

Non-monotonic Collective Decisions

Author

Cristani, M, Olivieri, F, Governatori, G

Published

2019

Conference Title

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Version

Accepted Manuscript (AM)

DOI

[10.1007/978-3-030-33792-6_24](https://doi.org/10.1007/978-3-030-33792-6_24)

Rights statement

© Springer Nature Switzerland AG 2019. This is the author-manuscript version of this paper. Reproduced in accordance with the copyright policy of the publisher. The original publication is available at www.springerlink.com

Downloaded from

<http://hdl.handle.net/10072/393002>

Griffith Research Online

<https://research-repository.griffith.edu.au>

Non-monotonic collective decisions

Matteo Cristani¹, Francesco Olivieri², Guido Governatori²

¹Department of Computer Science, University of Verona, Italy

²Data61-CSIRO, Brisbane Australia

Abstract. The social choice theory has focused in the past on the problem of devising methods to determine how individual preferences are transformed into collective ones. In some investigations, scholars provided methods for expressing the social choice function, that, given a set of individual preferences, computes the resulting collective choice. Other studies focused on determining under which conditions the social choice function is efficiently computable.

In this paper, we concentrate on the specific case of collective decisions, when we assume that the agents are *rational*: they do not express random preferences, and they do not make random choices. In this context, we define four logical problems derived and study their computational complexity: (1) Determining the rationality of a given choice, (2) Establishing a possible rational maximal subset of a given choice, (3) Computing the votes on a rational proposal, and (4) Determining a priori the winning conditions of a given rational choice.

1 Introduction

When groups of individuals collaborate, it is required that the individual choices are combined into collective decisions. A *choice* and a *decision* are not the same process. A choice is the “*expression of a preference*”, whilst a decision is the “*deliberation of a choice*”. For instance, we can have the preference of caviar and champagne for lunch, but not the money to buy them. We thus decide to have a club sandwich with a glass of wine, that is a valid (though not the optimal) alternative. This is a common sense notion of choice and decision for individuals, but does this apply to collectivities in the same way?

In the above simple scenario, decisions can be graded, based on money availability. Rational are those alternatives that can be afforded. In many cases this is not possible because some of alternatives are unacceptable, for they violate general assumptions (rules) about the world. Suppose that we have to choose what clothing to wear while going out for dinner. It would be acceptable, for the restaurant’s dress code, to wear a suit with a tie, or a tuxedo with a bowtie. We can make a choice between two groups of alternatives: $\{suit, tuxedo\}$, and $\{tie, bowtie\}$. Important here is to note that not all (four) combinations are acceptable: wearing a suit with a bowtie, or a tuxedo with a tie are *irrational* choices, as they do not match the dress code.

Both decision making processes devised above, in which we have measures on the alternatives, or conditions to exclude some of them, have been dealt with in *Decision Theory*, the field of study that focuses on the problems of making decisions in presence of multiple, and possibly conflicting, criteria. The combination of acceptability based on criteria matching and weights is incorporated in the well-known notion of *Pareto*

Optimality, where an alternative is said to be optimal when all the necessary criteria are matched and weighted with the maximal value among the admissible alternatives.

Although this field has been studied in depth (including studies on computational aspects of the above sketched problems), the problem of rationality of choices in collective decision making still lacks literature in the fields of multi-agent systems, as well as computational logic. A conceptualisation of collective decision making is provided by the *Social Choice Theory* [3], where individual choices are combined into collective decisions. The individual choices are analysed in order to guarantee social welfare, by extension of the notion of Pareto optimality, called *Social Choice Function*. A Social Choice Function takes the individual choices of a collectivity, and deliberates a decision. A well known result of Social Choice Theory is the famous *Arrow's Impossibility theorem*, stating that there is no Social Choice Function able to guarantee the social welfare (the satisfaction of members of a collectivity) under certain general conditions.

Collective decisions are more complex to make than individual ones. In collective decision making, we need not only to evaluate the alternatives, but also to deliberate one choice that is *accepted* by the individuals who collaborate. In this paper, we deal with the problem of making collective decisions with some degree of *rationality*, namely in a way that not only provides choices accepted by the collectivity, but also compatible with common sense, in particular with a commonly accepted set of general rules considered reasonable by the collectivity.

In Social Choice Theory, scholars usually assume that the agents act in a rational way: their choices are rational to some extent. One way of relaxing this unrealistic assumption (agents have, many times, irrational choices) is by assuming that the choices need to satisfy some logical constraints: a rational choice should be consistent with some logical assumptions. We hence devote our attention to non-monotonic logic systems, as they can cope with exceptions and potentially conflicting information.

If we assume that a set of individuals collaborate, then it is natural to imagine that they have a commonly accepted background. For instance, they share methods as in scientific collaborations, or ethical principles as in social movements, or legal background as in political activities. We envision a system in which collaborating individuals share the common knowledge with an intelligent system able to perform reasoning on the choices of the individuals against the common background, therefore helping the decision process. Such a system can be viewed as a voting system, as we show in the rest of the paper.

We assume a form of rationality that is of *higher order*: when a choice is conflicting with the commonly accepted background, then it is not considered and can be reduced to an acceptable subset. We show that if a system of decision support is able to filter the choices of the individuals acting on a platform as sketched above, it would be a valid method to improve the welfare in the system itself, always within the theoretical limitations offered by the Arrow's Impossibility theorem.

