# Social Choice Theory as a Foundation for Multiagent Systems

Ulle Endriss

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam

**Abstract.** Social choice theory is the study of mechanisms for collective decision making. While originally concerned with modelling and analysing political decision making in groups of people, its basic principles, arguably, are equally relevant to modelling and analysing the kinds of interaction taking place in a multiagent system. In support of this position, I review examples from three strands of research in social choice theory: fair division, voting, and judgment aggregation.

## 1 Introduction

Multiagent systems are systems composed of several autonomous agents, e.g., computer programs or robots, that interact with each other in a variety of ways, including coordination, cooperation, and competition. There are clear parallels to the ways in which individual people interact with each other as members of our society. It therefore is not surprising that formalisms and methods developed in the social and economic sciences, originally intended for the purposes of modelling and analysing humans and human society, have found rich applications also in the field of autonomous agents and multiagent systems.

A relatively recent example for this phenomenon is *social choice theory*, the study of mechanisms for collective decision making [1]. In this short paper I want to review some of the basic ideas developed in social choice theory and argue for their relevance to multiagent systems.

Social choice theory is usually considered part of economic theory, although besides economists also political scientists, philosophers, and mathematicians have contributed significantly to its development as a discipline. More recently, social choice theory, and specifically *computational social choice* [4, 2], has also become a hot topic in computer science and artificial intelligence, with many new conceptual and technical contributions coming from this community.

To sketch the scope of social choice theory, particularly in view of possible applications in multiagent systems, it is useful to consider it side by side with two neighbouring disciplines, namely *decision theory* and *game theory*. Decision theory provides us with tools to model and analyse the decision-making capabilities of a single intelligent agent. For an agent to make a decision it needs to understand how its actions will impact on the state of the world around it and it needs to realise its own preferences over these possible alternative states of the world, so as to be able to choose the best action. That is, decision theory is the right framework when we want to model a single agent *vis à vis* nature. Once we zoom in on nature and want to specifically model the fact that part of nature consists of other agents taking their own decisions, we enter the realms of game theory. This is the right framework to work in when, still taking the perspective on an individual agent, we want to model that agent's strategic behaviour *vis à vis* other agents. That is, while decision theory considers one agent at a time, game theory is concerned with such an agent's interactions with other agents. Social choice theory, finally, goes one step further in that direction and is concerned with such a group of interacting agents as a whole. In social choice theory we ask questions such as *"what is good for this group?"*, while we ask *"what is good for this agent?"* in decision theory and *"how can this agent do well, given that others try too?"* in game theory.

In the remainder of this paper I will briefly review three technical frameworks for social choice theory that model different aspects of collective decision making: deciding on a fair allocation of resources to agents, making a choice affecting all agents in view of their individual preferences, and aggregating the judgments of different agents regarding a number of logically related statements to come to a consistent view that appropriately reflects the position taken by the group as a whole. In other words, I will review the frameworks of *fair division* (Section 2), *voting* (Section 3), and *judgment aggregation* (Section 4).

## 2 Fair Allocation of Resources

Resource allocation plays a central role in multiagent systems research. For one, the problems people try to address by designing a multiagent system often just *are* problems of resource allocation (e.g., applications in electronic commerce). But also for other types of applications we often have to solve a resource allocation problem (e.g., allocating computing resources to different agents) along the way before our agents are in a position to solve the problem they have been designed to address.

Many resource allocation problems can be modelled as follows. We have a set of *agents* $N = \{1, \ldots, n\}$ and a set of (indivisible) *goods* $G = \{g_1, \ldots, g_m\}$. Each agent $i \in N$ is equipped with a *utility function* $u_i : 2^G \to \mathbb{R}$, mapping sets of goods to real numbers indicating the value the agent assigns to the set in question. An *allocation* is a function $A : N \to 2^G$, mapping each agent to a set of goods, that respects $A(i) \cap A(j) = \emptyset$ for all $i \neq j \in N$, i.e., no item may be assigned to more than one agent.

