9 CONCLUDING REMARKS

We studied consistent query answering for self-join-free conjunctive queries with respect to primary keys, within the paradigm of Answer Set Programming. In summary, contributions of our work are as follows:

- We showed how to construct programs in non-recursive datalog with negation for problems CERTAINTY(q) that are in **FO**. These programs have lower complexity upper bounds than ASP programs that use the generate-and-test approach.
- Experiments in clingo ASP show that these theoretical differences in complexity also reveal themselves, at least moderately, in practice. As we argued in the previous section, even a constant factor speedup on Boolean queries will give us a polynomial speedup on queries with free variables.
- Interestingly, we showed that CERTAINTY(q) is in **FO** for 95% of the queries over two common database schemas.

Therefore, we conclude that first-order rewriting is an advantageous alternative to generate-and-test.

We conclude by listing three issues for future research. First, the **FO**-boundary of CERTAINTY(q) has recently been revealed for queries q with negation [15], and for multiple keys per relation [18]. It is interesting to extend the results of the current paper to these more general settings. Second, although ASP is declarative by design, there exist nevertheless some good practices and tricks for achieving performance gains. The current paper stayed away from such tricks. In the future, we aim to investigate the optimization of ASP programs obtained from first-order rewriting. For example, if a query q in sjfBCQ can be split into two subqueries, say q_1 and q_2, that have no variables in common, then CERTAINTY(q) can be solved by solving CERTAINTY(q_1) and CERTAINTY(q_2) independently. Some preliminary experiments (not reported here) indicate that such splitting is an effective optimization step. Third, our experiments supported our hypothesis that "yes"- and "no"-instances pose different challenges to generate-and-test and first-order rewriting. For either approach, it remains an open question to more precisely identify and generate the hardest instances of CERTAINTY(q). Another perspective is to look at "typical" instances in particular real-world settings. In database systems environments, for example, almost consistent database instances (i.e., "almost yes"-instances) should be more common than highly inconsistent instances.

REFERENCES

[1] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. 1999. Consistent Query Answers in Inconsistent Databases. In *Proceedings of the Eighteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, May 31 - June 2, 1999, Philadelphia, Pennsylvania, USA*, Victor Vianu and Christos H. Papadimitriou (Eds.). ACM Press, 68–79. https://doi.org/10.1145/303976.303983

[2] Marcelo Arenas, Leopoldo E. Bertossi, and Jan Chomicki. 2003. Answer sets for consistent query answering in inconsistent databases. *Theory Pract. Log. Program.* 3, 4-5 (2003), 393–424. https://doi.org/10.1017/S1471068403001832

[3] Leopoldo E. Bertossi. 2019. Database Repairs and Consistent Query Answering: Origins and Further Developments. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*, Dan Suciu, Sebastian Skritek, and Christoph Koch (Eds.). ACM, 48–58. https://doi.org/10.1145/3294052.3322190

[4] Jan Chomicki, Jerzy Marcinkowski, and Slawomir Staworko. 2004. Hippo: A System for Computing Consistent Answers to a Class of SQL Queries. In *Advances in Database Technology - EDBT 2004, 9th International Conference on Extending Database Technology, Heraklion, Crete, Greece, March 14-18, 2004, Proceedings (Lecture Notes in Computer Science)*, Elisa Bertino, Stavros Christodoulakis, Dimitris Plexousakis, Vassilis Christophides, Manolis Koubarakis, Klemens Böhm, and Elena Ferrari (Eds.), Vol. 2992. Springer, 841–844. https://doi.org/10.1007/978-3-540-24741-8_53

[5] Akhil A. Dixit and Phokion G. Kolaitis. 2019. A SAT-Based System for Consistent Query Answering. In *Theory and Applications of Satisfiability Testing - SAT 2019 - 22nd International Conference, SAT 2019, Lisbon, Portugal, July 9-12, 2019, Proceedings (Lecture Notes in Computer Science)*, Mikolás Janota and Inês Lynce (Eds.), Vol. 11628. Springer, 117–135. https://doi.org/10.1007/978-3-030-24258-9_8

[6] Ariel Fuxman, Elham Fazli, and Renée J. Miller. 2005. ConQuer: Efficient Management of Inconsistent Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, Baltimore, Maryland, USA, June 14-16, 2005*, Fatma Özcan (Ed.). ACM, 155–166. https://doi.org/10.1145/1066157.1066176

[7] Ariel Fuxman and Renée J. Miller. 2005. First-Order Query Rewriting for Inconsistent Databases. In *Database Theory - ICDT 2005, 10th International Conference, Edinburgh, UK, January 5-7, 2005, Proceedings (Lecture Notes in Computer Science)*, Thomas Eiter and Leonid Libkin (Eds.), Vol. 3363. Springer, 337–351. https://doi.org/10.1007/978-3-540-30570-5_23