We now introduce the running example of this work, to show how the logical framework we are going to develop takes these aspects into consideration, and specifically on the concept of choice.

Example 1. A group of friends (Alice, Bob, Jill, John, Julia, Mary, and Mark) will spend the evening together, and needs to decide: (1) the food for dinner, (2) the beverages to drink, and (3) the movies to watch. The possible choices are as follows. (1) Dinner: *Indian*,

Italian, and Chinese. (2) Beverages: beer, wine, and coke. (3) Movies: *The Avengers* (A), *The Great Gatsby* (B), *The Hitchhikers' Guide to the Galaxy* (C), and *Red* (D). The individual choices are expressed in Table 1, while Figure 1 depicts such a scenario.

Agent	Dinner	Drink	Film
Alice	Italian	Coke	(C)
Bob	Indian	Beer	(D)
John	Italian	Wine	(B)
Mary	Chinese	Beer	(A)
Jill	Italian	Coke	(C)
Mark	Indian	Wine	(A)
Julia	Italian	Wine	(C)

Table 1: The agents' individual choices

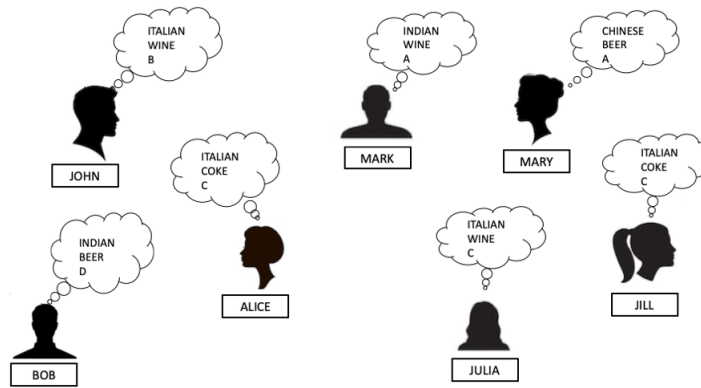


Fig. 1: The choices over the possible components for the a group of friends.

We assume that, despite the individual preferences, the group has to come up with collective decisions: they all must eat the same food, drink the same beverage, watch the same movie. These assumptions, whilst potentially unrealistic, are useful to highlight the results of this work.

When we deal with collective decisions, the Pareto optimal of one agent can be different from the Pareto optimal of another agent. Accordingly, we cannot try to use *object rank* as a means for solving the problem of optimal choice, as shown in [10].

We can combine the choices mentioned in Example 1 in $3 \times 3 \times 4 = 36$ different ways, but we have only 7 actual choices (precisely 1 for each individual). We may decide, for instance, to order Italian, drink wine and watch *The Hitchhikers's Guide to the Galaxy*. This corresponds to one specific combination of choices obtained by sequencing the choices on the columns from Dinner to Movie (as shown in [16]) and choose the most preferred of each column. Now, Julia is completely satisfied with this choice. To the

contrary, Alice has preferred dinner and film but not beverage, Bob has only preferred drink, John has only preferred dinner, Jill has preferred food and film, Mary has only preferred beverage, and Mark has no preference satisfied.

A method that makes a decision while preserving the *social welfare* should be able to guarantee some degree of satisfaction to each of the individuals. Social choice theory is a widely recognised method to provide room for those processes in which a group of individuals makes a collective decision. The problem of Social Choice Function has been dealt with different perspectives and for different purposes. For many of these investigations, the base of the study is some sort of assumption on the existence of one kind of *collective rationality* that, to some extent, guarantees the applicability of a Social Choice Function. In other terms, it provides room to a decision process that, while assuming individual rationality, extends it by guaranteeing collective rationality. We can consider these computational approaches, and specifically those cases in which the computation of the collective decision has been valued for multi-agent systems, *Artificial Intelligence* for collective decision making. Conversely, the idea of rationality is considered for systems that make use of Social Choice Function to compute an acceptable collective decision, and therefore belong to the field of Decision Theory.

In this paper, we look at the rationality of choices made both individually, and collectively, in a *logical framework* where we delimit the perimeter of rationality to the notion of *plausibility*. We consider rational choices in a very general way, and then embed them onto a well-known logical framework, Defeasible Logic, that has been used for non-monotonic reasoning, as well as for modelling multi-agent systems in many research efforts so far [15,13]. Based on this framework, we consider a general problem to be addressed: Is an individual choice *rational*? This problem can be shown to correspond to the notion of *rational voting*, when an individual evaluates their choice against a proposal in a collaboration setting. In the rest of the paper, we also discuss consequences of this setting: What happens when a choice is irrational?

We devise a method for reducing it to a rational subset. We show how a decision process can be implemented in one specific case: the *majority principle* as Social Choice Function. We also study a method for deciding whether a proposal to be advanced for individual choices is a valid collective decision under the *majority principle*. These problems result hard from a computational viewpoint.

The rest of the paper is organised as follows: Section 2 discusses the employed formalism, and shows how to adapt it to the context. Section 3 defines four computational problems, provides the corresponding algorithms, and proves their complexities. Section 4 discusses related work, while Section 5 summarises, and sketches further work.

2 Basic definitions

The basic setting of the theory presented in this paper is a multi-agent system: agents of the system employ a commonly accepted background in the form of a *defeasible theory*.

