

# Un bilan interprétatif de la théorie de l'agrégation logique<sup>1</sup>

Philippe Mongin<sup>1</sup>  
Centre National de la Recherche Scientifique & Groupe HEC

Franz Dietrich<sup>1</sup>  
London School of Economics

Juillet 2010

## Résumé

La théorie de l'agrégation des jugements, ou, comme elle est ici conçue, de l'agrégation logique, généralise celle du choix social en faisant porter la règle d'agrégation sur des jugements quelconques au lieu des seuls jugements de préférence. Elle procède du paradoxe doctrinal de Kornhauser et Sager et du dilemme discursif de Pettit, que l'article réexpose en soulignant leurs différences. Après ce préalable conceptuel, il reproduit les grandes étapes techniques de la théorie, depuis son premier théorème d'impossibilité, chez List et Pettit, jusqu'aux résultats entièrement généraux de Dietrich et Mongin. Il met en relief la réalisation collective du théorème canonique - par Dietrich et List, Dokow et Holzman, Nehring et Puppe - qui a donné sa méthode d'analyse à la théorie: elle consiste à caractériser mathématiquement les agendas d'impossibilité d'un agrégateur donné, c'est-à-dire les ensembles de propositions tels qu'il n'existe pas de fonction de jugement collectif vérifiant une certaine liste de conditions axiomatiques. La théorie est ici unifiée par le traitement logique du problème agrégatif, dont on défend la pertinence, et par la distinction initiale de ses deux variantes, le paradoxe doctrinal et le dilemme discursif, que l'on réélabore techniquement.

**Classification JEL:** D 70, D 71, D 79.

**Mots-clefs:** agrégation des jugements, agrégation logique, paradoxe doctrinal, dilemme discursif, logique générale, méthode des prémisses et méthode des conclusions, théorie du choix social.

## Abstract

The theory of judgment aggregation, or rather logical aggregation, as we conceive of it here, generalizes the theory of social choice by having the aggregation rule bear on judgments of any kind whatever, and not barely preference judgments. It derives from Kornhauser and Sager's doctrinal paradox and Pettit's

---

<sup>1</sup>Cet article développe la présentation de la théorie de l'agrégation logique faite par Mikael Cozic et les deux auteurs au colloque "New Developments in Social Choice and Welfare Theories. A Tribute to Maurice Salles" (Université de Caen Basse-Normandie, 10-12 Juin 2009). Il reprend aussi les présentations faites par le premier auteur au "Workshop on Judgment Aggregation" (Barcelone, 14-16 décembre 2009) et au colloque LOFT 2010 (Toulouse, 5-7 juillet 2010).

<sup>1</sup>Adresse: 1 rue de la Libération, 78350 Jouy-en-Josas, FR. Courriel: mongin@greg-hec.com

<sup>1</sup>Adresse: Houghton Street, London WC2A 2AE, GB. Courriel: f.dietrich@lse.ac.uk

discursive dilemma, two problems that we restate by emphasizing their differences. After this conceptual preliminary, we follow the main technical advances of the theory, from the first impossibility theorem proved by List and Pettit to the completely general results of Dietrich and Mongin. We stress the collective achievement of the canonical theorem - by Dietrich and List, Dokow and Holzman, Nehring and Puppe - which provided the theory with a specific method of analysis: it consists in mathematically characterizing the impossibility agendas of a given aggregator - i.e., the sets of propositions such that no collective judgment function exist with a certain list of axiomatic properties. The theory is unified here by the logical treatment of the aggregative problem, for which we claim relevance, and the initial distinction of this problem into the doctrinal paradox and the discursive dilemma, which we elaborate upon technically.

**JEL Reference Numbers:** D 70, D 71, D 79.

**Keywords:** Judgment Aggregation, Logical Aggregation, Doctrinal Paradox, Discursive Dilemma, General Logic, Premiss-Based vs Conclusion-Based Approach, Social Choice Theory.

# 1 Introduction

Les théories agrégatives contemporaines plongent leurs racines dans les analyses mathématiques du vote, développées en France dès la fin du XVIIIème siècle, et dans les formulations techniques de l'utilitarisme et de sa variante épurée, l'économie du bien-être, qui se succédèrent en Grande-Bretagne du XVIIIème au milieu du XXème siècle. Les économistes classiques, puis néo-classiques, firent grand cas de la seconde inspiration, mais ignorèrent complètement la première, jusqu'à ce qu'à ce qu'Arrow orchestre magistralement leur réconciliation dans *Social Choice and Individual Values* (1951). Le titre de l'ouvrage a fixé l'appellation reçue de la théorie qu'il expose, *social choice theory*, appellation inexacte en fait, parce que celle-ci part de la préférence, et non pas du choix, comme concept fondamental, et qu'elle s'énonce pour toute espèce de collectivité, la société entière n'étant qu'une illustration particulière. Avec la fonction dite non moins improprement de bien-être social, qui est définie sur les profils autorisés de préférences individuelles et qui prend ses valeurs dans l'ensemble autorisé de préférences collectives, avec les axiomes qui, posés sur cette fonction, en décrivent les propriétés désirables, avec les théorèmes de possibilité ou d'impossibilité qui résultent de conjonctions particulières de ces axiomes, la théorie arrowienne développe un formalisme susceptible de couvrir en une fois les problèmes agrégatifs que les deux traditions historiques, la française et la britannique, offraient dissociés. En effet, la notion de préférence, individuelle ou collective, peut incliner soit du côté de la fonction d'utilité, qui la "représente" suivant la conception moderne des économistes, soit du côté du choix, qui la "révèle" d'après cette même conception, et le vote n'est qu'une forme particulière de choix; c'est ainsi qu'Arrow et ses successeurs ont pu réunir Bentham avec Condorcet.