[8] Ariel Fuxman and Renée J. Miller. 2007. First-order query rewriting for inconsistent databases. *J. Comput. Syst. Sci.* 73, 4 (2007), 610–635. https://doi.org/10.1016/j.jcss.2006.10.013

[9] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. 2014. Clingo = ASP + Control: Preliminary Report. *CoRR* abs/1405.3694 (2014). arXiv:1405.3694 http://arxiv.org/abs/1405.3694

[10] Martin Gebser, Roland Kaminski, Arne König, and Torsten Schaub. 2011. Advances in *gringo* Series 3. In *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings (Lecture Notes in Computer Science)*, James P. Delgrande and Wolfgang Faber (Eds.), Vol. 6645. Springer, 345–351. https://doi.org/10.1007/978-3-642-20895-9_39

[11] Gianluigi Greco, Sergio Greco, and Ester Zumpano. 2003. A Logical Framework for Querying and Repairing Inconsistent Databases. *IEEE Trans. Knowl. Data Eng.* 15, 6 (2003), 1389–1408. https://doi.org/10.1109/TKDE.2003.1245280

[12] Phokion G. Kolaitis, Enela Pema, and Wang-Chiew Tan. 2013. Efficient Querying of Inconsistent Databases with Binary Integer Programming. *PVLDB* 6, 6 (2013), 397–408. https://doi.org/10.14778/2536336.2536341

[13] Paraschos Koutris and Jef Wijsen. 2015. The Data Complexity of Consistent Query Answering for Self-Join-Free Conjunctive Queries Under Primary Key Constraints. In *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, Tova Milo and Diego Calvanese (Eds.). ACM, 17–29. https://doi.org/10.1145/2745754.2745769

[14] Paraschos Koutris and Jef Wijsen. 2017. Consistent Query Answering for Self-Join-Free Conjunctive Queries Under Primary Key Constraints. *ACM Trans. Database Syst.* 42, 2 (2017), 9:1–9:45. https://doi.org/10.1145/3068334

[15] Paraschos Koutris and Jef Wijsen. 2018. Consistent Query Answering for Primary Keys and Conjunctive Queries with Negated Atoms. In *Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, Houston, TX, USA, June 10-15, 2018*, Jan Van den Bussche and Marcelo Arenas (Eds.). ACM, 209–224. https://doi.org/10.1145/3196959.3196982

[16] Paraschos Koutris and Jef Wijsen. 2019. Consistent Query Answering for Primary Keys in Logspace. In *22nd International Conference on Database Theory, ICDT 2019, March 26-28, 2019, Lisbon, Portugal (LIPIcs)*, Pablo Barceló and Marco Calautti (Eds.), Vol. 127. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 23:1–23:19. https://doi.org/10.4230/LIPIcs.ICDT.2019.23

[17] Paraschos Koutris and Jef Wijsen. 2020. Consistent Query Answering for Primary Keys in Datalog. *Theory of Computing Systems* (2020), 1–57. https://doi.org/10.1007/s00224-020-09985-6

[18] Paraschos Koutris and Jef Wijsen. 2020. First-Order Rewritability in Consistent Query Answering with Respect to Multiple Keys. In *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020*, Dan Suciu, Yufei Tao, and Zhewei Wei (Eds.). ACM, 113–129. https://doi.org/10.1145/3375395.3387654

[19] Nicola Leone, Gerald Pfeifer, Wolfgang Faber, Thomas Eiter, Georg Gottlob, Simona Perri, and Francesco Scarcello. 2006. The DLV system for knowledge representation and reasoning. *ACM Trans. Comput. Log.* 7, 3 (2006), 499–562. https://doi.org/10.1145/1149114.1149117

[20] Mónica Caniupán Marileo and Leopoldo E. Bertossi. 2010. The consistency extractor system: Answer set programs for consistent query answering in databases. *Data Knowl. Eng.* 69, 6 (2010), 545–572. https://doi.org/10.1016/j.datak.2010.01.005

[21] Jef Wijsen. 2012. Certain conjunctive query answering in first-order logic. *ACM Trans. Database Syst.* 37, 2 (2012), 9:1–9:35. https://doi.org/10.1145/2188349.2188351

[22] Jef Wijsen. 2019. Foundations of Query Answering on Inconsistent Databases. *SIGMOD Rec.* 48, 3 (2019), 6–16. https://doi.org/10.1145/3377391.3377393