Natural deduction for modal logic of judgment aggregation

Tin Perkov

Polytechnic of Zagreb

LAP 2014, Dubrovnik

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)?

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgment aggregation

- Judgment aggregation problem: how to make a fair decision based on individual judgments?
- Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.
- Judgments can be formalized as consistent sets of logical formulas.

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),
- *agenda* $\mathcal{A}$, a set of formulas of a fixed underlying logic (containing $\neg$ and $\rightarrow$ with standard semantics), s.t. if A is not of the form $\neg B$, then $A \in \mathcal{A}$ iff $\neg A \in \mathcal{A}$,

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),
- *agenda* $\mathcal{A}$, a set of formulas of a fixed underlying logic (containing $\neg$ and $\to$ with standard semantics), s.t. if A is not of the form $\neg B$, then $A \in \mathcal{A}$ iff $\neg A \in \mathcal{A}$,
- judgment set is a consistent $R_i \subseteq \mathcal{A}$ s.t. $A \in R_i$ or $\neg A \in R_i$ for all $A \in \mathcal{A}$ not of the form $\neg B$,

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),
- *agenda* $\mathcal{A}$, a set of formulas of a fixed underlying logic (containing $\neg$ and $\to$ with standard semantics), s.t. if A is not of the form $\neg B$, then $A \in \mathcal{A}$ iff $\neg A \in \mathcal{A}$,
- judgment set is a consistent $R_i \subseteq \mathcal{A}$ s.t. $A \in R_i$ or $\neg A \in R_i$ for all $A \in \mathcal{A}$ not of the form $\neg B$,
- *profile* is an n-tuple $R = (R_1, \ldots, R_n)$ of judgment sets,

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),
- *agenda* $\mathcal{A}$, a set of formulas of a fixed underlying logic (containing $\neg$ and $\rightarrow$ with standard semantics), s.t. if A is not of the form $\neg B$, then $A \in \mathcal{A}$ iff $\neg A \in \mathcal{A}$,
- judgment set is a consistent $R_i \subseteq \mathcal{A}$ s.t. $A \in R_i$ or $\neg A \in R_i$ for all $A \in \mathcal{A}$ not of the form $\neg B$,
- *profile* is an n-tuple $R = (R_1, \ldots, R_n)$ of judgment sets,
- judgment aggregation rule (JAR) is a function F which maps each profile R to a judgment set $F(R)$.

Judgment aggregation

Judgment aggregation problem: how to make a fair decision based on individual judgments?

Preference aggregation or social choice: how to determine society's preference (e.g. results of elections) from individual preferences (votes)? This is a special case of judgment aggregation.

Judgments can be formalized as consistent sets of logical formulas.

A framework for judgment aggregation:

- a set N of n individuals (agents, judges, voters),
- *agenda* $\mathcal{A}$, a set of formulas of a fixed underlying logic (containing $\neg$ and $\rightarrow$ with standard semantics), s.t. if A is not of the form $\neg B$, then $A \in \mathcal{A}$ iff $\neg A \in \mathcal{A}$,
- judgment set is a consistent $R_i \subseteq \mathcal{A}$ s.t. $A \in R_i$ or $\neg A \in R_i$ for all $A \in \mathcal{A}$ not of the form $\neg B$,
- *profile* is an n-tuple $R = (R_1, \ldots, R_n)$ of judgment sets,
- judgment aggregation rule (JAR) is a function F which maps each profile R to a judgment set $F(R)$.

Judgment set R_i represents judgments of agent i, while $F(R)$ represents resulting collective judgment.

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization.

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,
- a special propositional variable σ, representing the aggregated judgment,

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,
- a special propositional variable σ, representing the aggregated judgment,
- formulas are built inductively using Boolean connectives and modalities $\Box$ and $\blacksquare$.

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,
- a special propositional variable σ, representing the aggregated judgment,
- formulas are built inductively using Boolean connectives and modalities $\Box$ and $\blacksquare$.

Formulas are interpreted on the fixed Kripke frame:

- worlds are all pairs (R, A), where R is a profile and $A \in \mathcal{A}$,

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,
- a special propositional variable σ, representing the aggregated judgment,
- formulas are built inductively using Boolean connectives and modalities $\Box$ and $\blacksquare$.