Mais si grand que soit le pas de généralisation effectué, il n'est pas encore suffisant. Car les préférences des individus ou de la collectivité entre deux états de choses n'épuisent pas les jugements qu'ils sont susceptibles de porter sur ces états, et il n'est pas moins légitime d'examiner l'agrégation d'autres jugements que ceux de préférence. "Louis préfère une longue réunion mensuelle à de courtes réunions hebdomadaires" se paraphrase en disant que Louis juge l'une préférable aux autres, ce qui fait sentir que la préférence est une espèce du genre constitué par le jugement; elle se définit comme un jugement *comparatif* porté au point de vue *évaluatif* spécifique de la *préférabilité*. Sur les mêmes états de choses, Louis peut former d'autres jugements, absolus ou de nouveau comparatifs, qui se placent à d'autres points de vue spécifiques, factuels ou de nouveau évaluatifs: une longue réunion mensuelle est fatigante, plus fatigante que de courtes réunions hebdomadaires, réussie, plus réussie que de courtes réunions hebdomadaires, et ainsi de suite. Il est même inévitable que Louis forme d'autres jugements que celui de préférence, parce que celui-ci, comme toute évaluation, s'appuie sur des jugements factuels antérieurs. Quand des Louis multiples s'expriment sur les deux plans de la préférence et de ce qui la fonde, la théorie agrégative devrait-elle s'en tenir aux résultats en ignorant les intermédiaires? La restriction paraît contre-productive, et elle est pourtant imposée en théorie du choix social, puisque celle-ci ne connaît pas d'autres jugements que ceux de préférence.

Une théorie nouvelle, qu'on nomme *de l'agrégation des jugements*, surmonte la limitation de celle du choix social. Son formalisme lui permet de représenter les jugements individuels et collectifs au-delà du cas particulier de la préférence, et par ce moyen, d'attaquer une classe entière de problèmes agrégatifs négligés.

Les auteurs de cette théorie enchaînent donc un pas de généralisation sur celui qu'Arrow et ses disciples réalisèrent en leur temps. Tout en renouvelant l'objet d'étude de leurs prédécesseurs, ils en reproduisent strictement la démarche axiomatique. Leur méthode générale consiste en effet à soumettre les jugements individuels et collectifs à certaines propriétés normatives qui sont le pendant des contraintes imposées à la préférence, puis à définir une fonction de jugement collectif, allant des profils autorisés de jugements individuels vers les jugements collectifs autorisés, qui est l'analogue rigoureux la fonction de bien-être social. Après avoir énuméré les propriétés que la fonction de jugement collectif pourrait satisfaire, ils étudient les assemblages de ces propriétés au travers de théorèmes d'impossibilité ou de possibilité.

Ces recherches axiomatiques sont déjà suffisamment nombreuses et diversifiées pour qu'il ne soit plus possible d'en rendre compte sans fixer un angle d'attaque. En accord avec les travaux spécialisés de ses auteurs, ce bilan interprétatif souligne la pente logicienne de la théorie de l'agrégation des jugements et il en expose les résultats sous cet angle particulier; du coup, il la désignera comme étant la théorie *de l'agrégation logique*. Plus informatif que l'autre, l'intitulé présente aussi l'avantage de marquer une séparation nette avec la théorie antérieure de l'agrégation des probabilités. Dans une conception large du jugement, qui est en fait celle du sens commun, la certitude subjective ne lui est pas inhérente, et l'énoncé qu'il produit peut revêtir une force moindre que celle des valeurs "vrai" et "faux", seules considérées par la logique ordinaire. Si la théorie de l'agrégation des jugements relevait de la conception large, elle devrait englober la théorie l'agrégation probabiliste; mais il n'en est rien, la première s'étant constituée sans l'appui de la seconde, et les deux s'ignorant encore l'une l'autre sauf exceptions.<sup>1</sup>

De fait, c'est la logique, et non pas la théorie des probabilités, qui a fourni la notion technique de jugement qui généralise celle de préférence. En substance, le sujet individuel ou collectif qui porte un jugement s'exprime sur la proposition qui lui est associée: ou bien il l'affirme en lui donnant la valeur "vrai", ou bien il la nie en lui donnant la valeur "faux". Il est aussi possible que le sujet ne porte pas de jugement, auquel cas il n'attribue pas de valeur de vérité du tout à la proposition. Ce schéma s'étoffe lorsqu'on introduit un langage symbolique dans lequel chaque proposition est représentée par une formule; à partir de là, il est loisible de doser plus ou moins les composantes sémantique et syntaxique de l'analyse. La logique propositionnelle classique suffit aux premiers résultats, et même à quelques résultats avancés, mais nous pousserons cet article dans la direction de logiques moins élémentaires.

C'est avec List et Pettit (2002) que, pour la première fois rigoureusement, la théorie agrégative s'est incorporé l'analyse logique du jugement qui précède. Antérieurement, deux juristes américains, Kornhauser et Sager (1986, 1993) avaient ébauché une conception agrégative du jugement, mais pris dans le sens judiciaire et non pas dans le sens philosophiquement général. Ils avaient montré que les tribunaux collégiaux étaient exposés au *paradoxe doctrinal* (*doctrinal paradox*): les opinions individuellement cohérentes des juges peuvent mener le tribunal dans son ensemble à une incohérence. Dans un article-charnière qui

---

<sup>1</sup>La théorie de l'agrégation des probabilités remonte aux années 1960 avec des travaux de Stone, Raiffa et Winkler. Les principaux résultats ont été acquis rapidement, de sorte que la synthèse de Genest et Zidekh (1986) reste d'actualité. Chez McConway (1981), la théorie reproduit le modèle axiomatique de celle du choix social, tandis que, chez Lehrer et Wagner (1981), elle revêt la forme d'une théorie du consensus applicable à l'activité scientifique.

prépare le formalisme introduit avec List, Pettit (2001) réexamine le paradoxe de Kornhauser et Sager, et l'ayant jugé trop spécial, propose de le reformuler comme un problème universel qu'il nomme le *dilemme discursif* (*discursive dilemma*). Nous approfondirons cette étape sémantique et montrerons - c'est une thèse de l'article - qu'il n'y avait pas de nécessité à transformer le paradoxe doctrinal en dilemme discursif pour le traiter analytiquement. Il peut servir de point de départ à une branche de la théorie de l'agrégation logique dont les résultats sont encore plus généraux que ceux du rameau principal. Nous suivrons ici Dietrich et Mongin (2010) sans reproduire l'ensemble de leur développement technique.