The background is formed by indisputable facts and commonly accepted rules that are considered plausible by every agent involved in the discussion. This theory is specified by a Defeasible Logic. Defeasible Logic is a rule-based skeptical approach to non-monotonic reasoning. It is based on a logic programming-like language and is a simple, efficient, but

flexible formalism capable of dealing with many intuitions of non-monotonic reasoning in a natural and meaningful way [2].

Set PROP defines propositional atoms, while Lab is a set of arbitrary labels. The set $\text{Lit} = \text{PROP} \cup \{\neg p \mid p \in \text{PROP}\}$ denotes the set of literals. The *complement* of a literal q is denoted by $\sim q$; if q is a positive literal p , then $\sim q$ is $\neg p$, and if q is a negative literal $\neg p$ then $\sim q$ is p .

A defeasible theory D is a tuple $(F, R, >)$. $F \subseteq \text{Lit}$ are the facts which are always-true pieces of information. R contains three types of rules: strict rules, defeasible rules, and defeaters. A rule is an expression of the form ' $r : A(r) \leftrightarrow C(r)$ ', where r is the name of the rule, the *arrow* $\leftrightarrow \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}$ is to denote, resp., strict rules, defeasible rules and defeaters, $A(r)$ is the (finite) set of antecedents of the rule, and $C(r)$ is its consequent (a single literal). A strict rule is a rule in the classical sense: whenever the antecedent holds, so is the conclusion. A defeasible rule is allowed to assert its conclusion unless there is contrary evidence to it. A defeater is a rule that cannot be used to draw any conclusion, but can provide contrary evidence to complementary conclusions. Lastly, $> \subseteq R \times R$ is the *superiority relation*, a binary and antisymmetric relation, which exact purpose is to solve conflicts among rules with opposite conclusions. We use the following abbreviations on R : R_s is to denote the set of strict rules in R , R_{sd} the set of strict and defeasible rules, and $R[q]$ is the set of rules in R s.t. $C(r) = q$.

A *derivation* (or *proof*) is a finite sequence $P = P(1), \dots, P(n)$ of *tagged literals* of the type $+\Delta q$ (q is definitely provable), $-\Delta q$ (q is definitely not provable, or refuted), $+\partial q$ (q is defeasibly provable) and $-\partial q$ (q is defeasibly refuted). The proof conditions below define the logical meaning of such tagged literals. Given a proof P , we use $P(n)$ to denote the n -th element of the sequence, and $P(1..n)$ to denote the first n elements of P . $\pm\Delta$ and $\pm\partial$ are called *proof tags*. Given $\# \in \{\Delta, \partial\}$ a proof tag, the notation $D \vdash \pm\#q$ means that there is a proof P in D such that $P(n) = \pm\#q$ for an index n .

In what follows, we only present the proof conditions for the positive tags: the negative ones are obtained via the principle of *strong negation*. This is closely related to the function that simplifies a formula by moving all negations to an inner most position in the resulting formula, and replaces the positive tags with the respective negative tags, and the other way around.

The proof conditions for $+\Delta$ describe just forward chaining of strict rules.

$+\Delta$: If $P(n+1) = +\Delta q$ then

- (1) $q \in F$, or
- (2) $\exists r \in R_s[q]$ s.t. $\forall a \in A(r). +\Delta a \in P(1..n). P(1..n)$.

Literal q is definitely provable if either (1) is a fact, or (2) there is a strict rule for q , whose antecedents have all been definitely proved. Literal q is definitely refuted if (1) is not a fact, and (2) every strict rule for q has at least one definitely refuted antecedent.

The conditions to establish a defeasible proof $+\partial$ have a structure similar to arguments.

$+\partial$: If $P(n+1) = +\partial q$ then

- (1) $+\Delta q \in P(1..n)$, or
- (2) (2.1) $-\Delta \sim q \in P(1..n)$ and
- (2.2) $\exists r \in R_{sd}[q]$ s.t. $\forall a \in A(r) : +\partial a \in P(1..n)$, and
- (2.3) $\forall s \in R[\sim q]$. either

- (2.3.1) $\exists b \in A(s) : -\partial b \in P(1..n)$, or
(2.3.2) $\exists t \in R$ s.t. $\forall c \in A(t) : +\partial c \in P(1..n)$ and $t > s$

A literal q is defeasibly proved if, naturally, it has already strictly proved. Otherwise, we need to use the defeasible part of the theory. Thus, first, the opposite literal cannot be strictly proved (2.1). Then, there must exist an *applicable* rule supporting such a conclusion (2.3)¹. We now need to check that all counter-arguments (i.e., the rules supporting the opposite conclusion) are either *discarded* (condition (2.3.1)), or defeated by a stronger, applicable rule for the conclusion we want to prove (2.3.2).

We introduce here a defeasible theory that is the background of Example 1.

Example 2. We assume that the group of friends described in Table 1 has already decided to get together for the evening, and we represent that with the fact ‘*evening_together*’. Further on, we introduce the rules that express the components of the evening:

- R.1: *evening_together* \rightarrow *have_dinner*
- R.2: *evening_together* \rightarrow *drink_beverages*
- R.3: *evening_together* \rightarrow *watch_film*.

Each of the admissible choices for each of the components is provided in the form of a single defeasible rule, derived from the literals *have_dinner*, *drink_beverages*, *watch_film*.