The question then arises: what is a *good* allocation? The answer most frequently given in the context of multiagent systems is that it should be an allocation that maximises the sum of individual utilities. That is, it should be an allocation $A$ that maximises $\sum_{i \in N} u_i(A(i))$, the *utilitarian social welfare* of $A$. This certainly will be the right objective function to optimise in certain contexts. For example, if the utility experienced by each agent represents monetary revenue generated by that agent and as system designers we collect this revenue from all our agents, then what we are interested in is indeed the sum of individual utilities. But if the agents represent individual clients that make use of a multiagent platform we provide, then we may have rather different objectives. We may, for instance, wish to guarantee that individual agents are treated in a fair manner, so as to improve the user experience of our clients. Social choice theory offers a number of useful definitions for how to make the vague notion of *fairness* precise in the context of resource allocation [14]:

- We could look for an allocation $A$ that maximises the *egalitarian social welfare* $\min\{u_i(A(i)) \mid i \in N\}$, i.e., the utility of the agent that is worst off.
- Alternatively, we could maximise the *Nash social welfare* $\prod_{i \in N} u_i(A(i))$, which may be considered a compromise between the utilitarian and the egalitarian point of view, as it sanctions both increases in total efficiency and reallocations from rich to poor agents.
- Or we could look for an allocation $A$ that is *envy-free*, i.e., where $u_i(A(i)) \geqslant u_i(A(j))$ for all $i, j \in N$, meaning that every agent $i$ values the set $A(i)$ of goods assigned to it no less than the set $A(j)$ assigned to any other agent $j$.

Elsewhere, my coauthors and I have argued for a systematic exploitation of the rich variety of fairness criteria developed in the classical literature on social choice theory as design objectives for multiagent systems [3].

An important class of fairness criteria that is not yet widely used in multiagent systems are *inequality indices*, quantifying the degree of economic inequality in a group of agents (see [6] for an introduction aimed at computer scientists).

## 3 Voting

Voting is a framework for choosing a best alternative from a given set of available alternatives, given the preferences of a group of voters over these alternatives. The classical example is that of political elections, where the alternatives are the candidates standing for election and voters express their preferences on the ballot sheet. But voting rules can also be used in many other contexts, including multiagent systems. For example, the alternatives may be different plans available to a group of agents to execute together, and each agent may have their own preferences over alternative plans, determined by the information and reasoning capabilities available to them.

Social choice theory provides a simple mathematical framework for modelling the process of voting. Its ingredients are a set of *agents* $N = \{1, \ldots, n\}$ (the voters), a set of alternatives $X = \{x_1, \ldots, x_\ell\}$, and one *preference order* $\succ_i$ for each agent $i \in N$. Every such preference order $\succ_i$ is taken to be a strict linear order on $X$ (i.e., a binary relation that is irreflexive, transitive, and complete). We write $\mathcal{L}(X)$ for the set of all such linear orders. A *voting rule* is a function $F : \mathcal{L}(X)^n \to 2^X \setminus \{\emptyset\}$, mapping any given *profile* of preference orders (one for each agent) to a nonempty set of winning alternatives (due to the possibility of ties we cannot be sure to always obtain a single winner).

Voting theory provides many different such rules [16]. To exemplify the range, let us define two of them here:

- Under the *Borda rule*, every alternative obtains $\ell - k$ points whenever a voter ranks that alternative in position $k$ in her preference ordering. Thus, if there are 10 alternatives, then your most preferred alternative receive 9 points from you, your second most preferred alternative receives 8 points, and so forth. The alternative with the most points wins.
- Under the *Copeland rule*, we elect the alternative that wins the largest number of pairwise majority contests against other alternatives (with half a point awarded for a draw).