Sans que List et Pettit en aient eu connaissance, un spécialiste français de mathématique sociale, Guilbaud (1952), avait déjà procédé à la généralisation de la préférence en jugement. Inspiré par Arrow, qui venait de publier son ouvrage, mais aussi par Condorcet, qu'il aura contribué à tirer de l'oubli, Guilbaud reformule la théorie agrégative du premier comme l'aurait fait le second, c'est-à-dire non pas seulement pour des relations de préférences, mais pour tout ensemble d'"opinions" - Guilbaud préfère ce terme, venu de Condorcet, à celui de jugement, qu'il emploie cependant quelquefois. Hostile au boubakisme qui domine alors les mathématiques françaises, Guilbaud se refuse à la méthode axiomatique et même aux démonstrations générales, ce qui rend complexe d'évaluer ses anticipations, mais la tendance est aujourd'hui d'y voir un précurseur direct de la théorie de l'agrégation logique.<sup>2</sup>

S'il faut trouver au travail actuel son premier inspirateur, Condorcet s'impose avec *l'Essai sur l'application de l'analyse à la probabilité des décisions à rendre à la pluralité des voix* (1785) et ses autres mémoires ou articles sur la mathématique électorale. Sa méthode constante est de traiter la préférence comme l'acceptation ou le rejet de certaines propositions de préférabilité: l'électeur qui préfère A à B, B à C, A à C, accepte "A est préférable à B", "B est préférable à C", "A est préférable à C", et rejette les propositions contraires. On penserait, sous l'influence d'Arrow, que Condorcet ne fait que retranscrire un concept initial de préférence ordonnée, mais il n'en est rien. Il part des propositions et de leurs liaisons logiques supposées, et il n'attribue la propriété d'ordre à la relation de préférence qu'en vertu de ces liaisons logiques. La généralité supérieure de son raisonnement apparaît mieux dans le théorème du jury que dans le paradoxe du vote, mais celui-ci est déjà représentatif: il le commente en logicien, par exemple quand il écrit que les propositions choisies par la majorité forment un "système contradictoire" (*Essai*, p. LV-LVI). Que la contradiction logique porte sur la préférence est une particularité de l'exemple et non pas de la méthode. Relu de cette manière, qui est précisément celle de Guilbaud, Condorcet devient le précurseur lointain du paradoxe doctrinal, du dilemme discursif et de toutes les recherches actuelles.<sup>3</sup>

---

<sup>2</sup>Monjardet (2003) isolait chez Guilbaud un théorème et sa démonstration relevant du choix social arrowien. Eckert et Monjardet (2009) lui attribuent un théorème et une démonstration qui relèvent de l'agrégation logique, et cette lecture est plus fidèle que la précédente à l'inspiration que Guilbaud trouvait chez Condorcet. Sur le plan strictement technique, on peut apercevoir chez Guilbaud le premier usage démonstratif des filtres et des ultrafiltres, dont les notions venaient d'entrer dans les mathématiques de l'époque. Depuis Kirman et Sondermann (1972), la théorie de choix social et, désormais, celle de l'agrégation logique empruntent couramment ce procédé.

<sup>3</sup>Outre Guilbaud, Granger (1956) and Rashed (1974) ont participé à la redécouverte de la mathématique sociale de Condorcet. Leur interprétation du paradoxe du vote reste cependant marquée par la lecture arrowienne que nous proposons de dépasser.

L'article qui suit forme un long développement circulaire autour du paradoxe doctrinal. La section 2 l'expose en revenant à la source judiciaire initiale, puis le compare à sa réinterprétation comme dilemme discursif, laquelle motive le cadre formel posé à la section 3. A partir de là, les théorèmes d'impossibilité s'enchaînent. Celui de List et Pettit (Théorème 1) impose à la fonction de jugement collectif l'axiome discutable de systématisme, alors que ceux de la section 4, dus à Pauly et van Hees, Dietrich, Mongin, Nehring et Puppe (Théorèmes 2-5), ne supposent que l'axiome d'indépendance, normativement plus acceptable et d'ailleurs proche de la célèbre condition arrovienne d'indépendance des options non pertinentes (*independence of irrelevant alternatives*). La section 5 expose, en l'améliorant, la logique générale de Dietrich, qui a permis de dépasser la limitation des résultats précédents, formulés dans des logiques encore trop particulières. En s'aidant de cet outil, la section 6 énonce les théorèmes qui structurent le mieux le domaine aujourd'hui et qui peuvent, à ce titre, passer pour canoniques. Ils ont pour objet mathématique l'agenda, c'est-à-dire l'ensemble des formules logiques représentatives des propositions sur lesquelles opinent les individus et la collectivité, et comme ils imposent à cet objet des conditions non seulement suffisantes, mais nécessaires, pour que toute fonction de jugement collectif vérifiant une certaine liste d'axiomes soit dégénérée, ils constituent des théorèmes de possibilité non moins que d'impossibilité. Chacun des théoriciens de l'agrégation logique y a peu ou prou contribué, mais nous privilégierons la formulation synthétique de Dokow et Holzman (Théorèmes 6-7). La section 7 retourne au paradoxe doctrinal pour lui donner son développement théorique autonome; les résultats de Dietrich et Mongin (Théorèmes 8-9) sont l'analogie pour ce paradoxe de ce que les théorèmes canoniques sont pour le dilemme discursif.

## 2 Du paradoxe doctrinal au dilemme discursif

La théorie de l'agrégation logique trouve son origine immédiate dans l'analyse du fonctionnement collectif de l'institution judiciaire qu'ont proposée les deux juristes américains Kornhauser et Sager (1986, 1993; voir aussi Kornhauser, 1992). De cette analyse, les spécialistes de la théorie n'ont véritablement retenu que le *paradoxe doctrinal* (*doctrinal paradox*), et il ne leur sert guère qu'à introduire et motiver des théorèmes dont la généralité couvre beaucoup d'autres exemples, paradoxaux ou non. Nous restituons ici le problème juridique initial et les transformations qu'il a connues chez Pettit (2001), puis List et Pettit (2002). En le présentant comme un *dilemme discursif*, ces deux auteurs en ont facilité le traitement analytique, mais nous ferons sentir aussi les préoccupations conceptuelles qu'ils écartent.

