Rethinking the Neutrality Axiom in Judgment Aggregation

Extended Abstract

Zoi Terzopoulou
Institute for Logic, Language and Computation
Amsterdam, The Netherlands
z.terzopoulou@uva.nl

Ulle Endriss
Institute for Logic, Language and Computation
Amsterdam, The Netherlands
ulle.endriss@uva.nl

ABSTRACT

When aggregating the judgments of a group of agents, an important consideration concerns the fairness of the aggregation process. This is the fundamental idea behind the neutrality axiom in social choice theory: if two judgments enjoy the same support amongst the agents, either both or neither of them should be part of the collective decision. This is a reasonable requirement in many scenarios, but we argue that for scenarios in which agents are asked to judge very diverse kinds of propositions, the classical neutrality axiom is much too strong. We thus propose a family of weaker neutrality axioms, parametrised by binary relations between the propositions.

KEYWORDS

Judgment Aggregation; Social Choice Theory; Fairness

1 INTRODUCTION

How can we ensure that a group of (human or artificial) agents with diverse views will make collective decisions in a manner that is fair and that guarantees internal coherence? This question is addressed by Judgment Aggregation (JA), a formal framework for collective decision making in logically rich domains that has been studied across various disciplines, from philosophy to AI [1, 6, 7, 11]. In JA we can formally define the properties (also known as *axioms*) that we would like to be satisfied by a rule to aggregate individual judgments into a single collective judgment, and then examine which of the available such rules, if any, suit our purposes [5].

Here, we delve deeply into *neutrality*, a prominent axiom in the literature on JA [3, 4, 6, 7, 9].[1] Neutrality demands that all propositions judged by the members of a group should be treated symmetrically during aggregation—without any kind of differentiation. This clearly is a natural requirement for some applications, encoding a basic notion of fairness. But in most domains in which JA can be applied the propositions to be judged will be very diverse in nature and, therefore, the standard neutrality axiom is much too strong a requirement.[2] Consider the following (factitious) example.

---

[1]In fact, in the early literature on JA, neutrality is often combined with an *independence* axiom and appears only in disguise, as part of an axiom called *systematicity*.
[2]This observation has also been made by Slavkovik [12].

Example 1. The governor of Arizona is putting together a committee to help him decide (*i*) if a wall should be constructed separating Arizona and Mexico, and (*ii*) if the tuition fees for the state colleges should be increased—while it is known that in order for the wall's budget to be covered, money has to be collected (also) through more tuition. Suppose the committee consists of four pure democrats and six extreme conservatives, expressing judgments as follows:

	Wall?	Tuition?
6 members	Yes	Yes
4 members	No	No

Thus, the governor sees 60% of the committee members agree with building the wall, and exactly the same members support the increase on college tuition. Now, regarding the tuition, a simple majority by the committee members will suffice for the governor to make a decision in favour of an increase. However, a wall will have much more serious consequences to Arizona's finances and international relations with Mexico, so the governor would only decide for it if at least a 2/3 majority of the committee endorsed this. △

Apparently, in situations like the one described above, propositions may vary in *meta-value:* Judgments on certain propositions may have much more severe and large-scale effects than judgments on others. Then, we arguably should treat propositions of the same kind symmetrically, while we may still want to apply very different standards across the different kinds. We thus introduce a refinement of standard neutrality in JA that accounts for the diversity between the propositions under consideration.

2 NEUTRAL JUDGMENT AGGREGATION

In the formal framework of JA [6, 10], an *agenda* Φ contains all propositions on which a decision has to be made—modelled by formulas in propositional logic—and an individual *judgment* $J \subseteq \Phi$ is taken to be a (logically consistent) subset of that agenda.[3] Then, a profile $\boldsymbol{J} = (J_1, \ldots, J_n)$ represents the judgments of all agents $1, \ldots, n$. We denote by $N_\varphi^{\boldsymbol{J}}$ the set of agents who agree with proposition $\varphi \in \Phi$ in the profile $\boldsymbol{J}$ (i.e., the set of agents i for which $\varphi \in J_i$). A (resolute) aggregation rule is a function F that maps every possible profile $\boldsymbol{J}$ to the group's judgment $F(\boldsymbol{J}) \subseteq \Phi$. According to the axiom of neutrality, for *any* two propositions $\varphi, \psi \in \Phi$ and any profile $\boldsymbol{J}$, it must be the case that:

$$\left[N_\varphi^{\boldsymbol{J}} = N_\psi^{\boldsymbol{J}}\right] \text{ implies that } \left[\varphi \in F(\boldsymbol{J}) \Rightarrow \psi \in F(\boldsymbol{J})\right]$$

Clearly, the standard neutrality property proves too stringent when situations like the wall/tuition problem of Example 1 occur.

