Question entailment: a puzzle

Balder ten Cate

LogicBlox Inc. and UC Santa Cruz
btencate@ucsc.edu

One of the things I have enjoyed tremendously ever since my my first days a PhD student under supervision of Jeroen Groenendijk and Paul Dekker, is the impressive cross-fertilization between logic, linguistics, and computer science that permeates the work of many researchers at the ILLC, and, in in particular, is palpable in many of the influential publications of Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. In this note, as a tribute to Jeroen, Martin and Frank, and as a small contribution to this cross-fertilization, I will discuss a puzzle regarding question entailment, taken from recent literature in database theory (it will not come as a surprise that questions, answerhood conditions, and question entailment play an important role in database theory, too).

Question entailment and the Beth definability theorem The influential Groenendijk & Stokhof [4] account of the semantics of questions posits: in the same way that assertions can be viewed as denoting subsets of the space of possible worlds, questions can be viewed as denoting partitions of the space of possible worlds (intuitively, the question asks which of the blocks of the partition contains the actual world). This *partition semantics* of questions naturally gives rise to an entailment relation between questions, namely that of partition refinement.

Formally, let us fix a first-order signature consisting of relation symbols and constant symbols, and fix a set of possible worlds W, where each possible world is identified with a first-order structure, all sharing the same domain and the same interpretation for the constant symbols. Then we can view a first-order formula (with or without free variables) as expressing a question:

$$[[?\phi]]_W = \{(w, v) \in W^2 \mid \forall g \; w, g \models \phi \textit{ iff } v, g \models \phi\}.$$

We use the notation $?\phi$ here to stress that ϕ is interpreted as a question. Observe that $[[\phi]]_W$ is indeed an equivalence relation (i.e., a partition) of W.

Question entailment then amounts to partition refinement: $?\phi$ *entails* $?\psi$, relative to W (notation: $?\phi \models_W ?\psi$) if $[[?\phi]]_W \subseteq [[?\psi]]_W$. Thus, for example, the question *who is wearing shorts today?* entails the question *is anybody wearing shorts today?* because the first partition is a refinement of the second partition (in other words, every block of the second partition is a union of blocks of the first partition). The same definition naturally extends to entailment with sets of questions:

$$?\phi_1, \ldots, ?\phi_n \models_W ?\psi \text{ if } \bigcap_i [[?\phi_1]]_W \subseteq [[?\psi]]_W \,.$$

We write $?\phi_1, \ldots, ?\phi_n \models ?\psi$ if $?\phi_1, \ldots, ?\phi_n \models_W ?\psi$ for all sets of possible worlds W as described above. Question entailment (through the notion of *licensing*) features prominently in the logic of interrogation [5], where it is used in an *update semantics* [8] style framework to give a formal explication of Grice's *maxim of relation* [3].

In [7] we observed an intimate relationship between question entailment and the notion of *implicit definability* from the (projective) Beth definability theorem [2]. Indeed, $?\phi_1, \ldots, ?\phi_n \models ?\psi$ holds precisely if the formula ψ is implicitly definable in terms of the set of formulas

$\{\phi_1, \ldots, \phi_n, \} \cup \{x = c \mid c$ is a constant symbol$\}$ (the special appearance of the constant symbols in this characterization is due to the fact that they are treated as rigid designators here). The same holds relative to any given first-order theory (i.e., when restricted to models of the given theory).

This observation was then put to use in [7], by combining it with the actual statement of the projective Beth definability to give a syntactic characterization of question entailment. Let us say that a formula ϕ is *built up from* a set of formulas Ψ if ϕ is constructed out of formulas in Ψ (renaming of variables allowed) and equalities using the connectives and quantifiers of first-order logic. Thus, for example, $\exists x((P(x,y) \wedge Q(x,y)) \vee R(x,y))$ is built up from $\{P(x,y) \wedge Q(x,y), R(u,v)\}$. It can be shown by a straightforward induction that whenever ϕ is built up from $\{\psi_1, \ldots, \psi_n\}$, then ϕ is implicitly defined in terms of $\psi_1, \ldots, \psi_n$, and hence $?\psi_1, \ldots, ?\psi_n \models ?\phi$. The projective Beth definability theorem, implies the converse modulo-logical-equivalence, and therefore yields:

(†) $?\psi_1, \ldots, ?\psi_n \models ?\phi$ if and only if ψ is equivalent to a first-order formula built from $\{\psi_1, \ldots, \psi_n\} \cup \{x = c \mid c$ is a constant symbol$\}$

Again, the same holds relative to any first-order theory. There are various known proofs of the Beth definability theorem, both model theoretic ones and proof theoretic ones. The Craig interpolation theorem is typically used as a stepping stone. In [7], further consequences and variations of the above syntactic characterization of question entailment are studied.

