Une introduction à la Théorie de l'Agrégation Logique

Mikaël Cozic

IHPST (Paris I-ENS Ulm-CNRS)
GREGHEC (HEC-CNRS), Logiques de l'Agir (Besançon) & DEC (ENS Ulm)

30/VI/2008

1. Agrégation des jugements et paradoxe doctrinal

scénario du réchauffement (SR)

- voici le genre de scénario qui sert à illustrer la TAL :

  3 experts doivent se prononcer sur les questions suivantes:

  ✓ les émissions de CO_2 dépassent-elles x ?
  ✓ si les émissions de CO_2 dépassent x, l'atmosphère va-t-elle se réchauffer ?
  ✓ l'atmosphère va-t-elle se réchauffer ?

- les experts peuvent avoir des jugements différents et l'objectif de la TAL est d'explorer les manières d'*agréger* rationnellement ces jugements pour former un jugement collectif

théorie du choix social et théorie de l'agrégation logique

- comme la théorie du choix social (TCS), la TAL est une *théorie de l'agrégation* - elle ne traite pas, par exemple, de la façon dont les jugements individuels sont formés et/ou évoluent à la faveur d'une délibération collective

- *prima facie*, *ce qui* est agrégé par chaque théorie est différent :

  ✓ la TCS agrège (en général) des *préférences*
  ✓ la TAL agrège des *jugements*

les attitudes propositionnelles

- préférences et jugements sont des *attitudes propositionnelles* : type mental + contenu ou objet

  ✓ Paul [croit que]$_{typ}$ [la France gagnera l'Euro]$_{cont}$
  ✓ Paul [désire que]$_{typ}$ [la France gagne l'Euro]$_{cont}$

- on suppose souvent que l'objet ou le contenu d'une attitude prop. est une **proposition** i.e. qchose susceptible d'être vrai ou faux (un "porteur de valeur de vérité"), que l'on distingue de l'énoncé ou de la phrase qui l'exprime

les préférences et les jugements

- on peut voir les préférences comme une attitude prop. binaire (qui met en jeu deux contenus) et comparative

  ✓ Paul [préfère que]$_{typ}$ [la prude lise l'*Amant de Lady Chatterley*]$_{cont_1}$ plutôt que [le libertin ne lise l'*Amant de Lady Chatterley*]$_{cont_2}$

- les jugements sont des attitudes prop. unaires (comme les croyances ou les désirs) = ils ont un (seul) contenu

  ✓ Paul [juge que]$_{typ}$ [l'atmosphère va se réchauffer]$_{cont}$

les préférences et les jugements, cont.

- il s'agit d'une analyse de surface ; on verra à la fin de l'exposé que certains "plongent" les préférences dans les jugements

  ✓ Paul $[\text{juge qu'}]_{typ}$ [il est préférable que la prude lise l'*Amant de Lady Chatterley* plutôt que le libertin ne le lise $]_{cont}$

- est-ce équivalent à

  ✓ Paul $[\text{préfère que}]_{typ}$ [la prude lise l'*Amant de Lady Chatterley*$]_{cont_1}$ plutôt que [le libertin ne lise l'*Amant de Lady Chatterley* $]_{cont_2}$

les jugements

- quel est le type mental des jugements ?

  ✓ / acceptation : (i) jugement suggère délibération ; (ii) pas de "jugement à titre d'hypothèse"
  ✓ / croyance : (i) le jugement est conscient ; (ii) pas sûr que le jugement ait une "direction d'ajustement" (Searle) univoque
  ✓ lien étroit avec l'**assertion** : on asserte typiquement p lorsque l'on juge que $[p]$
  on dit parfois qu'asserter p, c'est affirmer que p est vrai ; mais...
  ✓ catégorie très générale qu'on peut diviser selon les grandes familles de contenus

les jugements, cont.

- quels sont les contenus ?

  ✓ contenus factuels : l'atmosphère va se réchauffer
  ✓ contenus prescriptifs : il faut abaisser le taux directeur
  ✓ contenus évaluatifs : MC n'est pas un bon candidat

les contenus

- la TAL, on le verra en détails, fait deux hypothèses (implicites) fondamentales sur les contenus des jugements:

(H1) *hypothèse de structure logique* : les contenus entretiennent des relations logiques entre eux

✓ Paul [juge que]$_{typ}$ [les émissions de CO2 dépassent x]$_{cont_1}$
✓ Paul [juge que]$_{typ}$ [si les émissions de CO2 dépassent x, l'atmosphère va se réchauffer]$_{cont_2}$
✓ Paul [juge que]$_{typ}$ [l'atmosphère va se réchauffer]$_{cont_3}$