- r.1: *have_dinner* \Rightarrow *italian_cuisine*
- r.2: *have_dinner* \Rightarrow *indian_cuisine*
- r.3: *have_dinner* \Rightarrow *chinese_cuisine*
- r.4: *drink_beverages* \Rightarrow *coke*
- r.5: *drink_beverages* \Rightarrow *beer*
- r.6: *drink_beverages* \Rightarrow *wine*
- r.7: *watch_film* \Rightarrow *The_avengers*
- r.8: *watch_film* \Rightarrow *The_great_Gatsby*
- r.9: *watch_film* \Rightarrow *The_hitchhikers_guide_to_the_galaxy*
- r.10: *watch_film* \Rightarrow *Red*.

Moreover, we introduce exclusion rules, to describe the fact that each of the dining, drinking and film options cannot be shared.

- r.11: *italian_cuisine* \Rightarrow \neg *indian_cuisine*
- r.12: *italian_cuisine* \Rightarrow \neg *chinese_cuisine*
- r.13: *indian_cuisine* \Rightarrow \neg *italian_cuisine*
- r.14: *indian_cuisine* \Rightarrow \neg *chinese_cuisine*
- r.15: *chinese_cuisine* \Rightarrow \neg *italian_cuisine*
- r.16: *chinese_cuisine* \Rightarrow \neg *indian_cuisine*
- r.17: *coke* \Rightarrow \neg *beer*
- r.18: *coke* \Rightarrow \neg *wine*
- r.19: *beer* \Rightarrow \neg *coke*
- r.20: *beer* \Rightarrow \neg *wine*

¹ We say that a rule is *applicable* when all its antecedents have been proved within the current derivation step. Symmetrically, we say that a rule is *discarded* if at least one of its antecedents has been defeasibly refuted.

- r.21: $wine \Rightarrow \neg coke$
- r.22: $wine \Rightarrow \neg beer$
- r.23: $The_avengers \Rightarrow \neg The_great_Gatsby$
- r.24: $The_avengers \Rightarrow \neg The_hitchhikers_guide_to_the_galaxy$
- r.25: $The_avengers \Rightarrow \neg Red$
- r.26: $The_great_Gatsby \Rightarrow \neg The_avengers$
- r.27: $The_great_Gatsby \Rightarrow \neg The_hitchhikers_guide_to_the_galaxy$
- r.28: $The_great_Gatsby \Rightarrow \neg Red$
- r.29: $The_hitchhikers_guide_to_the_galaxy \Rightarrow \neg The_avengers$
- r.30: $The_hitchhikers_guide_to_the_galaxy \Rightarrow \neg The_great_Gatsby$
- r.31: $The_hitchhikers_guide_to_the_galaxy \Rightarrow \neg Red$
- r.32: $Red \Rightarrow \neg The_avengers$
- r.33: $Red \Rightarrow \neg The_great_Gatsby$
- r.34: $Red \Rightarrow \neg The_hitchhikers_guide_to_the_galaxy$.

We have some rules for combining food, drink and film, shared by the friends having evening together.

- r.35: $italian_cuisine \Rightarrow wine$
- r.36: $wine \Rightarrow \neg indian_cuisine$
- r.37: $The_avengers \Rightarrow chinese_cuisine$

Agents express their preferences on the shared theory by means of choices as in Definition 1. Choices may be, or not, rational, in the sense that we provide in Definition 2.

Definition 1. (Choice) Given a defeasible theory T , that constitutes the common background of a multi-agent system, a choice, is a finite set of literals that does not contain any contradiction (e.g., both l and $\neg l$).

Definition 2. (Rational choice) We say that a choice C , expressed by an agent is rational with respect to a given background theory T , when it is possible to set the superiority relation on T , to define a theory T' in such a way that all the literals in C are defeasibly proved by T' .

We adopt here the definition of *choice* expressed in the Social Choice Theory literature. Consider a setting in which an individual advances a proposal for the decision, that consists in expressing their choice, in order to make the individuals collaborating in the system *vote* for that choice. Voting consists in choosing a proposal advanced by another individual who expresses her choice to the individuals collaborating in the system. To devise rationality of the vote we consider the notion of *compatibility*.

Given two choices C_1 and C_2 , we hereafter define the notion of *compatibility*.

- When $C_1 \subseteq C_2$, we say that C_1 is *strongly compatible* with C_2 .
- When $C_2 \subseteq C_1$, we say that C_1 is *weakly compatible* with C_2 .

hgConsider an agent that has choice C and votes for proposal P . If $P \subseteq C$ (the proposal is strongly compatible with the choice), the agent then obtains *only things they desire*, but possibly not everything (when $C \setminus P \neq \emptyset$). Conversely, if $C \subseteq P$ (P is weakly compatible with C), now the agent obtains *everything they desire*, but possibly also things that they did not include in their choice.

In fact, if an agent makes a choice that contains a literal l , they cannot vote for a rational proposal P that contains the opposite literal $\neg l$, their choice is strongly compatible with: this means that P contains both l and $\neg l$, therefore the proposal is irrational.

Guaranteeing satisfaction of the individuals under strong compatibility is more difficult than it is with weak one. When the compatibility is strong, and two individuals have no elements in common on their choices, then only the empty proposal is accepted by both. The setting we devise is as follows:

- Every agent has their own *choice*;
- Some agents advance their *proposals*, that consist in expressing their own choice;
- Every agent *votes* for one of the proposals, or none of them.

