# A Social Choice Theoretic Perspective on Database Aggregation

(Extended Abstract)

Francesco Belardinelli
Department of Computing, Imperial College London
Laboratoire IBISC, Université d'Evry
francesco.belardinelli@imperial.ac.uk

Umberto Grandi
IRIT, University of Toulouse
France
umberto.grandi@irit.fr

## KEYWORDS

Integrity Constraints; Axioms; Collective rationality

**ACM Reference Format:**
Francesco Belardinelli and Umberto Grandi. 2019. A Social Choice Theoretic Perspective on Database Aggregation. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

## 1 INTRODUCTION

Aggregating information coming from multiple sources is a long-standing problem in both knowledge representation and multi-agent systems (see, e.g., [28]). Depending on the chosen representation for the incoming pieces of knowledge or information, a number of competing approaches has seen the light in these literatures. Belief merging [21–23] studies the problem of aggregating propositional formulas coming from different agents into a set of models, subject to some integrity constraint. Judgment and binary aggregation [11, 12, 17] asks individual agents to report yes/no opinions on a set of logically-related binary issues – the agenda – in order to take a collective decision. Social welfare functions, the cornerstone problem in social choice theory (see, e.g., [2]), can also be viewed as mechanisms to merge conflicting information, namely the individual preferences of voters expressed in the form of linear orders over a set of alternatives. Other examples include graph aggregation [13], multi-agent argumentation [6–8], ontology merging [26], and clustering aggregation [15].

In this work we take a general perspective and represent individual knowledge coming from multiple sources as a profile of databases, modelled as finite relational structures [1, 25]. Our motivation lies in between two possibly conflicting views on the problem of information fusion. On the one hand, the study of information merging (typically knowledge or beliefs) in knowledge representation has focused on the design of rules that guarantee the consistency of the outcome, with the main driving principles inspired from the literature on belief revision. On the other hand, social choice theory has focused on agent-based properties, such as fairness and representativity of an aggregation procedure, paying attention as well to possible strategic behaviour by either the agents involved in the process or an external influencing source. While there already have been several attempts at showing how specific merging or aggregation frameworks could be simulated or subsumed by one another (see, e.g., [9, 14, 16, 18]), a more general perspective allows us to find a compromise between the two views described above.

*Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.* 

**Our Contribution.** We propose a number of rules for database aggregation, some inspired by existing ones from the literature on computational social choice and belief merging, and a new one adapted from representations of incomplete information in databases [24]. We first evaluate these rules axiomatically, using notions imported from the literature on social choice, to provide a first classification of the agent-based properties satisfied by our rules. Then, when integrity constraints are present, we study how to guarantee that a given aggregator "lifts" the integrity constraint from the individual to the collective level, i.e., the aggregated databases satisfy the same constraints as the individual ones.

## 2 DATABASES AND CONSTRAINTS

Our starting point is a set of finite relational structures on the same signature, coming from a group of agents or sources, and our research problem is how to obtain a collective database summarising the information received. Virtually all of the aggregation settings mentioned in the introduction (beliefs, graphs, preferences, judgments, …) can be represented as databases, showing the generality of our approach. Let us give the following basic definition:

*Definition 2.1 (Database Schema and Instance).* A *(relational) database schema* $\mathcal{D}$ is a finite set $\{P_1/q_1, \dots, P_m/q_m\}$ of relation symbols $P$ with arity $q \in \mathbb{N}$. Given database schema $\mathcal{D}$ and domain $U$, a $\mathcal{D}$*-instance* over $U$ is a mapping $D$ associating each relation symbol $P \in \mathcal{D}$ with a finite $q$-ary relation over $U$, i.e., $D(P) \underset{fin}{\subset} U^q$.