$cont_1$ et $cont_2$ impliquent logiquement $cont_3$

les contenus

(H2) *hypothèse de partage interindividuel* : les jugements des différents individus portent sur les mêmes contenus

✓ Paul $[\text{juge que}]_{typ}$ $[\text{MC est un bon candidat}]_{cont}$

✓ Jean $[\text{juge que}]_{typ}$ $[\text{MC est un bon candidat}]_{cont}$

- question : est-ce que (H1)-(H2) engagent à, ou excluent, des théories philosophiques ?
- un certain nombre de philosophes affirment que tout ou partie des énoncés évaluatifs n'expriment pas de proposition ou n'ont pas de valeur de vérité ; est-ce que cela pose pb pour (H1) ?
- probablement pas : tout le monde accorde que l'on raisonne (logiquement) sur des énoncés moraux par ex. (Geach 1965)

  ✓ Si une certaine action est mauvaise, alors encourager ton petit frère à la faire est mauvais
  ✓ Embêter le chat est une mauvaise action
  ✓ Encourager ton petit frère à embêter le chat est mauvais

- la propriété (non constestée) des énoncés évaluatifs suivant laquelle ils peuvent être des composants d'énoncés plus complexes et peuvent entrer dans des raisonnements logiques constitue surtout un problème pour les "non factualistes" (Frege-Geach Point)
- lesquels essaient typiquement de dériver cette propriété de contraintes sur la cohérences des attitudes propositionnelles engagées dans les jugements évaluatifs (Hare, Blackburn, Gibbard)

- concernant (H2), le point est plus délicat ; tension *prima facie* entre des théories subjectivistes/indexicales des énoncés évaluatifs et (H2)

  ✓ Paul *juge que* l'action de MC *est bonne*
  ✓ Paul *approuve* l'action de MC
  ✓ Jean *juge que* l'action de MC *n'est pas bonne*
  ✓ Jean *désapprouve* l'action de MC

  cf Frege (“La négation”, 1919) : “Le jury serait une sotte institution s'il n'était admis que chacun des jurés peut comprendre la question proposée dans le même sens que son voisin.”

scénario du réchauffement

- la TAL a son “paradoxe de Condorcet” : le paradoxe doctrinal ou paradoxe discursif
- 3 experts doivent se prononcer sur les propositions suivantes :

  p : les émissions de CO2 dépassent x
  q : l'atmosphère va se réchauffer
  $(p \rightarrow q)$: si les émissions de CO2 dépassent x, l'atmosphère va se réchauffer

- appliquons la règle majoritaire aux jugements des experts ("1" signifie que l'énoncé est accepté, "0" qu'il est rejeté) :

	p	$p \to q$	q
1	1	1	1
2	1	0	0
3	0	1	0
Maj.	1	1	0

scénario de la Commission de Spécialistes

- les 3 membres de la Commission de Spécialistes se prononcent sur la candidature de MC:

  p : MC est un bon enseignant
  q : MC est un bon chercheur
  r : MC est un bon candidat
  $r \leftrightarrow (p \wedge q)$: MC est un bon candidat ssi c'est un bon chercheur et un bon enseignant

	p	q	$r \leftrightarrow (p \wedge q)$	r
1	1	1	1	1
2	1	0	1	0
3	0	1	1	0
Maj.	1	1	1	0

2. Le cadre formel

le langage formel

- les contenus des jugements individuels sont représentés par des **formules** d'un langage formel L. Nous nous restreignons (sauf dans la dernière section) au langage de la logique propositionnelle.
- L est construit à partir de var.prop. L_0 et des connecteurs booléens $\vee, \wedge, \neg, \rightarrow, \leftrightarrow$:

(i) $L_0 \subseteq L$

(ii) si $\phi, \psi \in L$, alors $(\phi \vee \psi)$, $(\phi \wedge \psi)$, $(\phi \rightarrow \psi)$, $(\phi \leftrightarrow \psi)$ et $\neg\phi \in L$

(iii) seules les expressions engendrées par un nombre fini d'application de (i)-(ii) sont des formules

relations logiques

- les formules de L entretiennent certaines *relations logiques*:

(i) $(p \wedge q)$ et p: si $(p \wedge q)$ est vrai, alors p aussi. On dit que $(p \wedge q)$ a pour **conséquence logique** p, ce que l'on note $(p \wedge q) \models p$