To illustrate how the model we propose in this paper works, we provide a formalisation of Example 1.

Example 3. Choices introduced by Table 1 can be expressed as follows:

- Alice {italian_cuisine, coke, The_hitchhikers_guide_to_the_galaxy}
- Bob {indian_cuisine, beer, Red}
- John {italian_cuisine, wine, The_great_Gatsby}
- Mary {chinese_cuisine, beer, The_avengers}
- Jill {italian_cuisine, coke, The_hitchhikers_guide_to_the_galaxy}
- Mark {indian_cuisine, wine, The_avengers}
- Julia {italian_cuisine, wine, The_hitchhikers_guide_to_the_galaxy}

Example 4. Rational choices are introduced by the exclusion and the combination rules given in Example 2. In particular, Alice and Jill’s choices conflict with rule r.35, while Mark’s choice conflicts with rules r.36 and r.37. Therefore, the only rational choices can be Bob’s, John’s, Mary’s, and Julia’s. To settle the choice of Bob, we need to guarantee that a rule deriving *indian_cuisine* prevails over the rules deriving \neg *indian_cuisine*, and the same for *beer* and *Red*. This is done by the following superiority relation, superimposed on the background.

In Section 3, we use the above defined setting to discuss computational problems related to them.

Agent/Superiority	Agent/Superiority	Agent/Superiority	Agent/Superiority
Bob: r.2 > r.11	John: r.1 > r.13	Mary: r.3 > r.12	Julia: r.1 > r.13
Bob: r.2 > r.16	John: r.1 > r.15	Mary: r.3 > r.14	Julia: r.1 > r.15
Bob: r.5 > r.17	John: r.6 > r.18	Mary: r.5 > r.17	Julia: r.5 > r.16
Bob: r.5 > r.22	John: r.6 > r.20	Mary: r.5 > r.22	Julia: r.5 > r.17
Bob: r.10 > r.25	John: r.8 > r.23	Mary: r.7 > r.26	Julia: r.9 > r.24
Bob: r.10 > r.28	John: r.8 > r.30	Mary: r.7 > r.29	Julia: r.9 > r.27
Bob: r.10 > r.31	John: r.8 > r.33	Mary: r.7 > r.32	Julia: r.9 > r.34

Table 2: The choices listed in Example 1.

3 Collective decision making

We model the notion of *rational voting* by means of the setting devised in Section 2. In this section, we provide some relevant computational problems related to the introduced notion of rationality in vote and proposal expressions.

Definition 3. *Given a defeasible theory T that constitutes the commonly accepted background of a multi-agent system, and a choice C expressed by one of the agents, the problem of deciding whether C is rational w.r.t. T is named the Rationality Detection Problem (RDP).*

RDP can be solved by superimposing a superiority relation on the theory T , so that T derives each literal in the choice of the agent. To do so, since T can already have a superiority relation, it may be necessary to *revise* such a superiority relation until we either: obtain a theory that actually derives all the literals, or conclude that such a revision does not exist, and therefore establish that the choice is not rational.

Comparing this problem to other ones analogously defined, as in [12,11], we can prove the statement in Theorem 1. In the following, when referring to the problem of revising preference in a defeasible theory to derive one literal, we name this problem the *Preference Revision Problem*. That problem has been proven to be NP-complete in [11].

Theorem 1. *RDP is NP-complete.*

Proof. We prove this by showing that RDP can be polynomially reduced to the Preference Revision Problem of [11]. Given a set of literals, we aim at proving that these are derivable from the commonly accepted theory, once revised the superiority relation. We have a polynomial method on deterministic machine to establish whether a given superiority relation (superimposed on the commonly accepted theory) derives the whole set of literals, as this is the known problem of computing the extension of a defeasible theory, which is linear in the number of literals and rules in a theory. In [11], the authors provide an oracle for polynomially computing the superiority relation revision on a non-deterministic machine to derive a single literal, in the same configuration. Nevertheless, the number of literals does not influence the behaviour of the oracle. In fact, the oracle chooses the correct combination of superiority pairs for opposite literals. The number of these choices is, in the worst case, quadratic in the number of literals. This proves that the RDP problem is in NP. Moreover, the computation of the revision as shown above can be transformed into the revision of preferences in a linear number of steps, and this completes the proof that RDP is NP-complete.

We devise here a brute-force method to solve RDP on deterministic machines. Consider all the possible superiority relation pairs generated by combining every literal that appears at the head of at least one rule in the positive form, and in at least one rule in the negative form. In the worst case, these pairs are of the same order of the number of literals. Consequently, the number of rules obtained by the full combinatorics of the rules is $O(2^n)$ with n number of literals in the choice. For each combination computed as described above, we compute the extension of the theory, and we then check whether the choice is a subset of that extension. If we find such a combination, we have constructed the superiority relation and solved the RDP Problem.

Once a voter has analysed their choice, to evaluate the meaningfulness of their proposal they want to advance, and to realise whether that the choice is irrational, one possible reformulation of the problem may be: can I reduce the requests in the choice to a subset that results rational? If so, can I do this in a computationally effective manner? The problem defined informally above is formalised in Definition 4.