Avant de définir abstraitement le paradoxe doctrinal, comme le font Kornhauser et Sager (1993), nous emprunterons l'exemple, désormais classique et même rebattu, dont ils servent pour l'illustrer.<sup>4</sup> Une partie  $P$ , le plaignant, a intenté une action civile contre une autre partie  $D$ , le défendeur, en alléguant la rupture d'un contrat passé entre eux. Le tribunal est composé de trois juges  $A$ ,  $B$  et  $C$ , qui doivent, en s'appuyant sur le droit des contrats, déterminer si  $D$  doit verser des dommages-intérêts à  $P$ , ce qu'on représentera par les formules logiques  $d$  ou  $\neg d$ . Par hypothèse, l'affaire soulève les deux questions de savoir si le contrat était valide ou non ( $v$  ou  $\neg v$ ) et s'il a été rompu ou non ( $r$  ou  $\neg r$ ),

---

<sup>4</sup>Le paradoxe doctrinal est mentionné pour la première fois chez Kornhauser (1992). Les bibliographies citent abusivement Kornhauser et Sager (1986), qui mettaient en évidence des problèmes agrégatifs distincts du paradoxe.

et le droit tranche dans tous les cas de réponses à partir de la règle suivante:  $D$  doit des dommages-intérêts à  $P$  si, et seulement si, le contrat était valide et que  $D$  l'ait rompu. Supposons que les délibérations des juges les conduisent aux réponses et conclusions suivantes:

$A$	$v$	$\neg r$	$\neg d$
$B$	$\neg v$	$r$	$\neg d$
$C$	$v$	$r$	$d$

Si le tribunal décide à la majorité simple sur l'affaire directement, il rendra la conclusion  $\neg d$ , défavorable au plaignant; mais s'il décide, toujours à la majorité simple, les deux aspects de l'affaire et tire ensuite sa décision de la règle, il recueillera les réponses  $a$  et  $r$ , d'où la conclusion  $d$ , favorable au plaignant.

Par la suite, nous reviendrons souvent sur cet exemple de base, remarquablement choisi. Comme il est permis de s'y méprendre, il vaut la peine d'ajouter qu'il simplifie outrageusement le droit américain des contrats. Kornhauser et Sager examinent des cas authentiques, mais dont la complexité nuit à l'effet de paradoxe, et ils ont donc conçu didactiquement, et d'ailleurs tardivement, celui-ci. Leur intention dernière était d'approfondir l'effet sur les contenus jurisprudentiels de la forme collective de la décision judiciaire. Concrètement, le système judiciaire américain réserve la collégialité aux tribunaux d'appel (*appellate courts*) des Etats et de l'Union. Tout un chacun connaît la Cour Suprême, dont les neuf juges se prononcent par délibérations et par votes sur les affaires de nature fédérale, mais à un degré moindre de notoriété, les cours d'appel des Etats, souvent composées de trois membres, et les cours suprêmes correspondantes, en reproduisent le fonctionnement.

Nous rappellerons d'abord la méthode d'analyse judiciaire de Kornhauser et Sager, car elle fixe les concepts dont ils se servent pour définir le paradoxe doctrinal. Une *affaire* (*case*) portée devant un tribunal présente en général plusieurs *aspects* (*issues*) au regard du corpus juridique applicable. La jurisprudence (*common law*) pèse plus que les textes normatifs (*statutes*) dans la constitution du droit anglo-américain, du moins en matière civile, de sorte que le corpus juridique se ramène souvent à la *doctrine juridique* (*legal doctrine*) qui synthétise la jurisprudence aux yeux des juges. Kornhauser et Sager se placent dans ce cas et mobilisent une conception préalable de la doctrine juridique dans laquelle il n'est pas nécessaire d'entrer ici.<sup>5</sup> Simplement résumée, leur méthode repose sur une triple idéalisation du processus judiciaire: la décision du tribunal sur l'affaire serait une réponse par oui ou par non à une question univoque (le plaignant a-t-il tort ou raison contre le défendeur?), il en irait de même pour les décisions relatives à chaque aspect de l'affaire (le plaignant a-t-il tort ou raison dans telle partie de ses allégations? dans telle autre?), enfin, la doctrine juridique dicterait une et une seule réponse du premier type une fois que les réponses du second type sont connues.

La dernière hypothèse est sujette à interprétations. Les exposés courants du paradoxe associent la formule logique  $d \longleftrightarrow v \wedge r$  à la doctrine juridique et ils présentent les délibérations des juges comme des inférences déductives menées à l'aide de cette prémisse supplémentaire. Pareille modélisation est naturelle et commode, mais il peut s'en trouver d'autres qui répondent moins sommairement à l'idée que la doctrine juridique dicte aux juges la décision sur l'affaire une fois

<sup>5</sup>On trouve le développement explicatif chez Kornhauser (1992).

qu'ils ont tranché les questions des aspects, et les divers écrits de Kornhauser et Sager semblent justement pointer dans ces autres directions.