(ii) p et $\neg p$: si p est vrai, $\neg p$ ne l'est pas et inversement. p et $\neg p$ sont logiquement incompatibles. L'ensemble $\{p, \neg p\}$ est **logiquement incohérent**.

sémantique

- deux approches pour caractériser ces relations logiques : la syntaxe (règles de transformations, par ex. modus ponens) et la sémantique (fait appel à l'interprétation du langage)
- une **valuation** est une fonction $V : L \rightarrow \{0, 1\}$ qui associe une valeur de vérité (le vrai 1, le faux 0) à chaque formule du langage en respectant le sens des connecteurs logiques
- exemple : si $V(p) = V(q) = 1$, alors $V(p \wedge q) = 1$ (et réciproquement)

sémantique, cont.

- un ensemble de formules Γ a pour **conséquence (logique)** une formule ϕ (noté $\Gamma \models \phi$) ssi toute valuation qui rend vraies *toutes* les formules de Γ rend vraie ϕ.
- un ensemble de formules Γ est **(logiquement) cohérent** s'il existe une valuation qui rend vraies toutes ses formules
- exemples :
  SR : $\{p, p \rightarrow q\} \models q$
  SComSpé : $\{p, q, r \leftrightarrow (p \wedge q)\} \models r$

l'agenda

- les agents n'ont généralement pas à se prononcer sur l'ensemble des formules d'un langage donné, mais sur un ensemble restreint de ces formules que l'on appelle l'**agenda** et que l'on note $X \subseteq L$.
- si les individus doivent se prononcer sur ϕ, ils doivent pouvoir juger que $\neg\phi$ est le cas. Intuitivement, un agenda contient donc des paires $\phi/\neg\phi$.
- Pour simplifier, on élimine les double négations et l'on suppose que pour tout agenda X, si $\phi \in X$ alors $\sim \phi \in X$ où

  ✓ $\sim \phi = \neg\phi$ si ϕ n'est pas la négation d'une formule

  ✓ $\sim \phi = \psi$ si $\phi = \neg\psi$

l'agenda du réchauffement

- scénario du réchauffement
  p : les émissions de CO2 dépassent x
  q : l'atmosphère va se réchauffer
  $(p \rightarrow q)$: si les émissions de CO2 dépassent x, l'atmosphère va se réchauffer
- agenda conditionnel
  $X = \{p, (p \rightarrow q), q, \}^{+neg}$

l'agenda de la ComSpé

- scénario de la ComSpé
  p : MC est un bon enseignant
  q : MC est un bon chercheur
  r : MC est un bon candidat
  $r(p \leftrightarrow q)$: MC est un bon candidat ssi c'est un bon chercheur et un bon enseignant
- agenda conjonctif
  $X = \{p, q, r, r \leftrightarrow (p \wedge q)\}^{+neg}$

l'agenda et sa structure logique

- les relations logiques entre les formules de l'agenda jouent un rôle crucial dans les problèmes d'agrégation
- dans les agendas du SR et du SComSpé qui engendrent des jugements collectifs incohérents par la règle majoritaire, il y a des dépendances logiques entre les formules de l'agenda
- s'il n'y a pas de dépendance logique (non triviale) entre les formules de l'agenda, par ex. si l'agenda contient seulement des formules atomiques et leur négation, la règle majoritaire n'engendre pas de jugements collectifs incohérents

les ensembles de jugement

- les attitudes des individus relativement aux (contenus représentés par les) formules de l'agenda sont représentées de la manière suivante : à chaque individu i est associé l'ensemble $A_i \subseteq X$ des formules que i accepte (ou prétend accepter) - l'**ensemble de jugement** de i.
- deux hypothèses de "rationalité" (?) sur les A_i :

(i) cohérence logique

Dans SR, $A = \{p, p \rightarrow q, \neg q\}$ n'est pas cohérent

les ensembles de jugements, cont.

(ii) complétude ($\forall \phi \in X$, $\phi \in A_i$ ou $\sim \phi \in A_i$)

Dans SR, $A' = \{p\}$ n'est pas complet.