Observe that the so-called *plurality rule*, under which every voter can nominate one alternative and the alternative with the most nominations wins the election, also fits into this framework. The reason is that we can think of it as the rule under which the alternative ranked on top most often wins. The plurality rule is the rule used in most political elections, but it does in fact have many undesirable properties. For example, in the scenario below, $z$ wins under the plurality rule, yet every alternative other than $z$ is preferred to $z$ by a strict majority:

| |
|---|
| 2 agents: $x \succ y \succ z$ |
| 2 agents: $y \succ x \succ z$ |
| 3 agents: $z \succ x \succ y$ |

Observe that both the Borda rule and the Copeland rule will instead elect the intuitively "right" winner, namely $x$.

When designing a decision making mechanism for a multiagent system, it is important to refer back to the classical literature on voting theory, which has examined many of the questions of relevance here in depth. Having said this, multiagent systems will often give rise to requirements that are somewhat different from those arising in the context of elections amongst humans. For example, due to the bounded rationality of autonomous software agents (and due to the fact that we can model such bonded rationality much more precisely in such cases than for a human agent), we may wish to drop or alter some of the classical assumptions regarding preferences. Specifically, the assumption of completeness, i.e., an agent's ability to rank any two alternatives, will not always be appropriate. This insight has led to work in the artificial intelligence literature on *voting with nonstandard preferences* [15, 8]. In a similar vain, the alternatives that we need to choose between in the context of multiagent systems may not always be just simple "atomic" options, but they may come with some rich internal structure. This has lead to the research direction of *voting in combinatorial domains* in computational social choice [11, 5].

Observe that, in principle, we could use voting rules also in the context of resource allocation. If we think of the set of all possible allocations of resources to agents as the set of alternatives, then each agent's utility function defines its preference relation over these alternatives.

## 4 Judgment Aggregation

Preferences are not the only types of structures we may wish to aggregate in a multiagent system. Other examples include in particular individual judgments regarding the truth or falsehood of certain statements. Such questions have been investigated in the literature on *belief merging* [10] and *judgment aggregation* [13], as well as the closely related *binary aggregation* [9].

Let us briefly review the basic framework of judgment aggregation here. An *agenda* is a finite set of formulas of propositional logic that is of the form $\Phi = \Phi^+ \cup \{\neg\varphi \mid \varphi \in \Phi^+\}$, with $\Phi^+$ only including non-negated formulas. Now consider a set of *agents* $N = \{1, \ldots, n\}$ such that each agent $i \in N$ picks a *judgment set* $J_i \subseteq \Phi$ that is logically consistent and that includes either $\varphi$ or $\neg\varphi$ for every $\varphi \in \Phi^+$. Let $\mathcal{J}(\Phi)$ be the set of all such consistent and complete judgment sets for the agenda $\varphi$. An *aggregator* is a function $F : \mathcal{J}(\Phi)^n \to 2^\Phi$, mapping profiles of such judgment sets (one for each agent) to subsets of the agenda (i.e., to judgment sets that may or may not also be consistent and complete).

An example is the *majority rule*, under which we accept all those formulas that are accepted by a majority of the agents. Unfortunately, this simple rule is often not satisfactory, as it can lead to paradoxical outcomes. For example, suppose that $\Phi^+ = \{p, q, p \wedge q\}$. The table below show a profile with three agents (we assume that an agent accepts $\neg\varphi$ if and only if it does not accept $\varphi$):

| | $p$ | $q$ | $p \wedge q$ |
|---|---|---|---|
| Agent 1: | Yes | Yes | Yes |
| Agent 2: | Yes | No | No |
| Agent 3: | No | Yes | No |
| Majority: | Yes | Yes | No |

Thus, even though each individual agent declares a consistent judgment set (e.g., agent 2 claims $p$ is true and $q$ is false, i.e., $p \wedge q$ is also false), the majority judgment set we obtain is inconsistent: you cannot accept both $p$ and $q$ but also reject $p \wedge q$. There are several proposals for alternative methods of aggregation that avoid this kind of problem, albeit often at the price of significantly increased complexity [12]. This includes, for instance, distance-based methods where we choose as outcome a consistent judgment set that minimises, in some sense, the distance to the input profile. Another approach amounts to choosing from amongst the agents one that may be considered a good representative of the group, to then implement that agent's advice [7].