Les concepts qui précèdent ne suffisent que si l'on s'occupe d'un juge unique; dans le cas d'un tribunal collectif, il faut aussi décrire l'opération qui transforme les réponses individuelles en jugement final. Kornhauser et Sager n'en conçoivent que deux formes: ou bien, le tribunal enregistre directement *les réponses individuelles sur l'affaire* et il leur applique une procédure convenue de décision collective, qui est d'ordinaire le vote à la majorité simple; ou bien, il enregistre *les réponses individuelles sur chaque aspect*, et c'est à elles, considérées tour à tour, qu'il applique la procédure de décision collective avant de tirer une réponse globale en s'appuyant sur la doctrine juridique acceptée. Le paradoxe doctrinal tient dans le fait que la première méthode, qu'on peut dire *affaire par affaire* (*case-by-case*), ne donne pas toujours le même résultat que la seconde, qu'on peut dire *aspect par aspect* (*issue-by-issue*). Telle est la définition précise qu'en donnent Kornhauser et Sager (1993, p. 10-12; voir déjà Kornhauser 1992, p. 453) et les juristes qui les ont suivis.<sup>6</sup>

Chez Condorcet, le paradoxe du vote sert à motiver une solution particulière, ou peut-être plusieurs selon les interprètes, mais il ne contient pas en lui-même les réponses qu'on peut lui apporter. Le paradoxe doctrinal est tout autre: le problème fixe par avance les réponses, puisqu'il se définit justement par leur antagonisme. Il a une qualité structurelle qui le rapproche déjà d'un théorème d'impossibilité; du coup, les interprétations n'en sont pas fixées une fois pour toutes. Celle que manifeste l'exemple didactique, par un collègue qui déciderait simultanément, n'est ni la seule concevable, ni la plus intéressante au point de vue juridique. Ainsi, *A*, *B* et *C*, siégeant séparément, pourraient avoir porté les jugements indiqués plus haut, alors qu'un quatrième juge *D*, sollicité après eux, se demanderait comment tenir compte de la jurisprudence qui lui est transmise: peut-il ne retenir que les réponses aux questions d'affaire, ou bien doit-il tenir compte des réponses aux questions d'aspect? Au départ, sans avoir encore imaginé le paradoxe, Kornhauser et Sager (1986) se préoccupaient du fonctionnement collectif des tribunaux sous l'angle principal de leur cohérence *diachronique*, et si, le paradoxe enfin dégagé, leur travail privilégie la cohérence *synchronique*, c'est à notre sens pour une simple raison de commodité intellectuelle. Des deux problèmes, le premier importe plus que le second, parce qu'il affecte tous les tribunaux, qu'ils soient collectifs ou individuels, et il les affecte d'autant plus que la jurisprudence tend à l'emporter sur les textes dans les sources du droit. Mais il est aussi vrai que ce problème est très difficile et qu'il est de bonne politique de commencer par l'autre.<sup>7</sup>

Si l'antagonisme des méthodes du jugement collectif s'apparente à un *paradoxe*, c'est que chacune peut se prévaloir d'un solide argument normatif en sa faveur. En décidant affaire par affaire, le tribunal respecte les délibérations de chaque juge intégralement, c'est-à-dire jusqu'au stade de la décision qu'il prendrait s'il était seul à se prononcer. En décidant aspect par aspect, le tribunal garantit que la décision du tribunal s'appuie sur des raisons du même

---

<sup>6</sup> Voir notamment Post et Salop (1991-1992), qui écrivent peu après Kornhauser et Sager, et Nash (2003), qui synthétise la littérature juridique sur le paradoxe doctrinal.

<sup>7</sup> L'expression *affaire par affaire* ne devrait pas figurer dans la version synchronique, car l'affaire est alors fixée; on peut supposer que Kornhauser et Sager font écho à la version diachronique d'abord mise en exergue. Parmi les théoriciens de l'agrégation logique, il n'y a que List (2004) à s'être attaqué à cette version.

ordre formel que les décisions des juges - en l'occurrence, compte tenu des idéal-isations qui précèdent, sur la doctrine juridique, la caractérisation des aspects et une inférence logiquement valide qui part de ces prémisses. Selon Kornhauser et Sager, "lorsque le paradoxe doctrinal se produit, le jugement et la raison sont immédiatement et inexorablement tirés dans des directions différentes" (1993, p. 25). Par "raison", les auteurs entendent classiquement la capacité justificative à fonder des conclusions - à leur trouver des raisons - en utilisant la droite logique; il est entendu que cette capacité s'applique efficacement au niveau de chaque juge, de sorte que c'est la raison *collective* qui est ici en cause. Par "jugement", ils veulent sans doute signifier, mais alors trop allusivement, le respect des délibérations *individuelles* qui sont menées jusqu'au stade des conclusions. Dans la terminologie de List et Pettit (2002, p. 94), plus explicite que celle des juristes, les principes correspondants se nomment la "rationalité collective" (*collective rationality*) et l'"attention aux individus" (*individual responsiveness*).

Au-delà du paradoxe qu'il fait naître psychologiquement, l'antagonisme des deux méthodes pose un *dilemme*, au sens habituel d'un choix forcé entre deux options insatisfaisantes; car la raison de désirer l'une est aussi bien la raison de ne pas désirer l'autre. Il s'agit même d'un dilemme *théorique*, puisqu'il oppose, on le voit, des principes fondamentaux de la décision collective. Pettit (2001), puis List et Pettit (2002), tirent le paradoxe dans cette direction où leurs prédécesseurs ne s'engageaient pas tout à fait; cependant, la transformation du problème qu'annonce leur nouveau concept de *dilemme discursif* ne réside pas dans cette inflexion limitée. Il ne faut pas la chercher non plus dans l'observation de ces auteurs que le paradoxe doctrinal se manifesterait en dehors des tribunaux, car celle-ci est d'une évidence trop immédiate. Pettit signale notamment les assemblées délibératives ou les exécutifs collégiaux des institutions démocratiques, les comités d'expertise et les instances de régulation économique, les clubs ou les autres groupes dont les membres se cooptent, et même, dans une mesure à déterminer, les partis, les syndicats, les églises. Simple affaire de constatations empiriques, la liste peut s'allonger à volonté. La seule adjonction qui n'aille pas de soi est *la société politique entière*, que Pettit choisit précisément d'inclure au prix d'un détour argumentatif; il en appelle à la conception délibérative de la démocratie, qu'il a étoffée par ailleurs. Ce nouvel approfondissement du paradoxe doctrinal excède toutefois aussi la sphère d'application normale du dilemme discursif.