- différence substantielle entre cohérence (pas de contradiction) et complétude (prendre position sur chaque question posée) - cette dernière est critiquée par ex. par Gärdenfors (2006)
- contraintes plus faibles : cohérence + clôture déductive ($\phi \in X$ et $A_i \models \phi \Rightarrow \phi \in A_i$) - on obtient une notion formellement identique aux *belief sets* de la révision des croyances (AGM)
- On note $\mathcal{A}$ l'ensemble des ensembles de jugements cohérents et complets.

règle d'agrégation de jugements

- un **profil** d'ensembles de jugements individuels est un n-uplet $(A_1, ..., A_n) \in \mathcal{A}^n$
- une règle d'**agrégation de jugements** est une fonction F qui, à chaque profil d'ensemble des jugements $(A_1, ..., A_n)$ admissible, associe un *ensemble de jugements collectif* $F(A_1, ..., A_n) = A \subseteq X$.
- règle majoritaire formule par formule : $F(A_1, ..., A_n) = \{\phi \in X : |\{i : \phi \in A_i\}| > |\{j : \phi \notin A_j\}|\}$ (si n impair)

approche sémantique

- le cadre dominant est *logique et syntaxique* : les arguments et la valeur d'une règle d'agrégation de jugements, ce sont des ensembles de formules.
- approche logique et sémantique (Pauly & Van Hees 2007): au lieu d'ensembles de jugements on a des *valuations*
- sous les hypothèses de cohérence et de complétude, toute valuation $V \in \mathcal{V}$ sur L induit un ensemble de jugements pour l'agenda X ; et pour tout ensemble de jugements il existe une valuation V qui induit cet ensemble de jugement
- dans ce cadre, une règle d'agrégation des jugements est une fonction $F : \mathcal{V}^n \rightarrow \mathcal{V}$ où $\mathcal{V}$ est l'ensemble des valuations

exemple et généralisation

- l'expert 1 dans SR a pour ensemble de jugements $A_1 = \{p, p \rightarrow q, q\}$
  les valuations V tq $V(p) = V(q) = 1$ induisent A_1 i.e. $A_1 = \{\phi \in X : V(\phi) = 1\}$ pour tout V tq $V(p) = V(q) = 1$
- l'approche sémantique permet de généraliser naturellement à des valuations à t valeurs, ce qui permet de représenter des attitudes intermédiaires entre l'acceptation (valeur maximale, $t - 1$) et le rejet (valeur minimale, 0)

approche ensembliste

- à côté de ces approches logiques, une *approche ensembliste* (Dokow & Holzman 2008) inspirée par la théorie abstraite de l'agrégation (Wilson 1975, Fishburn & Rubinstein 1986)
- m objets sur lesquels se prononcer $\Rightarrow$ les contributions individuelles sont représentées par des m-uplets $x \in \{0,1\}^m$.
- les contraintes (sur les préférences, sur les relations logiques, etc.) sont représentées par un sous-ensemble $X \subseteq \{0,1\}^m$ de *vecteurs réalisables*.
- dans ce cadre, une règle d'agrégation est une fonction $F : X^n \to X$

exemples

- $X = \{(0,0,0),(0,1,0),(1,0,0),(1,1,1)\}$ est un ensemble de vecteurs réalisables que l'on peut voir comme la contrepartie ensembliste des ensembles de jugements possibles pour un agenda de type $\{p, q, p \wedge q\}$
- $X' = \{(0,0,1),(0,1,1),(1,0,0),(1,1,1)\}$ est un ensemble de vecteurs réalisables que l'on peut voir comme la contrepartie ensembliste des ensembles de jugements possibles pour un agenda de type $\{p, q, p \rightarrow q\}$

conditions U et R

- **Condition U** (domaine universel) : $Dom(F) = \mathcal{A}^n$
- **Condition R** (rationalité collective) : pour tout $(A_1, ..., A_n) \in Dom(F)$, $F(A_1, ..., A_n) \in \mathcal{A}$

- Le Paradoxe Doctrinal montre que la règle majoritaire formule par formule n'obéit pas à la Condition R puisqu'elle engendre des ensembles de jugements logiquement incohérents.

condition A (anonymat)

- **Condition A** (anonymat). Soit $\pi : \{1, ..., n\} \rightarrow \{1, ..., n\}$ une permutation et deux profils $(A_1, ..., A_n)$, $(A_{\pi(1)}, ..., A_{\pi(n)}) \in Dom(F)$. Alors $F(A_1, ..., A_n) = F(A_{\pi(1)}, ..., A_{\pi(n)})$
- la règle d'agrégation n'est pas sensible à l'identité des individus - ce qui exclut en particulier la dictature
- une RAJ est **dictatoriale** s'il existe un individu j tq $\forall (A_1, ..., A_n) \in Dom(F), F(A_1, ..., A_n) = A_j$