Question entailment in database theory Question entailment has been extensively studied, under a different name, in database theory. There, a common scenario is that users repeatedly query the database. To improve the performance of the system, it is sometimes helpful to store the answers to previous queries as "materialized views", with the goal to answer subsequent queries faster. Two concepts of central importance in this context are *view-based determinacy* and *view-based rewritings* [6]. For present purposes, a query is simply a first-order question, and a view is nothing more than the (stored) denotation of a query. We say that a query q is *determined* by a collection of views $V_1, \ldots, V_n$ if $V_1, \ldots, V_n$ provide enough information to answer q, that is, whenever two database instances are indistinguishable by the views, then also the query has the same answers in both database instances. Observe that if each view V_i is defined by the query q_i, then this holds precisely if $q_1, \ldots, q_n$, viewed as first-order questions, entail q. A *rewriting* of q in terms of $V_1, \ldots, V_n$ is a query that uses $V_1, \ldots, V_n$ as atomic relation symbols, and that is equivalent to q, given the view definitions. Observer that such a rewriting is nothing more than an explicit definition, i.e., a formula built up from $q_1, \ldots, q_n$ that is equivalent to q. Of course, as holds for most interesting problems regarding first-order formulas, determinacy and the existence of rewritings are undecidable.

In the above discussion of view-based determinacy and view-based rewritings I have skipped over some subleties that specific to the database context. In particular, two characteristic features of typical database querying scenarios are: (i) we are often interested not in arbitrary first-order queries, but in queries belonging to certain syntactic fragments of first-order logic. Of most central importance are the *conjunctive queries*. A conjunctive query is a query defined by a first-order formula built up from atomic formulas using only conjunction and existential quantification. These cover the most frequent type of queries in practice; (ii) we are particularly interested in finite structures, since these can be viewed as database instances. This makes a big difference because many results from the classic model theory of first-order logic, including the Beth definability theorem, break down when we restrict attention to finite structures. For an overview of the state-of-the-art and a list of open problems in the study of view-based determinacy and view-based rewritings for important query languages such as conjunctive queries, see [6].

A puzzle (from [1]): Consider the following first-order questions (concerning degrees of separation if we can think of R as expressing aquintance):

Q3: $?xy\ \exists uv(R(x,u) \wedge R(u,v) \wedge R(v,y))$

Q4: $?xy\ \exists uvw(R(x,u) \wedge R(u,v) \wedge R(v,w) \wedge R(w,y))$

Q5: $?xy\ \exists uvwz(R(x,u) \wedge R(u,v) \wedge R(v,w) \wedge R(w,z) \wedge R(z,y))$

Is it the case that Q3 and Q4 entail Q5 (in other words, using terminology from [5], do Q3 and Q4 license every answer to Q5)?

Hint: The characterization in (†) suggests an approach to solving this puzzle. Namely, if Q3 and Q4 entail Q5, then Q5 must be equivalent to a formula built up from Q3 and Q4. If, on the other hand, Q3 and Q4 do *not* entail Q5, then it follows from the definition of question entailment that we should be able to find a pair of first-order structures that is distinguished by Q5 and not by Q3 or Q4.

Solution: It turns out [1, 6] that Q5 *is* entailed by Q3 and Q4. Indeed, there is a path of length 5 from a node x to a node y if and only if *there is a path of length 4 from x to some node u and for every path of length 3 from a node v to u, there is a path of length 4 from v to y*. This shows that Q5 is equivalent to a formula built up from Q3 and Q4.

Analysis: What makes this puzzle interesting, from a database point of view, is that, even though Q3, Q4 and Q5 are all conjunctive queries, Q5 is not equivalent to any conjunctive query built up from Q3 and Q4. Indeed every first-order formula built up from Q3 and Q4 that is equivalent to Q5 must involve negation or universal quantification [1, 6]. This shows that the Beth definability theorem does not hold when restricted to conjunctive queries.

References

[1] F. N. Afrati. Determinacy and query rewriting for conjunctive queries and views. *Theor. Comput. Sci.*, 412(11):1005–1021, 2011.

[2] E. Beth. On Padoa's method in the theory of definition. *Indagationes Mathematicae*, 15:330–339, 1953.

[3] H. P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics*, volume 3. New York: Academic Press, 1975.

[4] J. Groenendijk and M. Stokhof. Questions. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, chapter 19, pages 1055–1124. Elsevier, 1997.

[5] J. A. G. Groenendijk. The logic of interrogation: Classical version. *Proceedings of the Ninth Conference on Semantics and Linguistic Theory*, 1999.

[6] A. Nash, L. Segoufin, and V. Vianu. Views and queries: Determinacy and rewriting. *TODS*, 35(3), 2010.

[7] B. ten Cate and C.-C. Shan. Axiomatizing groenendijk's logic of interrogation. In *Questions in Dynamic Semantics*, pages 63–82. Elsevier, 2007.

[8] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.