algebras, using the inverse image operator. The discrete nature of multi-teams allows us to talk about “discrete probabilities”, i.e. probabilities in the sense of the normalized counting measure. But what about using an arbitrary probability measure? I.e. is it possible to devise a version of team semantics where we evaluate formulas with respect to random vectors whose domain is an arbitrary probability space? This question motivates the introduction of a new notion: the notion of *measure team*. This is the topic of Article VI, where based on this notion we devise a very expressive logic, with many areas of applications, most notably biology.

We then saw how the abstract perspective of Article III lead to concrete developments in team semantics, with applications in biology and quantum physics, through the introduction of ideas extraneous to the original setting of the field. This motivates a strong connection between the second part of the thesis and the first.

Before introducing the specifics of Article VI and VII, we talk about Article V, where a new form of database dependency is studied, known as *G-dependence*. The focus of this paper is again on the solution of implication problems, as in Article I. Article V can in fact be read in direct continuity with Article I, where the notions and motivations behind the study of implication problems are treated in particular detail. This establishes another clear connection between the first and the second part of this thesis.

4.1. G-Dependence

In [14] Kurt Grelling lays the foundations of a mathematical theory of dependence, isolating several notions of dependence and noticing various properties that they satisfy. Among them, there is the notion of functional dependence, already mentioned above, and other notions of dependence familiar in combinatorics, specifically in the study of closure operators. Grelling also introduces a notion of dependence that fails to fit with any of the currently known forms of dependence. He calls this form of dependence *varequiddependence*, but we will refer to it simply as *G-dependence*. In [26] Väänänen, commenting the above-mentioned paper of Grelling, frames G-dependence in the context of database theory and team semantics. From the perspective of multi-teams, G-dependence is defined as follows.

Definition (G-dependence). Let $X = (I, \tau)$ be a multi-team and x and y finite sets of variables. We say that y G-depends on x , in symbols $X \models \varpi(x, y)$, if for every $s, s' \in I$, if there exists exactly one $x_i \in x$ such that $\tau(x_i)(s) \neq \tau(x_i)(s')$, then there exists at least one $y_j \in y$ such that $\tau(y_j)(s) \neq \tau(y_j)(s')$.

The notion of G-dependence is motivated by experimental scientific practice, in accordance with the general philosophical interests of Grelling, and it has a strong

intuitive appeal (think of the price of an hotel room as a function of its size and number of beds). In [26] Väänänen notices several properties of G-dependence, leaving open the problem of finding a complete axiomatization for this form of dependence, i.e. solving its implication problem, in the terminology of Section 3.1. In Article V we show that G-dependence admits a very natural finite complete axiomatization, in the style of Armstrong’s axiomatization of functional dependence [2]. Furthermore, we prove that G-dependence admits Armstrong relations, in the sense of [2], [5] and [4].

Theorem. Let $\Sigma \cup \{\sigma\}$ a set of G-dependence atoms. Then

$$\Sigma \vdash \sigma \quad \text{if and only if} \quad \Sigma \models \sigma$$

I.e. the G-dependence axioms (and rules) are complete. Furthermore, G-dependence admits Armstrong relations. I.e. given a set Σ of G-dependence atoms, there exists a team X such that

$$X \models \sigma \quad \text{iff} \quad \Sigma \vdash \sigma.$$

4.2. Quantum Teams

By definition, multi-teams can get values in any set A , but with respect to applications in information theory the context of multi-teams with values in $\{0, 1\}$ is already highly non-trivial. This is the context of Article VI, where, following the work of [1], team semantics meets the field of *quantum information theory*.