condition A et dictature

- exemple : $N = \{1, 2\}$, $\pi(1) = 2$ et $\pi(2) = 1$
  Soit F_2 la RAJ qui fait de l'individu 2 un dictateur :
  $F_2(A_1, A_2) = A_2$, mais
  $F_2(A_2, A_1) = A_1$
- la condition A est plus forte que la non-dictature ; par exemple, elle exclut également une règle majoritaire où l'un des individus aurait plus de poids dans le vote que les autres

la Condition S

- soit $part(\phi)$ **le parti** de ϕ i.e. l'ensemble des individus qui acceptent ϕ
- **Condition S** (systématicité) $\forall \phi, \psi \in X$, $\forall (A_1, ..., A_n)$, $(A'_1, ..., A'_n) \in Dom(F)$,
  si $part(\phi) = part'(\psi)$, alors $\phi \in A$ ssi $\psi \in A'$
- idée : le jugement collectif à l'égard de deux formules est le même si ces deux formules sont acceptées par les mêmes individus. Toute différence collective à l'égard de deux formules provient d'une différence entre le parti de l'une et le parti de l'autre. Les jugements collectifs surviennent sur les partis.

un premier théorème d'impossibilité

- **Théorème 1** (List & Pettit 2002)
  Si $n \geq 2$ et X contient au moins deux formules atomiques et leur conjonction (Condition X1), il n'existe aucune règle d'agrégation des jugements qui satisfasse les Conditions U, R, A et S.

un mot sur la preuve

- le point crucial est que les Conditions S et A impliquent conjointement que *la cardinalité du parti* d'une formule détermine son acceptation par le groupe (les jugements collectifs surviennent sur la cardinalité des partis) : si $|part(\phi)| = |part(\psi)|$, alors $\phi \in F(A_1, ..., A_n) \Leftrightarrow \psi \in F(A_1, ..., A_n)$.
- pour n pair, trivial ; pour n impair, il suffit de considérer le cas où $|part(p)| = |part(q)| = |part(\neg p \wedge q)|$

4. indépendance, manipulation et encore l'impossibilité

l'indépendance

- la Condition S a été critiquée : pourquoi faudrait-il que pour deux formules ϕ, ψ qui "parlent" de choses différentes l'identité du parti implique l'identité du jugement collectif ?
- dans le cas où $\psi = \phi$, la Condition S implique que l'attitude du groupe à l'égard d'une formule ne dépend que de l'attitude des individus à l'égard de cette formule - et pas de l'attitude des individus à l'égard des autres formules
- **Condition I** (Indépendance) $\forall \phi \in X$, $\forall (A_1, ..., A_n)$, $(A'_1, ..., A'_n) \in Dom(F)$,
  si $part(\phi) = part'(\phi)$ alors $\phi \in A$ ssi $\phi \in A'$

indépendance et manipulation de l'agenda

- comme en TCS, une des princpales motivations pour la Condition I est qu'elle empêche la manipulation

- exemple de manipulation de l'agenda sur le SR:
  • pour $X' = \{p, q\}^{+neg}$, $Maj(A'_1, A'_2, A'_3) = \{p, \neg q\}$ donc (trivialement)

$$Maj(A'_1, A'_2, A'_3) \models \neg q$$

  • pour $X = \{p, p \rightarrow q\}^{+neg}$, $Maj(A_1, A_2, A_3) = \{p, p \rightarrow q\}$ donc

$$Maj(A_1, A_2, A_3) \models q$$

- le passage de X à X' inverse le jugement collectif sur ϕ si $A(X) \models \phi$ et $A(X') \models \neg\phi$ ou inversement.

- **Proposition** (Dietrich 2006)
  Soit X' un agenda engendré par ajout ou suppression de formules ($\neq \phi$) à X et supposons que F et F' satisfassent les Conditions R et I.
  Si $\forall i$, A_i et A'_i sont compatibles et $\forall \psi \in X \cap X'$, $f_\psi = f'_\psi$ (voir ci-après), alors le passage de X à X' n'inverse pas le jugement collectif sur ϕ.