Properties and constraints on databases can be expressed as formulas of a suitable first-order language:

*Definition 2.2 (FO-formulas over $\mathcal{D}$).* Given a database schema $\mathcal{D}$, the formulas $\varphi$ of the first-order language $\mathcal{L}_{\mathcal{D}}$ are defined by the following BNF, where $P \in \mathcal{D}$, $x_1, \dots, x_q$ is a $q$-tuple of variables, as are $x, x'$:

$$\varphi \quad ::= \quad x = x' \mid P(x_1, \dots, x_q) \mid \neg\varphi \mid \varphi \to \varphi \mid \forall x \varphi$$

Let us illustrate one well-known class of integrity constraints on databases and its representation as a first-order formula. A *functional dependency* is an expression of type $\ell_1, \dots, \ell_k \mapsto \ell_{k+1}, \dots, \ell_q$. A database instance $D$ satisfies a functional dependency $\ell_1, \dots, \ell_k \mapsto \ell_{k+1}, \dots, \ell_q$ for predicate symbol $P$ with arity $q$ iff for every $q$-tuples $\vec{u}, \vec{u}'$ in $D(P)$, whenever $u_i = u'_i$ for all $i \leqslant k$, then we also have $u_i = u'_i$ for all $k < i \leqslant q$. If $k = 1$, we say that it is a *key dependency*. Clearly, any database instance $D$ satisfies a functional dependency $\ell_1, \dots, \ell_k \mapsto \ell_{k+1}, \dots, \ell_q$ iff it satifies the following FO-sentence:

$$\forall \vec{x}, \vec{y} \left( P(\vec{x}) \land P(\vec{y}) \land \bigwedge_{i \leqslant k} (x_i = y_i) \to \bigwedge_{k < i \leqslant q} (x_i = y_i) \right)$$

## 3 AGGREGATORS

We propose a number of rules for database aggregation, some of which are inspired by existing ones from the literature on computational social choice and belief merging. We fix a database schema $\mathcal{D}$ over a common domain $U$, and consider a *profile* $\vec{D} = (D_1, \ldots, D_n)$ of $n$ instances over $\mathcal{D}$ and $U$. Then, we define what is an aggregation procedure on such instances. We privilege computationally friendly aggregators, for which the time to determine the collective outcome is polynomial in the individual input received.

As an example, consider the following class of aggregators, well-known in judgment aggregation [10]: a *quota* rule is an aggregation rule $F$ defined via functions $q_P : U^q \to \{0, 1, \ldots, n+1\}$, associating each symbol $P$ and $q$-uple with a quota, by stipulating that $\vec{u} \in F(\vec{D})(P)$ iff $|\{i \leqslant n \mid \vec{u} \in D_i(P)\}| \geqslant q_P(\vec{u})$. $F$ is called *uniform* whenever $q$ is a constant function for all tuples and symbols.

Intuitively, if a tuple $\vec{u}$ appears in at least $q_P(\vec{u})$ of the initial databases, then it is accepted for symbol $P$. The well-known majority rule is a (uniform) quota rule for $q = \lceil (n+1)/2 \rceil$; while union and intersection are quota rule for $q = 1$ and $q = n$ respectively.

## 4 AXIOMS

Aggregation procedures are best characterised by means of axioms. In particular, we consider the following properties, where relation symbols $P, P' \in \mathcal{D}$, profiles $\vec{D}, \vec{D}' \in \mathcal{D}(U)^n$, tuples $\vec{u}, \vec{u}' \in U^+$ are all universally quantified.

**Anonymity** ($A$): for every permutation $\pi : \mathcal{N} \to \mathcal{N}$, we have $F(D_1, \ldots, D_n) = F(D_{\pi(1)}, \ldots, D_{\pi(n)})$.

**Independence** ($I$): if $N_{\vec{u}}^{\vec{D}(P)} = N_{\vec{u}}^{\vec{D}'(P)}$ then $\vec{u} \in F(\vec{D})(P)$ iff $\vec{u} \in F(\vec{D}')(P)$.

**Monotonicity** ($M$): if $\vec{u} \in F(\vec{D})(P)$ and for every $i \in \mathcal{N}$, either $D_i(P) = D'_i(P)$ or $D_i(P) \cup \{\vec{u}\} \subseteq D'_i(P)$, then $\vec{u} \in F(D')(P)$.

Combinations of the axioms above can be used to characterise some of the rules that we defined in Section 3. Some of these results, such as the following, lift to databases known results in judgement (propositional) aggregation.