As mentioned above, there is a canonical way to associate probabilities to a finite multi-team, i.e. using the normalized counting measure. Specifically, given a finite multi-team X with values in $\{0, 1\}$, and a propositional formula ϕ we can define the probability of ϕ in X as

$$[\phi]_X = \frac{|\{i \in \Omega : \tau(i)(\phi) = 1\}|}{|\Omega|} = P(\{i \in \Omega \mid \tau(i)(\phi) = 1\}).$$

In this way multi-teams allow for the modelling of many interesting information-theoretic phenomena, but multi-teams have a strong inherent feature, i.e. they are *classical*. That is, when we define a multi-team X as a multiset of assignments to the propositional variables in $\text{dom}(X)$, we assume that the variables in $\text{dom}(X)$ are a set of compatible observables, i.e. we can *simultaneously* give a value to every variable in $\text{dom}(X)$. But this is very far from being an innocuous assumption. In fact, under this hypothesis we have strong *constraints* on the possible correlations between probabilities arising from a multi-team (so-called *Bell’s inequalities*), which are violated by the mathematical formalism of quantum mechanics and by experimental verifications thereof. Thus, if we want to account for the “quantum world” we have to generalise the notion of multi-team to a new notion, what we call *quantum teams*.

Definition (Quantum team). Let Ω be a finite set and $\mathcal{Q} = (Q_i)_{i \in \Omega}$ a sequence of finite non-empty sets of proposition symbols. A *quantum team* on \mathcal{Q} is a pair $X = (\Omega, \tau)$ such that $\tau(i)$ is a truth-value assignment to the proposition symbols in Q_i for each $i \in \Omega$.

The intuitive interpretation of the definition above is that the sets $(Q_i)_{i \in \Omega}$ are the set of compatible observables in the observational scenario under analysis. From a technical point of view, the difference is that we are not computing probabilities with respect to a single fixed probability space, but with respect to a family of

probability spaces, one for each subset U of a given Q_i . More explicitly, given a propositional formulas ϕ and a subset U of some Q_i we let

$$[\phi]_{X,U} = \frac{|\{i \in \Omega : \tau(i)(\phi) = 1\}|}{|\Omega_U|} = P(\{i \in \Omega \mid \tau(i)(\phi) = 1\}),$$

where $\Omega_U = \{i \in \Omega \mid U \subseteq Q_i\}$. With this generalised notion of probability of a formula in a team X we can recover many non-classical phenomena of interest in quantum information theory. The point is then to devise a good logic for quantum teams. Based on the logic of [9], we build a propositional logic whose atomic formulas are expressions of the kind

$$a_0(\phi_0; V_0) + \dots + a_{k-1}(\phi_{k-1}; V_{k-1}) \geq c,$$

where we interpret $(\phi_0; V_0)$ as the probability $[\phi]_{X,U}$, in order to capture the phenomenon that there are limitations as to what observables can be measured simultaneously. Relativising the deductive system of [9] to this more complicated syntax, we prove a completeness theorem for this logic, which we call *quantum team logic*.

Theorem (Completeness). Let α be a formula of quantum team logic. Then

$$\vdash \alpha \Leftrightarrow \models \alpha.$$

As an application, we show how to resolve an apparent paradoxicality noticed by Feynman in the famous “Double-Slit Experiment” of [10].

4.3. Measure Teams

As mentioned above, a measure team is a random vector of assignments of a given set of variables. In order for this notion to make sense, we of course have to put some measurability conditions. This is done in the following way.

Definition (Measure team). Let L be a signature and \mathcal{A} an L -structure. A *measure team* X with values in \mathcal{A} and domain $\text{dom}(X) \subseteq \text{Var}$ is a quadruple $(\Omega, \mathcal{F}, P, \tau)$ such that (Ω, \mathcal{F}, P) is a probability space and $\tau : \Omega \rightarrow A^{\text{dom}(X)}$ is a measurable function, in the sense that

$$\{i \in \Omega : \mathcal{A} \models_{t(i)} \phi\} \in \mathcal{F}$$

for every first-order L -formula ϕ with free variables in $\text{dom}(X)$.

We then see that the function τ in the definition above is nothing but a random vector from (Ω, \mathcal{F}, P) into the space of definable subsets of \mathcal{A} , and thus that this notion captures exactly the intuitive idea described above.