indépendance et manipulation des jugements

- autre type de manipulation : il se peut qu'un individu ait intérêt à ne pas exprimer sincèrement ses jugements pour les faire accepter par la collectivité
- F est **manipulable** par l'individu i au profil $(A_1, ..., A_n)$ sur la formule $\phi \in X$ ssi
(i) A_i et $F(A_1, ..., A_n)$ sont en désaccord sur ϕ, et
(ii) il existe une i-variante $(A_1, ..., A*_i, ..., A_n)$ tq A_i et $F(A_1, ..., A_i^*, ..., A_n)$ s'accordent sur ϕ

non manipulabilité

- **Condition M** (monotonie) $\forall \phi \in X$ et $(A_1, ..., A_i, ..., A_n)$, $(A_1, ..., A_i^*, ..., A_n) \in Dom(F)$, où $\phi \notin A_i$ mais $\phi \in A_i^*$, si $\phi \in F(A_1, ..., A_i, ..., A_n)$ alors $\phi \in F(A_1, ..., A*_i, ..., A_n)$
- idée : si un individu se met à accepter ϕ (alors qu'il ne le faisait pas) et que le groupe acceptait déjà ϕ, alors le groupe accepte toujours ϕ
- **Théorème** (Dietrich & List 2007)
  Si F satisfait la Condition U, alors F est non-manipulable ssi F satisfait les Conditions I et M.

versions sémantiques

- on peut faciliter la comparaison entre S et I si l'on passe à leurs versions sémantiques
- **Condition S'** (systématicité) Il existe une méthode $f : \{0,1\}^n \longrightarrow \{0,1\}$ t.q $\forall (A_1, ..., A_n) \in Dom(F)$, $F(A_1, ..., A_n) = \{\phi : f(\delta_1(\phi), ..., \delta_n(\phi) = 1\}$ où $\delta_i(\phi) = 1$ ssi $\phi \in A_i$.

- **Condition I'** (Pauly & Van Hees) $\forall\phi \in X$, il existe une méthode $f_\phi : \{0,1\}^n \rightarrow \{0,1\}$ t.q $\forall(A_1, ..., A_n) \in Dom(F)$,

  $\phi \in F(A_1, ..., A_n) \Leftrightarrow f_\phi(\delta_1(\phi), ..., \delta_n(\phi)) = 1$

  où $\delta_i(\phi) = 1$ ssi $\phi \in A_i$.

- Attention : la Condition I n'exige pas que la méthode soit uniforme d'une formule à l'autre : il se peut que pour ϕ on utilise le vote majoritaire simple, pour ψ une majorité qualifiée, etc.

un autre théorème d'impossibilité

- une RAJ est **dictatoriale** s'il existe un individu j tq $\forall (A_1, ..., A_n) \in Dom(F)$, $F(A_1, ..., A_n) = A_j$
- **Condition NC**: il existe deux profils $(A_1, ..., A_n)$ et $(A'_1, ..., A'_n) \in Dom(F)$ tq $F(A_1, ..., A_n) \neq F(A'_1, ..., A'_n)$
- **Théorème 2** (Dietrich 2006, Pauly & Van Hees 2007) *Si X satisfait la Condition X2 ou X'2*, F satisfait les Conditions U, R, I et NC ssi F est dictatoriale.

les Conditions X

- Condition X2 : X est **atomique** et contient au moins deux formules non trivialement dépendantes
  Un **atome** (de X) est une formule ϕ (de X) cohérente et maximale i.e. $\forall \psi \in X$, $\phi \models \psi$ ou $\phi \models \neg\psi$.
  X est atomique si l'ensemble de ses atomes est exhaustif i.e. toute valuation rend au moins vrai l'un d'entre eux

exemple

- ex : $X = Cl^{\sim,\wedge}\{p, q\}$
- les conjonctions de littéraux $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$, $\neg p \wedge \neg q$ sont des atomes : $\forall \phi \in X$, $p \wedge q \models \phi$ ou $p \wedge q \models \neg\phi$
- $X = \{p \wedge q, p \wedge \neg q, \neg p \wedge q\}^{+neg}$ n'est pas atomique : il n'y a pas d'atome qui est vrai pour la valuation V : $V(p) = V(q) = 0$
- Pauly & Van Hees (2007) ont un résultat analogue avec la Condition X2' : (i) clôture par les var. prop. (si $\phi \in X$ et p atomique apparaît dans ϕ, alors $p \in X$ (ii) si ϕ et $\psi \in X$ sont des littéraux, $\phi \wedge \psi \in X$
  notre ex. satisfait X2'

- l'agenda du SR $X = \{p, (p \to q), q, \}^{+neg}$ ne satisfait ni la Condition X2 (pas d'atomes) ni la Condition X2' (par ex., la conjonction des deux littéraux p et $\neg q$ n'est pas dans X)

5. Pareto et indépendance

où est passé Pareto ?