Definition 4. *Given a defeasible theory T , that constitutes the commonly accepted background of a multiple agent system, and an irrational preference C expressed by one of the agents, the problem of computing the maximal rational subsets of C is named the Rationality Maximisation Problem (RMP).*

We can solve the problem with a rather natural technique: we systematically remove one literal at a time from the original choice until we obtain a subset that results rational. Once we have such a subset, we continue to remove literals from the original choice until we have completed the process for one single literal. We further look at pairs of literals, but only for those that did not succeed as singletons. When concluded, we now proceed by considering triples, excluding those that contain one pair already included. This step ends when we reach a number k of elements to be removed, for which no set of elements can be removed that does not contain a subset already included in the step $k - 1$. Once we have reached the step $k = n - 1$, we stop. At that step it would not be possible to find any further subset to delete.

Example 5. If we look at Alice's choice, the obvious selection is the literal *coke*, that, once removed, provides a rational choice, consisting in the expression: my preference is for Italian food, and I wish to watch the Hitchhikers Guide to the Galaxy, but I have no preference for beverages.

Theorem 2. *RMP is EXPTIME.*

Proof. RMP cannot be easily reduced to a NP-complete problem, and it has a non-polynomial space occupancy, due to the need to consider, in the worst case scenario, a number of subsets of the same size of the subsets of the choice, that is $O(2^n)$, with n number of literals in the choice. Also the size of the output of RMP can be exponential. We thus perform a total of steps that is bounded by the number of possible subsets of the choice. For each of these subsets, we can perform the algorithm sketched in the discussion above, to perform in an exponential time the correctness of the choice. Overall, we shall have a sequence of exponential size ($O(2^n)$) of exponential steps (each of $O(2^n)$), making therefore the result an exponential size of the original size of the choice ($O(2^{2^n}) = O(2^n)$).

The implementation of the method devised in the proof of Theorem 2 is described in Algorithm 1.

Let us step forward onto the other level of the problems exposed here. When the choices are known to be rational, as in the case in which every individual has made their choice compared against the background theory by means of the method devised above, we can formulate the notion of Definition 5.

Data: An irrational choice, a background defeasible theory.

Result: The set of all maximal subsets of the input choice that are rational.

```
for Every subset  $E$  of  $C$  of size  $i$ , incrementing  $i$  at each step do
  Consider  $E$  only when no subset  $E'$  recorded at step  $(i-1)$  is contained in  $E$ ;
  if  $C-E$  is rational then
    | record  $C-E$  and  $E$ 
  end
  if At step  $i$  we did not add elements to the collection then
    | break
  end
end
```

Return All the recorded subsets;

Algorithm 1: RMPComp: an algorithm to compute maximal rational choices from an irrational one.

Definition 5. Given a defeasible theory T that constitutes the commonly accepted background of a multiple agent system, and a finite set S of rational choices expressed as proposals, the problem of determining the result of the expression of rational choices by the agents, by applying a social choice function f , able to transform the individual votes for the proposals into a collective decision is called the Collective Decision Problem (CDP).

For this case, the input to Algorithm 2 are choices that are assumed to be rational, and so do the proposals to be voted.

Data: A set of rational choices, a set of rational proposals, a background defeasible theory, a parameter for Strong/Weak comparability.

Result: The majority choices, if any.

```
for every choice  $C$  do
  for every proposal  $P$  do
    | if  $C$  is Strongly/Weakly compatible with  $P$  then
      | | add a vote for  $P$ 
    | end
  end
end
```

Return All the proposals that received more than 50% of the votes;

Algorithm 2: RatVot: an algorithm to establish majority applied to MAS.

Theorem 3. The CDP is polynomially solvable on rational proposals and rational votes, for the majority Social Choice Function.

Proof. The proof that the algorithm RatVot correctly computes all, and only, the proposals that are rational and can pass the majority Social Choice Function filter is straightforward, and left therefore to the reader. The Algorithm clearly steps on a number of elements that are the choices, and a number of proposals. If we consider these as inputs, the computation is $O(p \cdot c)$ where p is the number of advances proposals, and c the number of rational choices involved in the vote. The claim is a direct consequence of the structure of the algorithm where two cycles are nested to each other of length respectively p and c .

To see how some proposals are rationally voted, we go back to Example 1.

Example 6. The friends of Example 1 vote in a rational way for three alternatives: [italian_cuisine, wine, The_hitchhikers_guide_to_the_galaxy] (A1, advanced by Julia), [indian_cuisine, beer, Red] (A2, advanced by Bob) and [chinese_cuisine, beer, The_avengers] (A3, advanced by Mary). The rational choices are obtained by reducing the expectations of Alice and Jill to dinner and film choices, and for Mark to film. Votes are in the third column. Note that the expectations of Alice and Jill are the only maximal subsets we can consider, whilst Mark may choose to reduce the expectations in two other ways: (1) only food choice to *Indian cuisine* (and thus voting for A2), or (2) only to *wine* (and thus voting for A1). If we value *strong* compatibility, and consider proposals for *Indian cuisine* and *beer* (B1, advanced by Bob), and for *Chinese cuisine* and *The Avengers* (B2, advanced by Mary), we have the votes on fourth column.

Agent	Choice	Vote (w)	Vote (s)
Alice	{Italian cuisine, The hitchhikers guide to the galaxy}	A1	-
Bob	{Indian cuisine, beer, Red}	A2	B1
John	{Italian cuisine, wine, The great Gatsby}	-	-
Mary	{Chinese cuisine, beer, The avengers}	A3	B2
Jill	{Italian cuisine, The hitchhikers guide to the galaxy}	A1	-
Mark	{The avengers}	-	B2
Julia	{Italian cuisine, wine, The hitchhikers guide to the galaxy}	A1	-

Table 3: The rational votes in Example 1.