Formulas are interpreted on the fixed Kripke frame:

- worlds are all pairs (R, A), where R is a profile and $A \in \mathcal{A}$,
- accessibility relations corresponding to modalities are:
  - $\Box$: "have the same agenda item,"

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Modal logic of judgment aggregation

Ågotnes et al.[1] defined a sound and complete modal logic JAL for reasoning about judgment aggregation, using a Hilbert-style axiomatization. The language of JAL is parametrized by N and $\mathcal{A}$:

- a propositional variable p_i for each $i \in N$,
- a propositional variable q_A for each $A \in \mathcal{A}$,
- a special propositional variable σ, representing the aggregated judgment,
- formulas are built inductively using Boolean connectives and modalities $\square$ and $\blacksquare$.

Formulas are interpreted on the fixed Kripke frame:

- worlds are all pairs (R, A), where R is a profile and $A \in \mathcal{A}$,
- accessibility relations corresponding to modalities are:
  - $\square$: "have the same agenda item,"
  - $\blacksquare$: "have the same profile."

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems*, 22(1):4–30, 2011

Semantics

A *model* is determined by a judgment aggregation rule F.

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',
- $F, R, A \Vdash \blacksquare\varphi$ iff $F, R, A' \Vdash \varphi$ for all agenda items A'.

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',
- $F, R, A \Vdash \blacksquare\varphi$ iff $F, R, A' \Vdash \varphi$ for all agenda items A'.

The validity (denoted $\Vdash \varphi$) and the global truth in a model ($F \Vdash \varphi$) is defined as usual.

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',
- $F, R, A \Vdash \blacksquare\varphi$ iff $F, R, A' \Vdash \varphi$ for all agenda items A'.

The validity (denoted $\Vdash \varphi$) and the global truth in a model ($F \Vdash \varphi$) is defined as usual.
Clearly $\Box\blacksquare$ is in fact the universal modality. We can use an abbreviation $U\varphi := \Box\blacksquare\varphi$.

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',
- $F, R, A \Vdash \blacksquare\varphi$ iff $F, R, A' \Vdash \varphi$ for all agenda items A'.

The validity (denoted $\Vdash \varphi$) and the global truth in a model ($F \Vdash \varphi$) is defined as usual.
Clearly $\Box\blacksquare$ is in fact the universal modality. We can use an abbreviation $U\varphi := \Box\blacksquare\varphi$. We also use diamonds $\Diamond$, $\blacklozenge$, and E, respectively.

Semantics

A *model* is determined by a judgment aggregation rule F. The truth of a formula is defined inductively:

- $F, R, A \Vdash p_i$ iff $A \in R_i$ ("i judges that A holds"),
- $F, R, A \Vdash q_B$ iff $A = B$,
- $F, R, A \Vdash \sigma$ iff $A \in F(R)$ ("group judges that A holds"),
- $F, R, A \Vdash \neg\varphi$ iff $F, R, A \nVdash \varphi$,
- $F, R, A \Vdash \varphi \vee \psi$ iff $F, R, A \Vdash \varphi$ or $F, R, A \Vdash \psi$,
- $F, R, A \Vdash \Box\varphi$ iff $F, R', A \Vdash \varphi$ for all profiles R',
- $F, R, A \Vdash \blacksquare\varphi$ iff $F, R, A' \Vdash \varphi$ for all agenda items A'.

The validity (denoted $\Vdash \varphi$) and the global truth in a model ($F \Vdash \varphi$) is defined as usual.
Clearly $\Box\blacksquare$ is in fact the universal modality. We can use an abbreviation $U\varphi := \Box\blacksquare\varphi$. We also use diamonds $\Diamond$, $\blacklozenge$, and E, respectively.
For $C \subseteq N$, we denote $p_C := \bigwedge_{i \in C} p_i \wedge \bigwedge_{i \in N \setminus C} \neg p_i$ ("exactly voters from C judge that A holds).

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$."

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF).