- Le Théorème 2 ressemble bcp au Théorème d'Arrow. Il manque cpdt l'équivalent de la Condition de Pareto.
- Mongin (2008) cherche à savoir "où est passée" la Condition P = factoriser la Condition I en deux Conditions dont l'une est une condition d'unanimité analogue à la Condition P.
- **Condition UP** (préservation de l'unanimité) $\forall \phi \in X$, $(A_1, ..., A_n) \in Dom(F)$, $part(\phi) = N \Rightarrow \phi \in F(A_1, ..., A_n)$

la Condition IA

- un **littéral** noté $\tilde{p}$ est une var.prop. ou la négation d'une var.prop.
- **Condition IA** (indépendance atomique) $\forall \tilde{p} \in X$, $\forall (A_1, ..., A_n), (A'_1, ..., A'_n) \in Dom(F)$, $part(\tilde{p}) = part'(\tilde{p}) \Rightarrow \tilde{p} \in A$ ssi $\tilde{p} \in A'$
- conséquence conditionelle :
  $\phi \models_Y \psi$ ssi $\{\phi\} \cup Y \models \psi$
  $\phi \models_d \psi$ ssi $\exists Y \subseteq X$ tq $\phi \models_Y \psi$ pour un Y cohérent avec ϕ et $\neg\psi$
  $\phi \models^* \psi$ ssi $\exists \phi_1, ..., \phi_k$ tq (i) $\phi_1 = \phi$, (ii) $\phi_k = \psi$ et (iii) $\phi_1 \models_d \phi_2 \models_d ... \models_d \phi_k$ (clôture transitive)

- **Théorème 4** (Mongin 2008)
  Si X satisfait la Condition X4, F satisfait les Conditions U, R, IA et UP ssi F est dictatoriale.
- Condition X4
  (i) X est clos par les var.prop.
  (ii) pour tous littéraux $\tilde{p}, \tilde{q} \in X$, $\tilde{p} \models^* \tilde{q}$
  (iii) il existe $\tilde{p}, \tilde{q} \in X$ et $Y \subseteq X$ tq $\tilde{p} \models_Y \tilde{q}$ et non $\tilde{q} \models_Y \tilde{p}$
  [*satisfait si* $\neg\tilde{p} \vee \tilde{q} \in X$]
  (iv) il existe $\tilde{p}, \tilde{q}, \tilde{r} \in X$ et $Y \subseteq X$ tq Y est compatible avec $\{\tilde{p}, \neg\tilde{q}, \neg\tilde{r}\}$, $\{\neg\tilde{p}, \neg\tilde{q}, \tilde{r}\}$, $\{\neg\tilde{p}, \tilde{q}, \neg\tilde{r}\}$ mais pas avec $\{\neg\tilde{p}, \neg\tilde{q}, \neg\tilde{r}\}$
  [*satisfait si* $\tilde{p} \vee \tilde{q} \vee \tilde{r} \in X$]

exemple

- $X = \{p, q, r, p \vee q \vee r, \pm p \vee \pm q, \pm p \vee \pm r, \pm q \vee \pm r\}^{+neg}$
  (i): immédiat
  (ii) et (iii): $\neg p \models^* q : \{\neg p, p \vee q\} \models q$
  (iv): p, q, r est un triplet privilégié

indépendance et indépendance atomique

- pourquoi la Condition IA plutôt que la Condition I ?
- objection à I : les raisons d'accepter ou de rejeter ϕ peuvent résider dans l'acceptation ou le rejet d'*autres* formules que ϕ
- défense de IA : l'objection précédente ne s'applique pas aux var.prop. car elles sont primitives. Les raisons qu'on aurait de les accepter seraient elles-mêmes primitives au sens où elle ne résideraient pas dans des raisons d'accepter d'autres formules de l'agenda
- les raisons d'accepter une formule suivent-elles sa forme logique ? Ne peut-on pas accepter q parce qu'on accepte p et $p \rightarrow q$?

caractérisation de la dictature

- les Thms précédents donnent des CS mais pas des CNS = caractérisation
- dans le cadre ensembliste, Dokow & Holzman cherchent (et trouvent !) des CNS sur l'ensemble des vecteurs réalisables pour qu'une règle d'agrégation qui satisfait les conditions U, P, I soit nécessairement dictatoriale.
- cette condition = Condition X'4 contient deux propriétés : la première est celle de "total blockedness" (Nehring & Puppe), la seconde est une condition algébrique ("non affineness")
- on va donner explicitement une caractérisation analogue mais formulée dans un cadre logique et syntaxique

le théorème

- **Théorème 4** (Dietrich & List 2007)
  Si X satisfait la Condition X4, F satisfait les Conditions U, R, I et UP ssi F est dictatoriale.