The winning alternatives are A1, and B2.

RMP problem is interesting from a methodological viewpoint. If a program to be voted is not rational, we can reduce it to a rational maximal subset. Once we have obtained a reduction of the choice, it becomes interesting to understand whether it is possible to aggregate a majority around the proposal. This problem is formalised in the definition below.

Definition 6. *Given a defeasible theory T , that constitutes the commonly accepted background of a multi-agent system, and the choices expressed by the agents in the system, the problem of computing a rational proposal that, under a given Social Choice Function f , satisfies f , is named the Winning Proposal Definition Problem (WPDP).*

The outcome of a voting system as devised here is not one single proposal that wins. We may observe also empty outcomes, without any winning proposals, or multiple outcome, where more than one proposal wins. Clearly, in both these cases we need some further selection process, but just the first step, under the control of rationality, is more complex than single winner systems (see [14] for the most general case).

Theorem 4. *WPDP is NP-hard for the majority Social Choice Function.*

Proof. To prove the claim, we need to devise a method that results exponential on deterministic machines. A simple approach consists in using the same method of

rational choices, and subsequently computing votes for the single proposal by means of Algorithm 2. We consider a rational proposal, and try to see whether this has a majority consensus. If this is the case, then the algorithm has finished. Therefore, if we have an oracle that can compute a proposal (the number of possible proposals is evidently exponential in the number of literals), then, by means of the mentioned algorithm, we can compute the solution in polynomial time. This proves the claim.

It might appear interesting to analyse the rationality in presence of revision operators able to introduce, eliminate, and/or substitute rules. However, this configuration is not harder than the one we described here. Assume that we wish to derive a set of literals, and that we can manipulate the rules. We can simply introduce one rule for each of the desired literals derived from the set of facts. This setting can always be obtained by the revision of the rules. Formally, we introduce an extended notion of rationality of a choice.

Definition 7. *A choice is r-rational (rational including revision of rules) when there is a revision of the rules that derives all the literals in the choice.*

r-RDP is the problem of deciding whether a choice is r-rational on a given background. The above reasoning can be used to derive the following.

Proposition 1. *r-RDP is polynomially solvable.*

4 Related work

In the recent past, many scholars have dealt with multi-agent systems from the viewpoint of preference aggregation and social choice theory. In [7], the authors deal with the basic issue that we addressed here: the collective decision problem in presence of varying preference expressions. The basic result of the investigation is that the application of multiple criteria methods, along with different methods to aggregate preferences, is generally very effective in defining a collective process of decision. What is missing in that study is that the authors do not address the specific problem of *logical analysis* of admissible choices, and the computational aspects of this logical analysis. This approach has been followed in this paper.

The field of social choice has been expanding rapidly as a research topic in the wider community of AI, and specifically in the multi-agent systems, especially with respect to socio-technical systems, where humans and bots can interact at the same level. Social choice theory is a very wide theme (for a rather complete analysis of the viewpoints developed in this field see [4]). Many investigations have focused upon the problems of how to define the social choice function, and on the difficulties determined by certain conditions that are superimposed on the social function itself, to the most famous negative result in this field, the Arrow's impossibility theorem [3], that has been investigated in terms of consequences in multi-agents systems widely as well [10,8].

There is also a wide literature in computational communities including multi-agent systems, that observes phenomena such the one we discussed, to an initial extent, in Section 3, when we introduced the Winning Proposal Definition Problem. In particular, the authors of [6] have dealt with the problem of *election manipulation*, a very complex

topic that also deserves consideration in multi-agent systems. Similarly, the manipulation of coalition has been studied from many different viewpoints (see [19] for an example).

In particular, there are many important recent investigations regarding *preferences* and their underlying mathematics [1,7,16,9], and on their specific applications to AI, as generally discussed by Rossi *et al.* in [17], and then applied to the definition of social choice problems [5,18,16].

Overall, we may look at the current literature regarding revision of preferences in non-monotonic logic, as an immature research topic, especially when applied to multi-agent systems. Scholars have dealt with many different aspects of social choice, but the notion of rationality, as developed in this investigation, is rather novel in the community of agents.

5 Conclusions and further work

This work introduced a method that can be used for implementing a form of rationality in collective decisions, based upon the notion of *rational choice*, and the concept of *rational voting*. The approach we adopted is completely centred on the revision of preferences in the non-monotonic setting of Defeasible Logic.

We showed that some relevant problems defined on the above mentioned framework are computationally hard, and that one specific problem, under certain conditions, is polynomially solvable. We also showed that the intrinsic computational complexity of the approach defines the worst case scenario: the revision of rules reduces the complexity to polynomial.

Further work may include many different aspects, including, in particular, an ample analysis of the simplified sub-cases for which some of the problems introduced in this paper have better computational complexity. In the second step, the investigation should consider extensions to the language. There are several possible extensions, including: (1) the introduction of *degrees of acceptance* in the literals of choices, (2) the introduction of *priorities* among the elements of one choice, and (3) the optimisation of the superiority revision, namely the minimisation of revision steps.