---

[3]For instance, in Example 1 the committee decides on propositions $w, t \in \Phi$ (capturing decisions on "wall" and "tuition", respectively), and we could also consider a proposition $w \rightarrow t$ (capturing the constraint that building a wall requires collecting tuition fees).

3 RELATIONAL NEUTRALITY

Consider a binary relation R between propositions. The corresponding R-neutrality axiom will impose equal treatment *only* for those propositions that are linked to each other by R.[4] Formally, for *any* two propositions $\varphi, \psi \in \Phi$ and any profile J, it must hold that:

$$\left[N_\varphi^J = N_\psi^J \text{ and } \varphi R \psi \right] \text{ implies that } \left[\varphi \in F(J) \Rightarrow \psi \in F(J) \right]$$

Hence, we directly weaken standard neutrality. Instead of talking about *all* propositions $\varphi, \psi \in \Phi$, we now only refer to propositions with $\varphi R \psi$. Note that for R being the total relation between propositions, relational neutrality is equivalent to standard neutrality.

Example 2. Having in mind an agenda like the one of the governor of Arizona we saw in Example 1, let us define an equivalence relation R that partitions the agenda into two classes of propositions of the same kind, that is, propositions with severe or light consequences, respectively. Then, an aggregation rule that accepts a proposition in a given class if its support exceeds a fixed (different for each class) quota is R-neutral but not classically neutral. △

More generally, given an agenda Φ, we can define a *type* $T \subseteq \Phi$ as a set of propositions "of the same kind". The *type* T then induces the (symmetric) relation $R_T = \{(\varphi, \psi) \in \Phi \times \Phi \mid \varphi, \psi \in T\}$, and the property of R_T-neutrality simply says that all propositions of *type* T will be treated equally by an aggregation rule.

The notion of a type can be useful for several JA scenarios of practical interest. For instance, in an agenda differences may arise between propositions that are either *subjective* or *objective* in character, meaning that they are or are not associated with a ground truth [12]. Propositions may for example express *preferences* rather than *opinions*. Preferences are subjective. On the other hand, when a committee of doctors need to evaluate whether a patient has cancer or not, their collective judgment will obviously be either right or wrong, depending on reality. Therefore, we may need to define the type of all objective propositions, distinguish from those that are subjective, and apply relational neutrality specifically on them.

Moreover, besides their conceptual interpretation, within the formal model of JA propositions are represented by specific logical formulas. We can thus compare them syntactically and semantically; propositions of the same type may be considered those that are logically equivalent or those that have the same syntax. For instance, for different propositions φ we can define the equivalence relation $R_\leftrightarrow^\varphi$ induced by the type $T_\leftrightarrow^\varphi = \{\psi \in \Phi \mid \varphi \leftrightarrow \psi\}$.[5]

Example 3. Consider an aggregation problem that involves the premises p, q, r, z and the conclusion $c = p \wedge q \wedge r \wedge z$. The premise-based rule (according to which every premise is collectively accepted if a majority agrees with it, and the conclusion is accepted if all premises are) is R_p-neutral, where R_p relates all premises to each other using the type $T_p = \{p, q, r, z\}$, but it is not neutral in general. Here is a profile violating neutrality:

	p	q	r	z		$p \wedge q \wedge r \wedge z$
Agent 1:	Yes	No	Yes	Yes	$\Rightarrow$	No
Agent 2:	Yes	Yes	No	Yes	$\Rightarrow$	No
Agent 3:	Yes	Yes	Yes	No	$\Rightarrow$	No

Propositions p and $\neg c$ are supported by all the agents. However, the premise-based rule will include p in the group's judgment (because a majority accepts it) but will not include $\neg c$ (since each premise is accepted by a majority, c will be included instead). △

Example 4. Given a proposition φ it may happen that neither φ nor $\neg \varphi$ are logically implied by any consistent subset of propositions that does not contain them. The Kemeny rule [6, 8] then simply follows the majority's opinion on φ. Defining the type T_{ind} to contain all such propositions and considering the corresponding relation R_{ind}, the Kemeny rule becomes R_{ind}-neutral, while it is not neutral in general in judgment aggregation [2] (even though it is neutral in the context of preference aggregation). △