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF). Consider some properties of SWF's which are expressive in JAL:

- We say that a social welfare function F satisfies the *Pareto condition* if, whenever all voters prefer x to y, then so does society.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF). Consider some properties of SWF's which are expressive in JAL:

- We say that a social welfare function F satisfies the *Pareto condition* if, whenever all voters prefer x to y, then so does society. Clearly, this holds iff $F \Vdash U(p_1 \wedge \cdots \wedge p_n \rightarrow \sigma)$.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF). Consider some properties of SWF's which are expressive in JAL:

- We say that a social welfare function F satisfies the *Pareto condition* if, whenever all voters prefer x to y, then so does society. Clearly, this holds iff $F \Vdash U(p_1 \wedge \cdots \wedge p_n \rightarrow \sigma)$.
- We call F a *dictatorship* if there is a voter whose preferences always agree with society's, i.e. $F \Vdash \bigvee_{i \in N} U(p_i \rightarrow \sigma)$.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF). Consider some properties of SWF's which are expressive in JAL:

- We say that a social welfare function F satisfies the *Pareto condition* if, whenever all voters prefer x to y, then so does society. Clearly, this holds iff $F \Vdash U(p_1 \wedge \cdots \wedge p_n \rightarrow \sigma)$.
- We call F a *dictatorship* if there is a voter whose preferences always agree with society's, i.e. $F \Vdash \bigvee_{i \in N} U(p_i \rightarrow \sigma)$.
- A SWF F is *independent of irrelevant alternatives* (IIA) if society's preference between two alternatives does not depend on any individual's ranking of any other alternative.

Preference aggregation

In the case of preference aggregation, agenda is defined w.r.t. a fixed set M of m alternatives. Agenda items are "$x \in M$ is preferred to $y \in M$." Underlying logic is first-order theory of strict linear orderings. Agenda items are expressed as formulas of the form $x < y$ or $\neg(x < y)$. A judgment set determines a strict linear ordering on M. A JAR is called a social welfare function (SWF). Consider some properties of SWF's which are expressive in JAL:

- We say that a social welfare function F satisfies the *Pareto condition* if, whenever all voters prefer x to y, then so does society. Clearly, this holds iff $F \Vdash U(p_1 \wedge \cdots \wedge p_n \rightarrow \sigma)$.
- We call F a *dictatorship* if there is a voter whose preferences always agree with society's, i.e. $F \Vdash \bigvee_{i \in N} U(p_i \rightarrow \sigma)$.
- A SWF F is *independent of irrelevant alternatives* (IIA) if society's preference between two alternatives does not depend on any individual's ranking of any other alternative. This is equivalent to $F \Vdash U \bigwedge_{C \subseteq N}(p_C \wedge \sigma \rightarrow \Box(p_C \rightarrow \sigma))$.

Arrow's Theorem

Denote the formulas from previous examples as follows:

- $Pareto := U(p_1 \wedge \cdots \wedge p_n \to \sigma)$,
- $IIA := U \bigwedge_{C \subseteq N} (p_C \wedge \sigma \to \Box(p_C \to \sigma))$,
- $Dictatorial := F \Vdash \bigvee_{i \in N} U(p_i \to \sigma)$.

We can now express (instances of) Arrow's impossibility theorem (if there are more then two alternatives, there is no non-dictatorial SWF that satisfies the Pareto condition and IIA): if $|M| \geqslant 3$, then $\Vdash \neg(Pareto \wedge IIA \wedge \neg Dictatorial)$. Ågotnes et al. make some steps towards a formal Hilbert-style proof. I propose an alternative approach – a natural deduction system – to formalize a proof of Arrow's Theorem adapted from Sen[2], as presented by Endriss[3].

[2] A.K. Sen. Social choice theory.
In K.J. Arrow and M.D. Intriligator, editors, *Handbook of Mathematical Economics, Volume 3*. North-Holland, 1986

[3] U. Endriss. Logic and social choice theory.
In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*. College Publications, 2011