Pour arracher le problème de Kornhauser et Sager à son origine judiciaire, Pettit écarte les deux termes d'aspect et d'affaire, et le reformule par le conflit de deux méthodes abstraites, celle qui se fonde sur les *prémisses* de la décision et celle qui se fonde sur ses *conclusions* (*premiss-driven way, conclusion-driven way*, 2002, p. 274). Diffusée par List et Pettit, cette nouvelle dénomination reflète mieux que la précédente l'extension effective du paradoxe doctrinal, et nous en ferons usage à notre tour, mais il faut admettre qu'elle n'en change pas la nature, et il ne saurait être question de ramener le dilemme discursif à cette banale clarification.

Ce qui distingue substantiellement les deux problèmes n'est pas tant la *sémantique*, comme le voudraient les suggestions qui viennent d'être écartées, que la *syntaxe*, comprise au sens des opérations sur les formules logiques représentatives des propositions. Voici comment List et Pettit (2002) reconstruisent l'exemple de base. Ils associent une formule à toutes les considérations qui influencent la délibération, et, pour la doctrine juridique, prennent  $d \longleftrightarrow v \wedge r$ .

Supposant alors que le tribunal  $T$  vote sur chaque formule simultanément, ils mettent en évidence la contradiction logique de l'ensemble de formules qu'il accepte:

$A$	$v$	$\neg r$	$\neg d$	$d \longleftrightarrow v \wedge r$
$B$	$\neg v$	$r$	$\neg d$	$d \longleftrightarrow v \wedge r$
$C$	$v$	$r$	$d$	$d \longleftrightarrow v \wedge r$
$D$	$v$	$r$	$\neg d$	$d \longleftrightarrow v \wedge r$

Alors que le paradoxe doctrinal se définissait par la divergence des moyens pour faire émerger une conclusion à partir de prémisses, *le dilemme discursif se définit par la contradiction au sein du jugement collectif total*, c'est-à-dire de l'ensemble des propositions délibérées, sans qu'on cherche à distinguer entre prémisses et conclusion. Ainsi présenté, le problème est facile à traiter par la logique propositionnelle ordinaire. Il ouvre la voie au formalisme de la théorie agrégative qu'on appelle, précisément, logique. Une fois cette théorie pleinement développée, il n'y aurait plus de raison de douter qu'on ait reformulé le problème comme il convenait.

Malgré tant d'avantages conceptuels et techniques, il reste que, si l'on se préoccupe d'approfondir le paradoxe doctrinal dans le contexte judiciaire où il est né, la voie qu'ouvrent List et Pettit n'est pas nécessairement la bonne. En effet, dans ce contexte, leur définition englobante du jugement collectif est discutable. Elle suppose que la doctrine juridique soit une proposition en tout point comparable aux propositions d'affaire et d'aspects, ce qui veut dire, plus précisément, (i) qu'elle se représente par une formule de la logique ordinaire, et (ii) qu'elle tombe comme les autres formules sous la règle de décision. La théorie de l'agrégation logique a suscité, chez Kornhauser et Sager, des réserves qu'ils n'ont pas complètement éclaircies et qui peuvent tenir soit aux hypothèses (i) et (ii), soit, plus radicalement, à l'hypothèse (iii) qu'une proposition donnée correspond à la doctrine juridique.<sup>8</sup> Nous ne reprendrons pas la dernière objection, qui interdirait pratiquement de recourir à la logique, mais la théorie présentée à la section 7 s'efforce de prendre en compte les deux premières.

La présente section avait pour but de comparer le paradoxe doctrinal et le dilemme discursif. Notre analyse de ce qui les distingue peut se résumer en la comparant avec celle de Pettit:

"Je décris le problème [le paradoxe doctrinal] sous sa forme générale comme un dilemme discursif. Je préfère le mot "discursif" parce que le problème en question n'est pas lié à l'acceptation d'une doctrine commune et ne suppose que l'entreprise de former des jugements de groupe sur la base de raisons. Je préfère le mot "dilemme" parce que le problème conduit à un choix dans lequel chaque options a ses difficultés propres et ne constitue pas un paradoxe dans un sens rigoureux" (Pettit, 2001, p. 272).

Nous avons dit que l'opposition du "paradoxe" et du "dilemme" était un simple choix d'accent et qu'il ne faisait pas de doute que le problème judiciaire se transposait ailleurs. La nouveauté du dilemme discursif nous semble résider dans une autre affirmation du passage: le dilemme dépasse la contrainte, incorporée au paradoxe, de la "doctrine commune". En effet, prise à l'échelle de la société entière, la délibération démocratique tend à ne rien laisser à l'extérieur, même ses propres conditions de fonctionnement; les libertés publiques, par exemple, ne restent fixes qu'aussi longtemps que le corps politique ne révisé pas la constitution qui les définit. A une moindre échelle collective, il est banal que les

<sup>8</sup>Le débat entre Kornhauser et Sager (2004) et List et Pettit (2005) n'identifie pas les différends avec une clarté entière.

organes délibératifs partent d'une table rase, en se donnant leurs règles et leurs missions. Ainsi, les désaccords ne sont pas toujours cantonnés comme ils le sont dans le paradoxe doctrinal. Cette observation est, selon nous, au fondement du dilemme discursif.

Plus clairement identifiée et hiérarchisée par List et Pettit (2002), elle devient chez eux la motivation d'une analyse généralisante du jugement collectif qui passe par les deux étapes suivantes: toutes les considérations présentes dans la décision, y compris ce qui tient lieu de doctrine commune, seront représentées par des formules, en l'occurrence de la logique ordinaire; et les formules ainsi réunies seront traitées comme si elles faisaient, toutes sans exception, l'objet de décisions individuelles et d'une décision collective. On retrouve, abstraitement reformulées, les hypothèses (i) et (ii) du paragraphe précédent. Une théorie agrégative d'un nouveau genre s'enchaîne à ce point du raisonnement.