In addition to symmetric relations between propositions discussed so far, we of course have the freedom to employ non-symmetric relations too. The *semantics-based* relations hinge on comparisons between the (number of) models of the propositions involved. We denote by $\mathcal{M}_\varphi$ the set of models of φ.[6] For $\star \in \{\geq, >, \leq, <\}$ and $\bullet \in \{\supseteq, \supset, \subseteq, \subset\}$, we define:

$$R_\star = \{(\varphi, \psi) \in \Phi \times \Phi \mid |\mathcal{M}_\varphi| \star |\mathcal{M}_\psi|\}$$
$$R_\bullet = \{(\varphi, \psi) \in \Phi \times \Phi \mid \mathcal{M}_\varphi \bullet \mathcal{M}_\psi\}$$

To illustrate, in certain JA scenarios distinctions may be made with respect to the *amount of evidence* different propositions require for their acceptance. For instance, a proposition can be very easy to agree with because it has more models in logic, while another one may be easier to reject. In some cases it makes sense that easily acceptable propositions require at least as much support as easily rejectable ones to be collectively accepted; but the converse can be reasonable too.

Example 5. Think of a group of movie experts judging the quality of films based on three criteria. Proposition (p_j) "film j passes at least one criterion" asks for less evidence to be satisfied than proposition (q_j) " film j passes all criteria". Suppose now that a subset S of the experts agree with propositions p_a and q_b for two films a and b, and based on the support received by S, the whole group announces that film a passes at least one criterion (p_a). But q_b gathered exactly the same support, and, most essentially, the evidence needed for that support was harder to obtain. So, one could argue, the group should also announce that film b passes all criteria, as a matter of fairness for the two films. On the other hand, it could also be argued that the group to make a logically stronger announcement like q_b should be harder (formally requiring more support) than a weaker announcement such as p_a. △

In conclusion, the clearly debatable property of neutrality has not yet received the direct attention it deserves in the literature. In this short note, we suggest how to close this gap by proposing a family of weaker neutrality axioms based on how the propositions under consideration relate to each other in various application domains.

---

[4]The idea of taking into account the relations between propositions when defining a neutrality axiom was already briefly mentioned in the seminal paper by List and Pettit [10] in their Footnote 4. However, they only discuss the option of limiting neutrality to propositions that belong to the same equivalence class, while our definition accounts for a much broader spectrum of relations that may occur.

[5]Note that $R_\leftrightarrow^\varphi$-neutrality is satisfied by all collectively rational aggregation rules.

[6]To be precise, for $\mathcal{M}_\varphi$ we only take into account valuations on variables that appear in propositions of the specific finite agenda of our aggregation problem.

REFERENCES

[1] Dorothea Baumeister, Gábor Erdélyi, and Jörg Rothe. 2016. Judgment Aggregation. In *Economics and Computation*, Jörg Rothe (Ed.). Springer.

[2] Marco Costantini, Carla Groenland, and Ulle Endriss. 2016. Judgment Aggregation under Issue Dependencies. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. 468–474.

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

[4] Franz Dietrich and Christian List. 2008. Judgment Aggregation without Full Rationality. *Social Choice and Welfare* 31, 1 (2008), 15–39.

[5] Franz Dietrich and Christian List. 2010. The Impossibility of Unbiased Judgment Aggregation. *Theory and Decision* 68, 3 (2010), 281–299.

[6] Ulle Endriss. 2016. Judgment Aggregation. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press.

[7] Davide Grossi and Gabriella Pigozzi. 2014. *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning, Vol. 8. Morgan & Claypool Publishers. 1–151 pages.

[8] John Kemeny. 1959. Mathematics Without Numbers. *Daedalus* 88 (1959), 577–591.

[9] Christian List. 2012. The Theory of Judgment Aggregation: An Introductory Review. *Synthese* 187, 1 (2012), 179–207.

[10] Christian List and Philip Pettit. 2002. Aggregating Sets of Judgments: An Impossibility Result. *Economics and Philosophy* 18, 1 (2002), 89–110.

[11] Christian List and Clemens Puppe. 2009. Judgment Aggregation: A Survey. In *Handbook of Rational and Social Choice*, Paul Anand, Prasanta Pattanaik, and Clemens Puppe (Eds.). Oxford University Press.

[12] Marija Slavkovik. 2014. Not all Judgment Aggregation Should be Neutral. In *Proceedings of the 2nd European Conference on Social Intelligence (ECSI)*. 198–211.