But does this abstract definition have any relevance for “concrete situations”? What are the canonical examples of measure teams? The first example that comes to mind is the case of finite multi-teams with the normalized counting measure, but this example does not justify the generality of the setting. The paradigm example of measure team is another one. This is the case of an idealised measurement of given variables v_0, \dots, v_n at all points of time starting at time 0 and ending at time 1. The values of the variables are e.g. real numbers, which change continuously with time, such as temperature, speed, pressure, amplitude, force, etc. When time progresses from 0 to 1, the vector $(s_t(v_0), \dots, s_t(v_n))$ flows from $(s_0(v_0), \dots, s_0(v_n))$ to $(s_1(v_0), \dots, s_1(v_n))$.

Thus, the notion of measure team has its canonical incarnation in the formalization of classical experimental situations. Consequently, we want to devise a syntax

capable of expressing as much as possible regarding statistics made on experimental scenarios. To this extent, we devise a syntax L_1 based on the language of the ordered field of real numbers $\{0, 1, +, -, \cdot, \leq\}$ (in order to express properties of our statistics), which is in function of an arbitrary countable signature L_0 , to be thought as expressing the fundamental aspects of our experimental setting. Technically, what we do is adding a constant symbol $|\phi(x)|$ to $\{0, 1, +, -, \cdot, \leq\}$ for every L_0 -formula $\phi(x)$.

This rich language allows for a broad field of applications. A leading example is within biology, in the study of pools of genes. This is the *Hardy-Weinberg Principle*, which shows that a conservation phenomenon takes place in a gene pool from generation to generation under certain assumptions, such as random mating. Other applications are in quantum mechanics, Markov chains, and the analysis of known paradoxes of probability logic.

The semantics for our language L_1 is given straightforwardly. In fact, for any team X with values in an L_0 -structure \mathcal{A} , we can naturally define the notion of probability of a formula in a team $X = (\Omega, \mathcal{F}, P, \tau)$, simply as

$$[\phi]_X = P(\{i \in \Omega : \mathcal{A} \models_{\tau(i)} \phi\}).$$

Now, given a measure team X with values in \mathcal{A} and $\text{dom}(X) = \{v_i : i < n\}$, we let \mathcal{R}_Q^X be the expansion of $\mathcal{R} = (\mathbb{R}, 0, 1, +, -, \cdot, \leq)$ to an L_1 -structure obtained by interpreting

$$|\phi(x)|^{\mathcal{R}_Q^X} = [\phi(x)]_X.$$

Definition 4.1 (Semantics). Let Σ be an L_1 -theory, \mathcal{A} an L_0 -structure, X a measure team with values in \mathcal{A} and $\text{dom}(X) = \{v_i : i < n\}$. We define

$$X \models \Sigma \Leftrightarrow_{def} \mathcal{R}_Q^X \models \Sigma.$$

The main result of Article VII is the elaboration of a complete deductive system for the language L_1 under this semantics.

Theorem (Completeness). Let T be an L_0 -theory and Σ a positive bounded L_1 -theory. Then the following are equivalent.

- (i) There exists $\mathcal{A} \models T$ and measure team $X = (\Omega, \mathcal{F}, P, \tau)$ with values in \mathcal{A} and $\text{dom}(X) = \{v_i : i < n\}$, such that $X \models \Sigma$.
- (ii) $(T, \Sigma) \not\perp$.
- (iii) As in (i), with $\Omega = [0, 1]$, \mathcal{F} the σ -algebra \mathcal{L} of Lebesgue measurable subsets of $[0, 1]$ and P the Lebesgue measure on $[0, 1]$.

Because of this completeness result, the Hardy-Weinberg Principle (and all other applications of our system) is not only a property of measure teams expressible semantically, but it is an actual *theorem* of our logic.

REFERENCES

- [1] Samson Abramsky and Lucien Hardy. Logical Bell Inequalities. *Physical Review A*, 85(062114):1–11, 2012.
- [2] William Ward Armstrong. Dependency Structures of Data Base Relationships. In *IFIP Congress*, pages 580–583, 1974.
- [3] J. T. Baldwin. *First-Order Theories of Abstract Dependence Relations*, Ann. Pure Appl. Logic 26, 215–243 (1984).
- [4] Catriel Beeri, Martin Dowd, Ronald Fagin and Richard Statman. On the Structure of Armstrong Relations for Functional Dependencies. *Journal of the ACM*, 31(1):30–46, 1984.