### 3 Le cadre formel de la théorie de l'agrégation logique

La théorie se développe à partir des trois notions spécifiques de l'agenda, des ensembles de jugement et de la fonction de jugement collectif, ainsi que d'un petit groupe d'axiomes, posés sur la fonction de jugement collectif, qui rappellent immédiatement ceux que la théorie du choix social fait porter sur la fonction de bien-être social. Dans cette section et la suivante, nous présentons la théorie avec un minimum de détails logiques; nous n'y développons en fait que la composante du langage formel, renvoyant à la section 5 la définition complète de l'inférence et des notions logiques associées.

Par définition, un *langage de la théorie*  $\mathcal{L}$  est un ensemble de *formules*  $\varphi, \psi, \chi, \dots$  contenant des *symboles logiques* pris dans un certain ensemble  $\mathcal{S}$ . Il n'est pas nécessaire de préciser la nature des formules au-delà de cette unique restriction:  $\mathcal{S}$  contient le symbole de la négation booléenne  $\neg$  ("non") et  $\mathcal{L}$  est fermé pour ce symbole, ce qui veut dire que, si  $\varphi \in \mathcal{L}$ , alors  $\neg\varphi \in \mathcal{L}$ . Si l'ensemble  $\mathcal{S}$  contient d'autres symboles, ils seront pris parmi les connecteurs booléens restants,  $\vee$  ("ou"),  $\wedge$  ("et"),  $\rightarrow$  ("si..., alors..."),  $\leftrightarrow$  ("si et seulement si..., alors...") ou parmi les symboles d'opérateurs non booléens représentant des modalités ("il est obligatoire de ...", "il est désirable que...", "il est connu que ...", "si ... alors..." pris dans un sens non booléen, par exemple contrefactuel). Pour chaque symbole supplémentaire de  $\mathcal{S}$ , on pose la règle de fermeture correspondante; s'il s'agit de  $\wedge$ , on demandera que si  $\varphi \in \mathcal{L}$  et  $\psi \in \mathcal{L}$ , alors  $\varphi \wedge \psi \in \mathcal{L}$ , et de même pour les autres symboles.

Dans la classe très vaste des langages autorisés, un cas particulier se dégage, celui des *langages propositionnels*  $\mathcal{L}_{\mathcal{P}}$ . Leur définition fait intervenir un ensemble  $\mathcal{P}$  de formules élémentaires, ou *variables propositionnelles*, qui ne contiennent pas de symboles logiques, et un ensemble de symboles logiques  $\mathcal{S}$  égal à celui des cinq symboles de connecteurs booléens. En fait, comme ces connecteurs sont interdéfinissables, il est identique de poser  $\{\neg, \cdot\} \subset \mathcal{S}$ , en remplissant la place vide indifféremment par  $\vee, \wedge, \rightarrow$  ou  $\leftrightarrow$ . Nous distinguerons les langages propositionnels *classiques* et *non-classiques* selon que, respectivement,  $\mathcal{S}$  ne contient que des symboles booléens ou en contient d'autres en sus. Techniquement, un langage  $\mathcal{L}_{\mathcal{P}}$  se définit comme le plus petit ensemble qui contienne  $\mathcal{P}$  et qui vérifie les règles de fermeture pour les symboles de  $\mathcal{S}$ .

Par *calcul* de la théorie, nous entendons la donnée, simultanément, d'un langage  $\mathcal{L}$  et d'un système d'axiomes et de règles qui détermine les liaisons logiques entre les formules de ce langage. Il n'est pas plus nécessaire de préciser entièrement ce système - la logique proprement dite - que le langage à l'étape antérieure. La section 5 montrera qu'il suffit de se donner un symbole de *relation d'inférence*  $B \vdash \psi$ , défini pour  $B \subset \mathcal{L}$  et  $\psi \in \mathcal{L}$ , qui réponde à des restrictions très générales. Dans cette section et la suivante, nous nous en tenons au cas particulier des calculs propositionnels classiques. De tels calculs comportent en effet, pour leur partie logique, des systèmes d'axiomes et de règles bien connus et conformes à l'intuition, et ce sont eux qui fixeront temporairement le sens de la règle d'inférence - également symbolisée par  $\vdash$  - et de ses notions formelles associées: vérité logique, contradiction logique, cohérence. Simples cas particuliers au regard de la théorie générale, les calculs propositionnels classiques se désignent à l'attention par leurs propriétés élémentaires. Nous nous en sommes servis comme M. Jourdain de la prose lorsque nous avons formalisé l'exemple de base: notre langage était alors  $\mathcal{L}_{\mathcal{P}}$ , construit sur  $\mathcal{P} = \{v, r, d\}$  et  $S = \{\neg, \leftrightarrow\}$ , et l'ensemble  $\{v, r, \neg d, d \leftrightarrow v \wedge r\}$  était contradictoire au sens d'un système propositionnel classique.

Dans le langage  $\mathcal{L}$ , la théorie fixe un sous-ensemble  $X$  représentatif des propositions que les membres du groupe mettent à l'examen; c'est l'*agenda*. Il peut être vaste ou restreint suivant les applications, mais en toute généralité, la théorie a seulement besoin qu'il soit non-vide et que, à l'instar de  $\mathcal{L}$ , il satisfasse la clôture par négations. L'agenda que List et Pettit retiennent pour l'exemple de base s'écrit:

$$\overline{X} = \{v, r, d, d \leftrightarrow v \wedge r, \neg v, \neg r, \neg(d \leftrightarrow v \wedge r)\}.$$

Si l'on ajoutait dans  $\overline{X}$  les négations multiples, comme  $\neg\neg v, \neg\neg\neg v, \dots$ , la logique en assurerait finalement la simplification. Il vaut mieux anticiper le processus et définir les agendas comme des ensembles:

$$X = \{\varphi, \chi, \psi, \dots\}^{\pm},$$

dont les éléments  $\varphi, \chi, \psi, \dots$  sont des formules *positives*, c'est-à-dire ne commençant pas par  $\neg$ , chacune étant doublée de sa négation (ce qu'indique le symbole  $\pm$  mis en exposant). Pour simplifier, nous imposerons une restriction supplémentaire à celles dont la théorie se contente minimalement: nos agendas se composeront de formules *contingentes*, c'est-à-dire qui ne sont ni des vérités logiques, ni des contradictions logiques.