In the same way as resource allocation problems can, in principle, be studied as voting problems, voting can, in principle, be embedded into judgment aggregation. The basic idea is to work with propositional variables of the form $p_{x \succ y}$, acceptance of which would then signal that an agent prefers $x$ to $y$.

## 5 Conclusion

We have reviewed three frameworks for collective decision making and discussed their relevance to multiagent systems. Judgment aggregation and the closely related binary aggregation are the most general amongst them: they, in principle, allow us to aggregate any kind of information. The other two frameworks, voting and fair division, deal with one specific type of information to be aggregated, namely information on the preferences of the agents. Fair division, again, is more specific than voting, as it imposes specific constraints on the types of preferences held by agents over alternative outcomes. Despite this relative narrowness of scope, or maybe because of it, of the ideas discussed here, those pertaining to resource allocation are clearly the most widely used in multiagent systems research today. But also the others have clear potential for multiagent systems.

## References

1. Arrow, K.J., Sen, A.K., Suzumura, K. (eds.): Handbook of Social Choice and Welfare. North-Holland (2002)
2. Brandt, F., Conitzer, V., Endriss, U.: Computational social choice. In: Weiss, G. (ed.) Multiagent Systems, pp. 213–283. MIT Press (2013)
3. Chevaleyre, Y., Endriss, U., Estivie, S., Maudet, N.: Welfare engineering in practice: On the variety of multiagent resource allocation problems. In: Engineering Societies in the Agents World V. LNAI, vol. 3451, pp. 335–347. Springer-Verlag (2005)
4. Chevaleyre, Y., Endriss, U., Lang, J., Maudet, N.: A short introduction to computational social choice. In: Proc. 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM). LNCS, vol. 4362, pp. 51–69. Springer-Verlag (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. Endriss, U.: Reduction of economic inequality in combinatorial domains. In: Proc. 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2013)
7. Endriss, U., Grandi, U.: Binary aggregation by selection of the most representative voter. In: Proc. 28th AAAI Conference on Artificial Intelligence (2014)
8. Endriss, U., Pini, M.S., Rossi, F., Venable, K.B.: Preference aggregation over restricted ballot languages: Sincerity and strategy-proofness. In: Proc. 21st International Joint Conference on Artificial Intelligence (IJCAI) (2009)
9. Grandi, U., Endriss, U.: Binary aggregation with integrity constraints. In: Proc. 22nd International Joint Conference on Artificial Intelligence (IJCAI) (2011)
10. Konieczny, S., Pino Pérez, R.: Merging information under constraints: A logical framework. Journal of Logic and Computation 12(5), 773–808 (2002)
11. Lang, J.: Logical preference representation and combinatorial vote. Annals of Mathematics and Artificial Intelligence 42(1–3), 37–71 (2004)
12. Lang, J., Slavkovik, M.: How hard is it to compute majority-preserving judgment aggregation rules? In: Proc. 21st European Conference on Artificial Intelligence (ECAI) (2014)
13. List, C., Puppe, C.: Judgment aggregation: A survey. In: Anand, P., Pattanaik, P., Puppe, C. (eds.) Handbook of Rational and Social Choice. Oxford University Press (2009)
14. Moulin, H.: Axioms of Cooperative Decision Making. Cambridge University Press (1988)
15. Pini, M.S., Rossi, F., Venable, K.B., Walsh, T.: Aggregating partially ordered preferences. Journal of Logic and Computation 19(3), 475–502 (2008)
16. Taylor, A.D.: Social Choice and the Mathematics of Manipulation. Cambridge University Press (2005)