- [5] Catriel Beeri, Ronald Fagin and John H. Howard. A Complete Axiomatization for Functional and Multivalued Dependencies in Database Relations. In: *SIGMOD '77 Proceedings*, pp. 47-61, 1977.
- [6] Itai Ben Yaacov. On Theories of Random Variables. *Israel J. Math.*, 194(2):957–1012, 2013.
- [7] Henry H. Crapo and Giancarlo Rota. *On the Foundations of Combinatorial Theory: Combinatorial Geometries*. M.I.T. Press, Cambridge, Mass, 1970.
- [8] Henry H. Crapo. Single-Element Extensions of Matroids. *J. Res. Nat. Bur. Standards, Sect. B*, v. 69B, 1965, pp. 55-65. MR 32 # 7461.
- [9] Ronald Fagin, Joseph Y. Halpern, and Nimrod Megiddo. A Logic for Reasoning about Probabilities. *Information and Computation*, 87:78–128, 1990.
- [10] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman lectures on physics. Vol. 3: Quantum mechanics*, Addison-Wesley Publishing Co., Inc., Reading, Mass.-London, 1965.
- [11] Pietro Galliani and Jouko Väänänen. On Dependence Logic. Trends in Logic: Outstanding Contributions (ed. A. Baltag and S. Smets), Springer, 2014.
- [12] Dan Geiger, Azaria Paz, and Judea Pearl. Axioms and Algorithms for Inferences Involving Probabilistic Independence. *Inform. and Comput.*, 91(1):128–141, 1991.
- [13] Erich Grädel and Jouko Väänänen. Dependence and Independence. *Studia Logica*, 101(2):399–410, 2013.
- [14] Kurt Grelling. A Logical Theory of Dependence. Included in: Barry Smith, editor. Foundations of Gestalt Theory. Philosophia, Munich and Vienna, 1988.
- [15] Christian Herrmann. On the Undecidability of Implications between Embedded Multivalued Database Dependencies. *Inform. and Comput.*, 122(2):221–235, 1995.
- [16] Christian Herrmann. Corrigendum to "On the Undecidability of Implications between Embedded Multivalued Database Dependencies" [*Inform. and Comput.* 122 (1995) 221–235]. *Inform. and Computat.*, 204(12):1847–1851, 2006.
- [17] Wilfrid Hodges. Compositional Semantics for a Logic of Imperfect Information. *Log. J. IGPL*, 5:539-563, 1997.
- [18] D. Stott Parker, Jr. and Kamran Parsaye-Ghomi. Inferences Involving Embedded Multivalued Dependencies and Transitive Dependencies. In *Proceedings of the 1980 ACM SIGMOD international conference on Management of data*, SIGMOD '80, pages 52–57, New York, NY, USA, 1980. ACM.
- [19] Judea Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman, San Mateo CA 1988.
- [20] Anand Pillay. *Geometric Stability Theory*. Oxford University Press, 1996.
- [21] Yehoshua Sagiv and Scott F. Walecka. Subset Dependencies and a Completeness Result for a Subclass of Embedded Multivalued Dependencies. *J. ACM*, 29(1):103–117, 1982.
- [22] S. Shelah. *Classification Theory: and the Number of Non-Isomorphic Models* (North-Holland, Amsterdam, 1990).
- [23] Saharon Shelah, *Classification Theory for Abstract Elementary Classes*. College Publications, London, 2009.
- [24] Milan Studeny. Conditional independence relations have no finite complete characterization. In: Transactions of the 11th Prague Conference on Information Theory, pp. 377 – 396. Kluwer, 1992.
- [25] Jouko Väänänen. *Dependence logic*, volume 70 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2007.
- [26] Jouko Väänänen. Grelling on Dependence. In: Dependence Logic: Theory and Applications. Springer (to appear), 2015.