Natural deduction rules

Let $Prof = \{R_1, R_2, \dots\}$ and $Var = \{X_1, X_2, \dots\}$ be countable sets of symbols. A *proof* is a sequence of clauses of the form $R, X : \varphi$, where $R \in Prof$, $X \in Var \cup \mathcal{A}$, and φ is a formula of the language of JAL, built using the following rules:

$$\frac{\begin{array}{c} R, X : \varphi \\ R, X : \psi \end{array}}{R, X : \varphi \wedge \psi} \ (\wedge\mathrm{I})$$

$$\frac{R, X : \varphi \wedge \psi}{R, X : \varphi} \ (\wedge\mathrm{E}) \qquad \frac{R, X : \varphi \wedge \psi}{R, X : \psi} \ (\wedge\mathrm{E})$$

$$\frac{R, X : \varphi}{R, X : \varphi \vee \psi} \ (\vee\mathrm{I}) \qquad \frac{R, X : \psi}{R, X : \varphi \vee \psi} \ (\vee\mathrm{I})$$

$$\frac{R, X : \neg\neg\varphi}{R, X : \varphi} \ (\mathrm{DN})$$

$$\frac{\begin{array}{c} R, X : \varphi \to \psi \\ R, X : \varphi \end{array}}{R, X : \psi} \ (\to\mathrm{E}) \qquad \frac{\begin{array}{c} R, X : \varphi \\ R, X : \neg\varphi \end{array}}{R, X : \bot} \ (\neg\mathrm{E})$$

Natural deduction rules

$$\frac{\boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R, X : \psi \end{array}}}{R, X : \varphi \to \psi} \;(\to\mathrm{I})$$

$$\frac{R, X : \varphi \vee \varphi' \quad \boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R, X : \psi \end{array}} \quad \boxed{\begin{array}{c} R, X : \varphi' \\ \vdots \\ R, X : \psi \end{array}}}{R, X : \psi} \;(\vee\mathrm{E})$$

$$\frac{\boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R', X' : \bot \end{array}}}{R, X : \neg\varphi} \;(\neg\mathrm{I})$$

Natural deduction rules

$$\dfrac{\boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R, X : \psi \end{array}}}{R, X : \varphi \rightarrow \psi}\ (\rightarrow \mathrm{I})$$

$$\dfrac{R, X : \varphi \vee \varphi' \quad \boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R, X : \psi \end{array}} \quad \boxed{\begin{array}{c} R, X : \varphi' \\ \vdots \\ R, X : \psi \end{array}}}{R, X : \psi}\ (\vee \mathrm{E})$$

$$\dfrac{\boxed{\begin{array}{c} R, X : \varphi \\ \vdots \\ R', X' : \bot \end{array}}}{R, X : \neg \varphi}\ (\neg \mathrm{I})$$

$$\frac{R, X : \Box \varphi}{R', X : \varphi}\ (\Box \mathrm{E}) \qquad \frac{R, X : \blacksquare \varphi}{R, X' : \varphi}\ (\blacksquare \mathrm{E}) \qquad \frac{R, X : U \varphi}{R', X' : \varphi}\ (U\mathrm{E})$$

$$\frac{R, X : \varphi}{R', X : \Diamond \varphi}\ (\Diamond \mathrm{I}) \qquad \frac{R, X : \varphi}{R, X' : \blacklozenge \varphi}\ (\blacklozenge \mathrm{I}) \qquad \frac{R, X : \varphi}{R', X' : E \varphi}\ (E\mathrm{I})$$

where R' and X' are any (including R and X).

Natural deduction rules

$$\begin{array}{l} R, X : \Diamond\varphi \\ \boxed{\begin{array}{l} R', X : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\ R, X : \psi \quad (\Diamond\text{E}) \end{array}$$

$$\begin{array}{l} R, X : \blacklozenge\varphi \\ \boxed{\begin{array}{l} R, X' : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\ R, X : \psi \quad (\blacklozenge\text{E}) \end{array}$$

$$\begin{array}{l} R, X : E\varphi \\ \boxed{\begin{array}{l} R', X' : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\ R, X : \psi \quad (E\text{E}) \end{array}$$