Lemma 4.1. *An aggregation procedure satisfies A, I, and M iff it is a quota rule.*

## 5 COLLECTIVE RATIONALITY

We present a notion of *collective rationality* that aims to capture the appropriateness of a given aggregator $F$ w.r.t. some constraint $\varphi$ on the input instances $D_1, \ldots, D_n$. Let $\varphi$ be a sentence in the first-order language $\mathcal{L}_{\mathcal{D}}$ associated to $\mathcal{D}$, interpreted as a common constraint that is satisfied by all $D_1, \ldots, D_n$. Consider the following:

*Definition 5.1 (Collective Rationality).* A constraint $\varphi$ is lifted by an aggregation procedure $F$ if whenever $D_i$ satisfies $\varphi$ for all $i \in \mathcal{N}$, then also $F(\vec{D})$ satisfies $\varphi$. An aggregation procedure $F : \mathcal{D}(U)^n \to \mathcal{D}(U)$ is *collectively rational* (CR) with respect to $\varphi$ iff $F$ lifts $\varphi$.

Intuitively, an aggregator is CR w.r.t. constraint $\varphi$ iff it lifts, or preserves, $\varphi$. Consider the following:

*Example 5.2.* We now provide an illustrative example of first-order collective (ir)rationality with the majority rule. Consider agents 1 and 2 with database schema $\mathcal{D} = \{P/1, Q/2\}$. Two database instances are given as $D_1 = \{P(a), Q(a,b)\}$ and $D_2 = \{P(a), Q(a,c)\}$. Clearly, both instances satisfy the constraint $\varphi = \forall x(P(x) \to \exists y Q(x,y))$. However, their aggregate $D = F(D_1, D_2) = \{P(a)\}$, obtained by the majority rule, does not satisfy $\varphi$. This example, which can be considered a *paradox* in the sense of [17], shows that not every constraint in the language $\mathcal{L}_{\mathcal{D}}$ is collective rational w.r.t. majority, thus obtaining a first, simple negative result.

For the classes of aggregators presented, we investigate collective rationality with respect to arbitrary integrity constraints. As an example, consider the following result relating the quota rule aggregators with constraints of functional dependency:

Proposition 5.3. *A quota rule lifts a functional constraint iff for all relation symbols P occurring in the functional constraint we have that* $q_P > \frac{n}{2}$*, where n is the number of agents.*

As immediate applications of Prop. 5.3, the intersection rule clearly lifts any functional dependency, while the union lifts none.

## 6 RELATED WORK AND CONCLUSIONS

The closest approach to ours is the work of Baral *et al.* [3, 4] and Konieczny [20]. Baral *et al.* [4] considers the problem of merging information represented in the form of first-order theories, taking a syntactic rather than a semantic approach, and focusing on finding maximally consistent sets of the union of the individual theories received. In [20], the author applies techniques from belief merging to the equivalent problem of aggregating knowledge bases of first-order formulas, proposing a number of rules analysed axiomatically. Both contributions stem from a long tradition on combining inconsistent theories, especially in the domain of paraconsistent logics [5, 27]. However, all these approaches focus on merging syntactic representations (e.g., logic programs, first-order theories), while here we operate on semantical instances, i.e., databases. More recently, connections between social choice theory and database querying have been explored in [19], which enriches the tasks currently supported in computational social choice by means of a relational database, thus allowing for sophisticated queries about voting rules, candidates, and voters.

In this work we propose a framework for the aggregation of conflicting information coming from multiple sources in the form of finite relational databases. We propose a number of aggregators inspired by the literature on social choice theory, and adapt a number of axiomatic properties. We focus on a natural question which arise when dealing with the aggregation of databases. Specifically, we study what kind of integrity constraints are lifted by some of the rules we propose, i.e., what constraints are true in the aggregated database supposing that all individual input satisfies the same constraints.

**Acknowledgments**. U. Grandi acknowledges the support of the ANR JCJC project SCONE (ANR 18-CE23-0009-01). F. Belardinelli acknowledges the support of the ANR JCJC Project SVeDaS (ANR-16-CE40-0021-01).

## REFERENCES


[1] S. Abiteboul, R. Hull, and V. Vianu. 1995. *Foundations of Databases*. Addison-Wesley.

[2] K. J. Arrow. 1963. *Social Choice and Individual Values* (2nd ed.). John Wiley and Sons.

