A comment on Samson Abramsky's paper "Relational Databases and Bell's Theorem"

Jouko Väänänen

May 13, 2014

Abramsky points out a connection between a non-locality phenomena in quantum physics and a non-locality phenomenon in database theory. This seems surprising as one would not expect—a priori—such a connection to exist. The proof of the relevant quantum theoretic phenomenon turns out in Abramsky's analysis to be a mathematical rather than a physical phenomenon. It is based on the nonexistence, in the mathematical sense, of probability distributions satisfying certain conditions. The probability distributions can even be replaced by 2-valued distributions with the same effect.

A similar mathematical non-existence phenomenon exists in database theory. The phenomenon is twofold. On the one hand there is the decomposition problem: how to decompose a large database into smaller parts, which would form an instance of a database schema with the associated dependencies. On the other hand we may start from an instance of a schema and ask whether there is a universal relation.

Abramsky calls this *contextuality* and suggests that a common *logic of contextuality* is emerging from these examples. We may also ask, is contextuality an even more general phenomenon? If it is, there is all the more reason to take a logical approach to the question. By a logical approach I mean establishing a language, a semantics of the language and an attempt to figure out the logical properties.

The characteristic feature of contextuality, be it quantum physics or database theory, is the presence of several tables of data and the question is whether they arise from just one universal table by projection. This co-existence of several limited "realities" is not an unusual theme in the history of ideas, and neither is the question whether there is one ultimate "truth". This is the case in different areas of humanities for rather obvious reasons, but it exists even in mathematics. For example, in set theory there is discussion about a multiverse position according to which the fact that we have not been able to solve questions such as the Continuum Hypothesis is a consequence of the circumstance that the set theoretical reality of mathematical objects is a multiverse, i.e. a collection of universes, very "close" to each other, some satisfying the Continuum Hypothesis and some not. One may indeed reformulate first order logic so that the semantics is not based on the concept of a model satisfying a sentence, but on the concept of a multiverse model satisfying a sentence [5].

The question of contextuality arises in database theory rather naturally. When dealing with large databases one observes that a certain datum, e.g. the telephone number of a person, occurs repeatedly, and it is more efficient to have a separate small database for telephone numbers. This decomposition of a large database into smaller pieces is possible because of *dependencies*, in our case that of the person functionally determining his or her telephone number. For an ideal result one would also want the components to be *independent* of each other. The decomposition results in an *instance* of a database schema. The question is, given an instance, does it arise from the decomposition of a big "universal" database into smaller ones. Abramsky gives strong evidence to category theory, especially the language of presheaves and sheaves, being a natural framework for studying this kind of issues concerning schemas and relational databases.

The logic generally used to express dependencies in database theory is first order logic or a fragment of it called the *relational calculus.* Individual databases of instances are treated as *relations.* The existence of a *functional dependence* $x \to y$ between attributes x and y in a relation R can be expressed in first order logic as

$$\forall x, y, y', \vec{z}, \vec{z'}((R(x, y, \vec{z}) \wedge R(x, y', \vec{z'})) \to y = y'). \tag{1}$$

Another kind of dependence is *inclusion*, $x \subseteq y$, expressing the concept that every value of the attribute x in relation R occurs as a value of the attribute y in relation S. This can be expressed in first order logic as

$$\forall x, \vec{z}(R(x, \vec{z}) \to \exists \vec{u} S(x, \vec{u})). \tag{2}$$

However, to express that a universal relation exists for binary relations R_1, R_2, R_3 seems in general—not least because of its NP-completeness in the general case—to require existential second order logic:

$$\begin{aligned} \exists S(\forall x, y(R_1(x, y) \leftrightarrow \exists z S(x, y, z)) \wedge \\ \forall y, z(R_2(y, z) \leftrightarrow \exists x S(x, y, z)) \wedge \\ \forall x, z(R_3(x, z) \leftrightarrow \exists y S(x, y, z))). \end{aligned} \tag{3}$$

So here we go beyond first order logic. Is existential second order logic the right "logic of contextuality"? As it is, existential second order logic is badly non-axiomatizable: the set of valid sentences is complete Π_2 in the Levy hierarchy, which means that to check the validity of a sentence requires scanning the cumulative hierarchy up to inaccessible cardinals (if they exist) and beyond. In the database context this is not so drastic as databases correspond to finite models and even first order logic itself is non-axiomatizable in finite models.