Natural deduction rules

$$
\begin{array}{l}
R, X : \Diamond\varphi \\
\boxed{\begin{array}{l} R', X : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\
R, X : \psi \quad (\Diamond\mathrm{E})
\end{array}
\qquad
\begin{array}{l}
R, X : \blacklozenge\varphi \\
\boxed{\begin{array}{l} R, X' : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\
R, X : \psi \quad (\blacklozenge\mathrm{E})
\end{array}
\qquad
\begin{array}{l}
R, X : E\varphi \\
\boxed{\begin{array}{l} R', X' : \varphi \\ \quad\vdots \\ R, X : \psi \end{array}} \\
R, X : \psi \quad (E\mathrm{E})
\end{array}
$$

$$
\begin{array}{l}
\boxed{\begin{array}{l} \quad\vdots \\ R', X : \varphi \end{array}} \\
R, X : \Box\varphi \quad (\Box\mathrm{I})
\end{array}
\qquad
\begin{array}{l}
\boxed{\begin{array}{l} \quad\vdots \\ R, X' : \varphi \end{array}} \\
R, X : \blacksquare\varphi \quad (\blacksquare\mathrm{I})
\end{array}
\qquad
\begin{array}{l}
\boxed{\begin{array}{l} \quad\vdots \\ R', X' : \varphi \end{array}} \\
R, X : U\varphi \quad (U\mathrm{I})
\end{array}
$$

where R' and $X' \in \mathit{Var}$ are new, i.e. did not appear in the proof before.

Natural deduction rules

The following rules reflect the semantics of propositional variables, and consistency and completeness of judgment sets.

$$\frac{}{R, A : q_A}\ (Q1) \qquad \frac{R, X : q_A}{R, X : \neg q_B}\ (Q2) \qquad \frac{}{R, X : \bigvee_{A \in \mathcal{A}} q_A}\ (Q3)$$

$$\frac{R, X : q_A}{R', X : q_A}\ (Q4) \qquad \frac{\begin{array}{l} R, X : q_A \\ R, X' : q_A \\ R, X : \varphi \end{array}}{R, X' : \varphi}\ (Q5)$$

where $A, B \in \mathcal{A}$, $B \neq A$, $R, R' \in Prof$, $X, X' \in Var \cup \mathcal{A}$.

Natural deduction rules

The following rules reflect the semantics of propositional variables, and consistency and completeness of judgment sets.

$$\frac{}{R, A : q_A}\ (Q1) \qquad \frac{R, X : q_A}{R, X : \neg q_B}\ (Q2) \qquad \frac{}{R, X : \bigvee_{A \in \mathcal{A}} q_A}\ (Q3)$$

$$\frac{R, X : q_A}{R', X : q_A}\ (Q4) \qquad \frac{\begin{array}{l} R, X : q_A \\ R, X' : q_A \\ R, X : \varphi \end{array}}{R, X' : \varphi}\ (Q5)$$

where $A, B \in \mathcal{A}$, $B \neq A$, $R, R' \in \mathit{Prof}$, $X, X' \in \mathit{Var} \cup \mathcal{A}$.

$$\frac{\begin{array}{l} R, A_1 : p \\ \vdots \\ R, A_k : p \end{array}}{R, B : p}\ (\mathit{Cons}) \qquad \frac{R, X : \neg p}{R, \tilde{X} : p}\ (\mathit{Compl})$$

where $A_1, \ldots, A_k \vdash B$ in the underlying logic, p is any p_i or σ, and $\tilde{X}$ is $\neg X$ if X is not of the form $\neg Y$, otherwise it is Y.

Universal domain rules

An individual can judge about agenda items in any possible way, so a JAR must provide a group decision for any possible profile (*universal domain assumption*).

Universal domain rules

An individual can judge about agenda items in any possible way, so a JAR must provide a group decision for any possible profile (*universal domain assumption*). This allows a type of argument in an informal proof, which begins like this: let R be a profile such that individuals from $C_1 \subseteq N$ judge A_1, individuals from C_2 judge A_2, and so on.

Universal domain rules

An individual can judge about agenda items in any possible way, so a JAR must provide a group decision for any possible profile (*universal domain assumption*). This allows a type of argument in an informal proof, which begins like this: let R be a profile such that individuals from $C_1 \subseteq N$ judge A_1, individuals from C_2 judge A_2, and so on. To address this, we add the following rules:

$$\dfrac{\boxed{\begin{array}{l} R', A_1 : p_{C_1} \\ \quad\vdots \\ R', A_k : p_{C_k} \\ \quad\vdots \\ R, X : \varphi \end{array}}}{R, X : \varphi}\ (UD1) \qquad \dfrac{\begin{array}{c} R_1, X' : p_i \\ R_2, X' : \neg p_j \\ \boxed{\begin{array}{l} R', X' : p_C \\ \quad\vdots \\ R, X : \varphi \end{array}} \end{array}}{R, X : \varphi}\ (UD2)$$

where R' is new, $X' \in Var$, $C_1, \ldots, C_k, C \subseteq N$ and $A_1, \ldots, A_k \in \mathcal{A}$ s.t. for all $i \in N$, $\{A_j : i \in C_j\} \cup \{\neg A_j : i \notin C_j\}$ is consistent in the underlying logic.

