Addendum to "Current approaches to handling imperfect information in data and knowledge bases"

Simon Parsons

Department of Electronic Engineering,
Queen Mary and Westfield College,
University of London,
London E1 4NS,
United Kingdom.

November 29, 1999

In a recent paper [10] I discussed, among other things, the different ways in which other authors had proposed handling incomplete information in databases using null values. This discussion contained one mistake and a number of omissions. The purpose of this note is to correct the mistake and to fill in the omissions.

The mistake was to attribute the idea of replacing missing information with null values to Codd's 1979 paper [2]. In fact, by the time that this paper was written, the use of null values was widespread. Indeed, the first use of null values in the context of relational algebra was contained in a paper written by Codd in 1975 paper [3], and the ANSI/X3/SPARC interim report [1] published the same year included 14 different ways in which null values might be interpreted suggesting wide acceptance of their use.

The omissions were various. The most serious was that of the work of Grant, who was working on the problem of incomplete information as early as 1974 [8], and who took Codd's initial results on null values in relational algebra and extended them considerably. The first extension [7] involved a slight alteration to Codd's three-valued logic to ensure that it retrieved the right tuples in all situations. The next extension [6] was to handle what Grant calls "partial values", that is values which are known to lie in an interval and so are partially determined, and this idea was further developed in [5]. This line of research was finally extended [9] to deal with indefinite information (that is information which in [10] I called disjuntively imprecise) using a generalisation of the closed world assumption.

Other important work on incomplete information in databases which I was unaware of when writing [10] was that of Vassiliou, Goldstein and Sagiv. Vassiliou showed how to handle both the "unknown" and "nonexistant" interpretations of null values [13] and investigated how null values can be used in conjunction with functional dependencies [12]. Thus his work can be considered as investigating how models can be extended to take account of incomplete infor-

mation. Goldstein [4], on the other hand, was concerned with the ways in which the use of null values can be limited, and discussed means by which database designers may restrict the use of null values, while Sagiv [11] investigated the ramifications of the decision not to use null values when building data models.

References

[1] ANSI/X3/SPARC "Study Group on Data Base Management Systems, Interim Report," *FDT Bulletin of ACM-SIGMOD*, 7(2), 1975.

[2] E. F. Codd, "Extending the database relational model to capture more meaning," *ACM Transactions on Database Systems*, 4(4), 397–434, 1979.

[3] E. F. Codd, "Understanding relations (Installment #7)," *FDT Bulletin of ACM-SIGMOD*, 7(3/4), 23–28, 1975.

[4] B. S. Goldstein, "Constraints on null values in relational databases", *Proceedings of the Conference on Very Large Databases*, 101–110, 1981.

[5] J. Grant, "Incomplete information in a relational database", *Fundamenta Informaticae*, III (3), 363–378, 1980.

[6] J. Grant "Partial values in a tabular database model," *Information Processing Letters*, 9(2), 97–99, 1979.

[7] J. Grant, "Null values in a relational database," *Information Processing Letters*, 6(5), 156–157, 1977.

[8] J. Grant, "Incomplete models," *Notre Dame Journal of Formal Logic*, XV (4), 601–607, 1974.

[9] J. Grant and J. Minker, "Answering queries in indefinite databases and the null problem," *Advances in Computing Research*, 3, 247–267, 1986.

[10] S. Parsons,"Current approaches to handling imperfect information in data and knowledge bases," *IEEE Transactions on Knowledge and Data Engineering*, 8(3), 353–372, 1996.

[11] Y. Sagiv, "Can we use the universal instance assumption without using nulls?" *Proceedings of the ACM-SIGMOD International Conference on the Management of Data*, 108–120, 1981.

[12] Y. Vassiliou, "Functional dependencies and incomplete information", *Proceedings of the Conference on Very Large Databases*, 260–269, 1981.

[13] Y. Vassiliou, "Null values in data base management: a denotational semantics approach," *Proceedings of the ACM-SIGMOD International Conference on the Management of Data*, 162–169, 1979.