[3] C. Baral, S. Kraus, and J. Minker. 1991. Combining Multiple Knowledge Bases. *IEEE Transactions on Knowledge and Data Engineering* 3, 2 (June 1991), 208–220.

[4] C. Baral, S. Kraus, J. Minker, and V. S. Subrahmanian. 1992. Combining knowledge bases consisting of first-order theories. *Computational Intelligence* 8, 1 (1992).

[5] H. A. Blair and V. S. Subrahmanian. 1989. Paraconsistent Logic Programming. *Theoretical Computer Science* 68, 2 (Oct. 1989), 135–154.

[6] R. Booth, E. Awad, and I. Rahwan. 2014. Interval Methods for Judgment Aggregation in Argumentation. In *Proceedings of the 14th International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

[7] M. Caminada and G. Pigozzi. 2011. On judgment aggregation in abstract argumentation. *Autonomous Agents and Multiagent Systems* 22, 1 (2011), 64–102.

[8] W. Chen and U. Endriss. 2017. Preservation of Semantic Properties during the Aggregation of Abstract Argumentation Frameworks. In *Proceedings of the 16th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*.

[9] F. Dietrich and C. List. 2007. Arrow's Theorem in Judgment Aggregation. *Social Choice and Welfare* 29, 1 (2007), 19–33.

[10] F. Dietrich and C. List. 2007. Judgment Aggregation by Quota Rules: Majority Voting Generalized. *Journal of Theoretical Politics* 19, 4 (2007), 391–424.

[11] E. Dokow and R. Holzman. 2010. Aggregation of binary evaluations. *Journal of Economic Theory* 145, 2 (2010), 495–511.

[12] U. Endriss. 2016. Judgment Aggregation. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). Cambridge University Press.

[13] U. Endriss and U. Grandi. 2017. Graph Aggregation. *Artificial Intelligence* 245 (2017), 86–114.

[14] P. Everaere, S. Konieczny, and P. Marquis. 2015. Belief Merging versus Judgment Aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[15] A. Gionis, H.i Mannila, and P. Tsaparas. 2007. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data* 1, 1 (2007), 4.

[16] U. Grandi and U. Endriss. 2011. Binary Aggregation with Integrity Constraints. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.

[17] U. Grandi and U. Endriss. 2013. Lifting Integrity Constraints in Binary Aggregation. *Artificial Intelligence* 199–200 (2013), 45–66.

[18] E. Gregoire and S. Konieczny. 2006. Logic-based approaches to information fusion. *Information Fusion* 7, 1 (2006), 4 – 18. Logic-based Approaches to Information Fusion.

[19] B. Kimelfeld, P. G. Kolaitis, and J. Stoyanovich. 2018. Computational Social Choice Meets Databases. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*.

[20] S. Konieczny. 2000. On the Difference between Merging Knowledge Bases and Combining them. In *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR)*.

[21] S Konieczny, J Lang, and P Marquis. 2004. DA2 merging operators. *Artificial Intelligence* 157, 1 (2004), 49 – 79.

[22] S. Konieczny and R. Pino Pérez. 2002. Merging Information Under Constraints: A Logical Framework. *Journal of Logic and Computation* 12, 5 (2002), 773–808.

[23] P. Liberatore and M. Schaerf. 1998. Arbitration (or How to Merge Knowledge Bases). *IEEE Transactions on Knowledge and Data Engineering* 10, 1 (1998), 76–90.

[24] L. Libkin. 2015. How to Define Certain Answers. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*. The AAAI Press, 4282–4288.

[25] D. Maier, J. Ullman, and M. Vardi. 1984. On the Foundations of the Universal Relation Model. *ACM Transactions on Database Systems* 9, 2 (1984), 283–308.

[26] D. Porello and U. Endriss. 2012. Ontology merging as social choice: Judgment aggregation under the open world assumption. *Journal of Logic and Computation* 24 (11 2012), 1229–1249.

[27] V. S. Subrahmanian. 1992. Paraconsistent Disjunctive Deductive Databases. *Theoretical Computer Science* 93, 1 (1992), 115–141.

[28] F. van Harmelen, V. Lifschitz, and B. Porter. 2007. *Handbook of Knowledge Representation*. Elsevier Science, San Diego, USA.