Soundness and completeness

A proof can end at any point, provided all boxes are completed. A formula φ is a *theorem* (we write $\vdash \varphi$) if there is a proof (with all boxes completed) which ends with a clause $R, X : \varphi$.

Soundness and completeness

A proof can end at any point, provided all boxes are completed. A formula φ is a *theorem* (we write $\vdash \varphi$) if there is a proof (with all boxes completed) which ends with a clause $R, X : \varphi$.

Theorem

Let φ be any formula of the language of JAL. Then $\vdash \varphi$ iff $\Vdash \varphi$.

Soundness and completeness

A proof can end at any point, provided all boxes are completed. A formula φ is a *theorem* (we write $\vdash \varphi$) if there is a proof (with all boxes completed) which ends with a clause $R, X : \varphi$.

Theorem

Let φ be any formula of the language of JAL. Then $\vdash \varphi$ iff $\Vdash \varphi$.

Proof.

($\Rightarrow$) The claim follows by induction from the apparent soundness of the rules.

Soundness and completeness

A proof can end at any point, provided all boxes are completed. A formula φ is a *theorem* (we write $\vdash \varphi$) if there is a proof (with all boxes completed) which ends with a clause $R, X : \varphi$.

Theorem

Let φ be any formula of the language of JAL. Then $\vdash \varphi$ iff $\Vdash \varphi$.

Proof.

($\Rightarrow$) The claim follows by induction from the apparent soundness of the rules.

($\Leftarrow$) We prove the axioms and simulate the inference rules from Ågotnes et al. in the natural deduction system.

□

Soundness and completeness

A proof can end at any point, provided all boxes are completed. A formula φ is a *theorem* (we write $\vdash \varphi$) if there is a proof (with all boxes completed) which ends with a clause $R, X : \varphi$.

Theorem

Let φ be any formula of the language of JAL. Then $\vdash \varphi$ iff $\Vdash \varphi$.

Proof.

($\Rightarrow$) The claim follows by induction from the apparent soundness of the rules.

($\Leftarrow$) We prove the axioms and simulate the inference rules from Ågotnes et al. in the natural deduction system.

□

Further work: questions regarding complexity, implementation etc.

An admissible rule for preference aggregation

Recall that agenda items in the case of preference aggregation are of the form $x < y$ or $\neg(x < y)$, so we can consider agenda items to be pairs of alternatives.

An admissible rule for preference aggregation

Recall that agenda items in the case of preference aggregation are of the form $x < y$ or $\neg(x < y)$, so we can consider agenda items to be pairs of alternatives. So, in proofs we can write $R, (x, y) : \varphi$ instead of $R, X : \varphi$.

An admissible rule for preference aggregation

Recall that agenda items in the case of preference aggregation are of the form $x < y$ or $\neg(x < y)$, so we can consider agenda items to be pairs of alternatives. So, in proofs we can write $R, (x, y) : \varphi$ instead of $R, X : \varphi$. It is easy to see that the following variant of universal domain rule is admissible for preference aggregation:

$$\frac{\boxed{\begin{array}{l} R', (x_1, y_1) : p_{C_1} \\ \quad\vdots \\ R', (x_k, y_k) : p_{C_k} \\ \quad\vdots \\ R, (x, y) : \varphi \end{array}}}{R, (x, y) : \varphi} \quad (UD)$$

where for each $i \in N$, $\{x_j < y_j : i \in C_j\} \cup \{\neg(x_j < y_j) : i \notin C_j\}$ is consistent for all possible choices of $x_1, y_1, \ldots, x_k, y_k \in M$.

An example of derivation

If $C \subseteq D$, then $\vdash (\mathit{Pareto} \land \mathit{IIA}) \to (\Box(p_C \to \sigma) \to U(p_D \to \sigma))$.