There is, however, an aspect of further work that deserves some extensive discussion, for it is one of the crucial changes in the form of defeasible logics as studied in the recent past: the distinction between defeasible rules employed to represent plausible interpretation of the reality, and those that are used to introduce prescriptions to the behaviour of agents. Although this distinction has been dealt with in the current literature of logic and AI (see [2] for a large discussion), the applications of these aspects to the representation of voting issues and collective decisions is yet lacking.

In such a distinction, a given rule might state '*Usually women like beer less than men*' is a factual description, though, possibly, just plausible, and therefore significantly different from a rule that states '*Every action has an opposite and equal reaction*', which has more the flavour of a *strict* rule. This should be differentiated from a normative rule such as '*Citizens cannot wear war weapons*'. Law makers aim, where possible, at changing the second type of rules, and therefore a political proposal consists in the introduction of new normative rules. However, the ultimate purpose of political actions is to affect the reality. Consider that a law is introduced that increases the level of recognised

competence for a certain class of workers (for instance, in Australia, yoga instructors must now have 500 hours of certified training, whereas before was only 200). This will impact on the life of many people: some may undergo further training and education to continue within their current field, some may chose not to undertake further study and will consequently loose their job, since less people will have the required level to teach, the job will likely pay more.

It may also be the case that votes are expressed not only based on the rationality (but possibly within the rationality boundary) but also considering the *intention* included in the norm change proposals.

The purpose of such an investigation would be threefold:

- Determine what difference may be devised in terms of revision process, when such a distinction is introduced;
- Establish whether a revision process applies to normative part without interfering with the non-normative part;
- Establish whether the reality could be modified by the introduction of a normative set of rules, and by specifically introducing those rules by collective decisions.

Another important aspect that has not yet been investigated is the distinction between negative and positive choice elements. A negative literal expresses two different desires: the desire for the literal to become false, regardless of the nature of the positive literal corresponding to it, and the desire for the literal *not to become true*. For instance, consider the situation where, in a normative system, there is a law that forces women to wear certain types of clothing. Certain people may be disappointed that such an illiberal norm is not suppressed.

References

1. Andréka, H., Ryan, M., Schobbens, P.Y.: Operators and laws for combining preference relations. *Journal of Logic and Computation* 12(1), 13–53 (2002)
2. Antoniou, G., Billington, D., Governatori, G., Maher, M.J.: Representation results for defeasible logic. *ACM Trans. Comput. Log.* 2(2), 255–287 (2001)
3. Arrow, K.J.: A difficulty in the concept of social welfare. *Journal of Political Economy* 58(4), 328–346 (1950)
4. Chevaleyre, Y., Endriss, U., Lang, J., Maudet, N.: A short introduction to computational social choice. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4362 LNCS, 51–69 (2007)
5. Chevaleyre, Y., Endriss, U., Lang, J., Maudet, N.: Preference handling in combinatorial domains: From ai to social choice. *AI Magazine* 29(4), 37–46 (2008)
6. Conitzer, V., Sandholm, T., Lang, J.: When are elections with few candidates hard to manipulate. *Journal of the ACM* 54(3) (2007)
7. Dong, Y., Zhang, H.: Multiperson decision making with different preference representation structures: A direct consensus framework and its properties. *Knowledge-Based Systems* 58, 45–57 (2014)
8. Endriss, U.: *Judgment aggregation* (2016)
9. Fono, L., Andjiga, N.: Fuzzy strict preference and social choice. *Fuzzy Sets and Systems* 155(3), 372–389 (2005)

10. Geist, C., Endriss, U.: Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research* 40, 143–174 (2011)
11. Governatori, G., Olivieri, F., Cristani, M., Scannapieco, S.: Revision of defeasible preferences. *International Journal of Approximate Reasoning* 104, 205–230 (2019)
12. Governatori, G., Olivieri, F., Scannapieco, S., Cristani, M.: The hardness of revising defeasible preferences. In: Bikakis, A., Fodor, P., Roman, D. (eds.) *Rules on the Web. From Theory to Applications - 8th International Symposium, RuleML. Lecture Notes in Computer Science*, vol. 8620, pp. 168–177. Springer (2014)
13. Governatori, G., Olivieri, F., Scannapieco, S., Rotolo, A., Cristani, M.: The rationale behind the concept of goal. *TPLP* 16(3), 296–324 (2016)
14. Munda, G.: Social multi-criteria evaluation: Methodological foundations and operational consequences. *European Journal of Operational Research* 158(3), 662–677 (2004)
15. Olivieri, F., Cristani, M., Governatori, G.: Compliant business processes with exclusive choices from agent specification. *Lecture Notes in Computer Science*, vol. 9387, pp. 603–612. Springer (2015)
16. Pini, M., Rossi, F., Venable, K., Walsh, T.: Aggregating partially ordered preferences. *Journal of Logic and Computation* 19(3), 475–502 (2009)
17. Rossi, F., Venable, K., Walsh, T.: A short introduction to preferences: Between artificial intelligence and social choice. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 14, 1–102 (2011)
18. Wang, Y.M., Yang, J.B., Xu, D.L.: A preference aggregation method through the estimation of utility intervals. *Computers and Operations Research* 32(8), 2027–2049 (2005)
19. Zuckerman, M., Procaccia, A., Rosenschein, J.: Algorithms for the coalitional manipulation problem. *Artificial Intelligence* 173(2), 392–412 (2009)