la Condition X4

- condition X3(1) : il existe un $Y \subseteq X$, minimalement incohérent et un $Z \subseteq Y$ de taille paire tq $Y - Z \cup \{\neg\zeta : \zeta \in Z\}$ est cohérent.
- condition X3(2) : pour toutes formules contingentes $\phi, \psi \in X$, $\phi \models^* \psi$

exemple

- $X = \{p, (p \rightarrow q), q\}^{+neg}$
- condition X3(1) : OUI : soit $Y = \{p, (p \rightarrow q), \neg q\}$; $Z = \{p, (p \rightarrow q)\}$ est tq $\{\neg p, \neg(p \rightarrow q), \neg q, \}$ est cohérent
- condition X3(2) : NON : on ne peut pas trouver de $\models_d$-chaîne qui aille, par ex., de $\neg p$ à $\neg q$.

6. TAL et TCS

un langage pour les préférences

- la TAL et la TCS ont manifestement d'étroites ressemblances ; quelles sont les relations exactes ?
- Dietrich & List (2007) : représenter les préférences individuelles strictes dans un langage du premier ordre avec identité construit à partir d'un prédicat binaire P, d'un symbole de constante par option a, b, c, de variables et des constantes logiques booléennes + $\forall$
- agenda X : les formules atomiques construites à partir des constantes du type aPb et leurs négations

rationalité des préférences

- les formules de l'agenda ne contiennent pas les contraintes de rationalité sur les préférences (asymétrie, transitivité, complétude)
- les propriétés sur les préférences forment une théorie T au sens logique :
  ✓ $\forall xy(xPy \rightarrow \neg yPx)$
  ✓ $\forall xyz((xPy \wedge yPz) \rightarrow xPz)$
  ✓ $\forall xy(\neg x = y \rightarrow (xPy \vee yPx))$
- ces contraintes sont intégrées dans la relation de conséquence logique : $\Gamma \models \phi$ iff $\Gamma \cup T \models_{LPO} \phi$
- on obtient alors le Théorème d'Arrow comme corrolaire du Théorème 4

Conclusion

quelques questions en chantier

- agrégation probabiliste
- distinction prémisses/conclusions
- délibération et révision des jugements individuels (voir tout à l'heure l'exposé de C. List)

agrégation probabiliste

- en général, la TAL considère des jugements binaires (oui/non), sans prendre en compte les *degrés* d'assentiment ou d'acceptation
- si l'on veut prendre en compte les degrés, le cadre probabiliste s'impose naturellement.
- dans la version sémantique de la TAL, on peut voir les ensembles de jugement (des valuations) comme des cas-limites de probabilité ("fully opiniated"), des **probabilités 0-1**

quelques résultats

- on dispose de plusieurs résultats en TAP, mais, attention, pas d'agenda !
- Théorème de possibilité : pour des probabilités quelconques, si F satisfait les (analogues probabilsites des) Conditions U, I, NC et la 0-Unanimité, alors il existe $\lambda_1, ...\lambda_n$ tq $\sum \lambda_i = 1$ et $F(\pi_1, ..., \pi_n)(E) = \sum \lambda_i \pi_i(E)$
- Théorème d'impossibilité : pour des probabilités 0-1, si F satisfait les (analogues probabilistes des) Conditions U, I et NC, alors F est dictatoriale

quelques suggestions

- si l'on compare les deux théorèmes, on peut se demander si les Théorèmes d'impossibilité de la TAL ne sont pas dûs à un cadre trop pauvre
- cela suggère un développement parallèle à la TCS avec les théorèmes de possibilité apparus à la suite de l'introduction d'attitudes prop. plus fines (comparaisons interpersonnelles d'utilité)

prémisses et conclusion

- dès les débuts de la TAL, on a distingué deux approches possibles de l'agrégation, une approché *fondée sur les prémisses* et une approche *fondée sur les conclusions*
- cela nécessite évidemment de distinguer dans l'agenda les prémisses des conclusions ; dans SR, supposons que prémisses = $\{p, p \rightarrow q\}$ et conclusion = q
- approche (majoritaire) fondée sur les prémisses :

	p	$p \rightarrow q$	q
1	1	1	1
2	1	0	0
3	0	1	0
Maj.	1	1	X

si l'on clôt déductivement, on obtient q

- approche (majoritaire) fondée sur la conclusion :

	p	$p \rightarrow q$	q
1	1	1	1
2	1	0	0
3	0	1	0
Maj.	X	X	0

- Dietrich & Mongin (2008, venez demain !) explore techniquement l'approche fondée sur les prémisses