An example of derivation

If $C \subseteq D$, then $\vdash (Pareto \wedge IIA) \rightarrow (\Box(p_C \rightarrow \sigma) \rightarrow U(p_D \rightarrow \sigma))$.
This is an important part of a proof of Arrow's Theorem.

An example of derivation

If $C \subseteq D$, then $\vdash (Pareto \wedge IIA) \rightarrow (\Box(p_C \rightarrow \sigma) \rightarrow U(p_D \rightarrow \sigma))$.
This is an important part of a proof of Arrow's Theorem.
A natural deduction proof should end with
$R, (x, y) : (Pareto \wedge IIA) \rightarrow (\Box(p_C \rightarrow \sigma) \rightarrow U(p_D \rightarrow \sigma))$.

An example of derivation

$$
\boxed{\begin{array}{l} R,(x,y) : \mathit{Pareto} \land \mathit{IIA} \\ \vdots \\ R,(x,y) : \Box(p_C \to \sigma) \to U(p_D \to \sigma) \end{array}}
$$

$$
R,(x,y) : (\mathit{Pareto} \land \mathit{IIA}) \to (\Box(p_C \to \sigma) \to U(p_D \to \sigma)) \quad (\to\mathrm{I})
$$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \rightarrow \sigma)$

$\vdots$

$R, (x, y) : U(p_D \rightarrow \sigma)$

$R, (x, y) : \Box(p_C \rightarrow \sigma) \rightarrow U(p_D \rightarrow \sigma) \quad (\rightarrow \text{I})$

$R, (x, y) : (Pareto \wedge IIA) \rightarrow (\Box(p_C \rightarrow \sigma) \rightarrow U(p_D \rightarrow \sigma)) \quad (\rightarrow \text{I})$

An example of derivation

$$R, (x, y) : Pareto \wedge IIA$$

$$R, (x, y) : \Box(p_C \to \sigma)$$

$$\vdots$$

$$R', (x', y') : p_D \to \sigma$$

$$R, (x, y) : U(p_D \to \sigma) \quad (U\mathrm{I})$$

$$R, (x, y) : \Box(p_C \to \sigma) \to U(p_D \to \sigma) \quad (\to\mathrm{I})$$

$$R, (x, y) : (Pareto \wedge IIA) \to (\Box(p_C \to \sigma) \to U(p_D \to \sigma)) \quad (\to\mathrm{I})$$

An example of derivation

$$R, (x, y) : Pareto \wedge IIA$$

$$R, (x, y) : \Box(p_C \to \sigma)$$

$$R'', (x, y) : p_C$$

$$R'', (x', y') : p_D$$

$$R'', (y, y') : p_N$$

$$R'', (x', x) : p_N$$

$$\vdots$$

$$R', (x', y') : p_D \to \sigma$$

$$R, (x, y) : U(p_D \to \sigma) \quad (U\mathrm{I})$$

$$R, (x, y) : \Box(p_C \to \sigma) \to U(p_D \to \sigma) \quad (\to\mathrm{I})$$

$$R, (x, y) : (Pareto \wedge IIA) \to (\Box(p_C \to \sigma) \to U(p_D \to \sigma)) \quad (\to\mathrm{I})$$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \rightarrow \sigma)$

$R'', (x, y) : p_C$

$R'', (x', y') : p_D$

$R'', (y, y') : p_N$

$R'', (x', x) : p_N$

$\vdots$

$R', (x', y') : p_D \rightarrow \sigma$

An example of derivation

$R, (x, y) : \mathit{Pareto} \wedge \mathit{IIA}$

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$
$R'', (x', y') : p_D$
$R'', (y, y') : p_N$
$R'', (x', x) : p_N$
$R'', (x, y) : p_C \to \sigma \quad (\Box\mathrm{E})$
$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \rightarrow \sigma)$

$$\boxed{\begin{array}{l} R'', (x, y) : p_C \\ R'', (x', y') : p_D \\ R'', (y, y') : p_N \\ R'', (x', x) : p_N \\ R'', (x, y) : p_C \rightarrow \sigma \quad (\Box E) \\ R'', (x, y) : \sigma \quad (\rightarrow E) \\ \vdots \end{array}}$$