The approach of *generalized quantifiers*, introduced by A. Mostowski in 1957, is to consider particular second order definable quantifiers, such as "for infinitely many x" or "for uncountably many x", or in finite models "for an even number of x", and add them to first order logic. In many cases this has resulted in fragments of second order logic with nice properties, such as effective axiomatization and countable compactness. Perhaps we can do the same here, not by adopting generalized quantifiers, but by adopting *generalized atomic formulas*. Since we are talking about *logical notions* here, the atomic formulas that are most naturally subject to generalisation are the identities, such as $x^1 = y^1$. Accordingly, let us take an atomic formula

$$=(x^1, y^1) \tag{4}$$

with (1) as the meaning with the understanding that the variables x^1,y^1,$\vec{z}^1$ denote the attributes of the relation R. Thus a relation R satisfies (4) iff (1) is true. Likewise, let us take an atomic formula

$$x^1 \subseteq x^2 \tag{5}$$

with (2) as the meaning with the understanding that the variables x^1,$\vec{z}^1$ denote the attributes of the relation R and the variables x^2,$\vec{z}^2$ denote the attributes of the relation S. Thus an instance (R, S) satisfies (5) iff (2) is true. If we add these new atomic formulas to first order logic and define the logical operations in a canonical way we get a "schema"-version of what is in the single database (or universal database) case known as *dependence logic* [6], if only (4) is added, *inclusion logic* [1], if only (5) is added, and (essentially) *independence logic* [3], if both are added.

One may reasonably ask, in what way are dependence logic, inclusion logic and independence logic better than mere existential second order logic, of which they all are fragments? The point is that we have more control and can make finer distinctions. For example, inclusion logic has on finite models exactly the expressive power of fixed point logic [2]. Dependence logic formulas can express exactly all second order properties of (single) relations which are closed downwards (as (1) is but (2) is not) [4]. Finally, independence logic can in fact express all existential second order properties of (single) relations [1].

The origin of dependence logic is in game theoretic semantics. There a typical database consists of plays of the semantic game associated with a sentence and a model. If the variables x, y, z take on values during the game, then subsequent plays of the game yield a database with x, y, z as attributes. From this database we can see e.g. whether the player that picked z was playing a strategy (expressed by $=(x, y, z)$), whether the player used partial information (expressed e.g. by $=(x, z)$), whether the player was committed to play only values that the player picking y uses (expressed by $z \subseteq y$) etc. These examples show how the above generalized atoms arise naturally in game theoretic semantics. They arise also naturally in experimental science ("the time of descent is functionally determined by the height but independent of the weight"), social choice (the value of the social welfare function on the choice between a and b is functionally determined by the choices between a and b of the voters), biology (Mendel's Laws), and philosophy (e.g. inquisitive logic). In all these application areas one can see relevant universal relation type questions, even if they may have not been subjected to mathematical study. Laws governing logics arising from atoms such as $=(x, y)$ and $x \subseteq y$ may offer the beginning of a new "logic of contextuality", but this requires development of "multi-relation" versions of these logics.

Let us return to (3), which plays a central role in Abramsky's paper. The condition (3) is an existential second order property of the relations R_1, R_2 and R_3, but over which domain? We can construe it as being over the so-called *active domain*, which occurs in (3) only implicitly. Currently this seems to go beyond the logical apparatus of even the strongest of the above logics, the independence logic. However, we can take it as a new atom, in the spirit of $=\!(x, y)$ and $x \subseteq y$. So let us adopt a new atom

$$\bowtie(x^1y^1, y^2z^2, x^3z^3), \tag{6}$$

with (3) as the meaning with the understanding that the variables x^1,y^1 denote the attributes of the relation R^1, the variables y^2,z^2 denote the attributes of the relation R^2, and the variables x^3,z^3 denote the attributes of the relation R^3. Thus an instance (R^1, R^2, R^3) satisfies (6) iff (3) is true[1]. If this atom is added to first order logic, we still remain within existential second order logic. Can we axiomatize this atom in the same sense as (4) and (5) have been axiomatised?

My contention is that the surprising connection Abramsky's paper establishes between the mathematics of quantum physics and database theory is an indication that there may be a general multiverse logic underlying phenomena not only in physics and computer science, but also in social choice, biology and philosophy. Such a logic does not exist yet, but I believe that it will exist.

Pietro Galliani made the interesting observation that (6) added to first order logic gives exactly inclusion logic and hence on finite models fixed point logic.

References

[1] Pietro Galliani. Inclusion and exclusion dependencies in team semantics: On some logics of imperfect information. *Annals of Pure and Applied Logic*, 163(1):68 – 84, 2012.

[2] Pietro Galliani and Lauri Hella. Inclusion logic and fixed point logic. CSL 2013 to appear, arXiv:1304.4267 [cs.LO].

[3] Erich Grädel and Jouko Väänänen. Dependence and independence. *Studia Logica*, 101(2):233–236, 2013.

[4] Juha Kontinen and Jouko Väänänen. On definability in dependence logic. *J. Log. Lang. Inf.*, 18(3):317–332, 2009.

[5] Jouko Väänänen. Multiverse set theory and absolutely undecidable propositions. In Juliette Kennedy, editor, *Interpreting Gödel*. Cambridge University Press. to appear.

[1]The new atom resembles but is different from the so called *join dependency* atom $\star(xy, yz, xz)$ which would say of *one* relation that it is the join of its projections to xy, xy and xz, respectively. Atoms of the type $\star(xy, yz, xz)$ can be expressed in independence logic.

[6] Jouko Väänänen. *Dependence logic*, volume 70 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2007.