$R', (x', y') : p_D \rightarrow \sigma$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$
$R'', (x', y') : p_D$
$R'', (y, y') : p_N$
$R'', (x', x) : p_N$
$R'', (x, y) : p_C \to \sigma \quad (\Box E)$
$R'', (x, y) : \sigma \quad (\to E)$
$R'', (y, y') : \sigma \quad (Pareto)$
$R'', (x', x) : \sigma \quad (Pareto)$
$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$

$R'', (x', y') : p_D$

$R'', (y, y') : p_N$

$R'', (x', x) : p_N$

$R'', (x, y) : p_C \to \sigma \quad (\Box\mathrm{E})$

$R'', (x, y) : \sigma \quad (\to\mathrm{E})$

$R'', (y, y') : \sigma \quad (Pareto)$

$R'', (x', x) : \sigma \quad (Pareto)$

$R'', (x', y') : \sigma \quad (Cons)$

$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$R, (x, y) : Pareto \wedge IIA$

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$

$R'', (x', y') : p_D$

$R'', (y, y') : p_N$

$R'', (x', x) : p_N$

$R'', (x, y) : p_C \to \sigma \quad (\Box E)$

$R'', (x, y) : \sigma \quad (\to E)$

$R'', (y, y') : \sigma \quad (Pareto)$

$R'', (x', x) : \sigma \quad (Pareto)$

$R'', (x', y') : \sigma \quad (Cons)$

$R'', (x', y') : p_D \wedge \sigma \quad (\wedge I)$

$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$

$R'', (x', y') : p_D$

$R'', (y, y') : p_N$

$R'', (x', x) : p_N$

$R'', (x, y) : p_C \to \sigma \quad (\Box\text{E})$

$R'', (x, y) : \sigma \quad (\to\text{E})$

$R'', (y, y') : \sigma \quad (Pareto)$

$R'', (x', x) : \sigma \quad (Pareto)$

$R'', (x', y') : \sigma \quad (Cons)$

$R'', (x', y') : p_D \wedge \sigma \quad (\wedge\text{I})$

$R'', (x', y') : p_D \wedge \sigma \to \Box(p_D \to \sigma) \quad (IIA)$

$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$R, (x, y) : \Box(p_C \to \sigma)$

$R'', (x, y) : p_C$

$R'', (x', y') : p_D$

$R'', (y, y') : p_N$

$R'', (x', x) : p_N$

$R'', (x, y) : p_C \to \sigma \quad (\Box E)$

$R'', (x, y) : \sigma \quad (\to E)$

$R'', (y, y') : \sigma \quad (Pareto)$

$R'', (x', x) : \sigma \quad (Pareto)$

$R'', (x', y') : \sigma \quad (Cons)$

$R'', (x', y') : p_D \wedge \sigma \quad (\wedge I)$

$R'', (x', y') : p_D \wedge \sigma \to \Box(p_D \to \sigma) \quad (IIA)$

$R'', (x', y') : \Box(p_D \to \sigma) \quad (\to E)$

$\vdots$

$R', (x', y') : p_D \to \sigma$

An example of derivation

$$R, (x, y) : \Box(p_C \to \sigma)$$

$R'', (x, y) : p_C$
$R'', (x', y') : p_D$
$R'', (y, y') : p_N$
$R'', (x', x) : p_N$
$R'', (x, y) : p_C \to \sigma \quad (\Box\mathrm{E})$
$R'', (x, y) : \sigma \quad (\to\mathrm{E})$
$R'', (y, y') : \sigma \quad (Pareto)$
$R'', (x', x) : \sigma \quad (Pareto)$
$R'', (x', y') : \sigma \quad (Cons)$
$R'', (x', y') : p_D \wedge \sigma \quad (\wedge\mathrm{I})$
$R'', (x', y') : p_D \wedge \sigma \to \Box(p_D \to \sigma) \quad (IIA)$
$R'', (x', y') : \Box(p_D \to \sigma) \quad (\to\mathrm{E})$
$R', (x', y') : p_D \to \sigma \quad (\Box\mathrm{E})$

$$R', (x', y') : p_D \to \sigma \quad (UD)$$