On représentera les jugements que portent soit les membres du groupe, soit le groupe lui-même, par des sous-ensembles de  $X$ , les *ensembles de jugement*, qui répondent éventuellement à certaines contraintes logiques détaillées ci-dessous; la plus attendue est la cohérence. Les ensembles de jugement seront notés par  $B, B', \dots$  génériquement, et par  $A_i, A'_i, \dots, A, A', \dots$  quand ils appartiennent à des individus  $i$  et au groupe qu'ils forment, respectivement. Une formule  $\varphi$  de ces ensembles représente une proposition, au sens d'un objet sémantique doté d'une valeur de vérité, et si elle sert à représenter aussi un jugement, au sens d'une opération cognitive, c'est en vertu de la règle d'interprétation naturelle: (R)  $i$  juge que  $\varphi$  ssi  $\varphi \in A_i$ , et la collectivité juge que  $\varphi$  ssi  $\varphi \in A$ . Nous traitons la formule  $\varphi$  dans cet énoncé comme si elle était *elle-même* la proposition qu'elle exprime; la facilité de style est banale en logique et nous nous la permettrons souvent par la suite. Grâce à la règle (R), les jugements

obéissent à une distinction de deux types de négation, interne et externe, qui n'a pas d'analogue au niveau des propositions ou des formules, celles-ci étant niées d'une seule manière. En effet, "juger que ne pas" ( $\neg\varphi \in A_i$ ) diffère de "ne pas juger que" ( $\varphi \notin A_i$ ). Une fois la relation d'inférence définie, la cohérence logique des ensembles de jugement reliera une négation à l'autre comme on l'attendrait: "juger que ne pas" entraîne "ne pas juger que", sans que la réciproque ait toujours lieu.

Par ce qui précède, on voit que la théorie de l'agrégation logique se rattache à une conception philosophique déterminée du jugement et de la proposition. Le langage  $\mathcal{L}$  représente toutes les propositions exprimables, donc susceptibles de devenir objets de jugements, mais seules les propositions représentées dans  $X$  le seront effectivement. Ainsi, les deux concepts sont dissociés à l'avantage de la proposition, qui couvre plus largement, ce qui est caractéristique de la conception moderne du jugement - depuis Frege et Russell - par opposition à l'ancienne - celle d'Aristote, qu'on retrouve encore chez Kant.<sup>9</sup> La théorie de l'agrégation logique mobilise en fait le concept frégeen d'*assertion*, ici rendu par  $\varphi \in A_i$  ou  $\varphi \in A$ , et l'on vérifie qu'elle honore le principe, qui est aussi typiquement frégeen et moderne, voulant que l'assertion reste indifférente aux transformations opérées par la logique. Peu importe, en effet, que  $\varphi$  soit positive ou négative, conditionnelle ou inconditionnelle, de type modal ou non; l'indicateur choisi pour l'assertion, qui est l'appartenance à un ensemble, fonctionne à l'identique. Sans doute est-il permis à un moderne de faire porter les distinctions logiques par le jugement, mais c'est alors par voie de conséquence; de cette manière, il parlera de "jugement positif" ou de "jugement négatif" sur  $\varphi$  suivant que  $\varphi$  ou  $\neg\varphi$  appartient à l'ensemble de jugement considéré; c'est uniquement par la négation mise ou non devant la formule qu'il différencie les deux cas.

Reprenant la description du cadre formel, nous définirons *l'ensemble des individus*  $\{1, \dots, n\}$ , avec  $n \geq 2$ , et la *fonction de jugement collectif*, qui associe un ensemble de jugement collectif à chaque configuration, ou *profil*, d'ensembles de jugement pour les  $n$  individus:

$$A = F(A_1, \dots, A_n).$$

Comme celle du choix social, la théorie de l'agrégation logique privilégie les groupes de cardinalité finie.<sup>10</sup> Généralisant la fonction arrovienne de bien-être social,  $F$  formalise les règles de décision que le groupe est susceptible d'appliquer aux formules de l'agenda. Dans sa définition ordinaire, la seule considérée ici, elle part du *domaine universel*, c'est-à-dire de l'ensemble de tous les profils concevables, étant donné les contraintes logiques posées sur les ensembles de jugement; ces contraintes, que nous allons maintenant préciser, influencent aussi le choix de l'espace d'arrivée.<sup>11</sup>

Un ensemble de jugement  $B$  répond normalement à des contraintes logiques, dont voici les principales, exprimées à l'aide de la relation d'inférence:

<sup>9</sup>La *Logique* de Kant parue en 1800, est une synthèse remarquable de la conception dite ancienne. C'est dans les *Recherches logiques* de Frege (1918-1919) qu'apparaît le mieux la conception dite moderne.

<sup>10</sup>Dietrich et Mongin (2007), suivis par Herzberg et Eckert (2010) et Herzberg (2010), ont traité des groupes infinis. Les résultats transposent ceux déjà obtenus en théorie du choix social, notamment par Kirman et Sonderman (1972).

<sup>11</sup>La théorie de l'agrégation logique a seulement commencé à étudier les domaines restreints (avec les travaux de List, 2003, et Dietrich et List, 2010).

- $B$  est *déductivement clos*, i.e., pour tout  $\varphi \in X$  et tout  $B$ , si  $B \vdash \varphi$ , alors  $\varphi \in B$
- $B$  est *cohérent*, i.e., pour tout  $\varphi \in X$ , non ( $B \vdash \varphi$  et  $B \vdash \neg\varphi$ )
- $B$  est *complet*, i.e., pour tout  $\varphi \in X$ , ou bien  $\varphi \in B$ , ou bien  $\neg\varphi \in B$

Ces propriétés se combinent entre elles pour donner des concepts inégalement forts d'ensembles de jugement. Voici les deux familles principales:

- l'ensemble  $D$  des ensembles de jugements *cohérents et complets*, qui satisfont les trois propriétés ou, de manière équivalente, les deux dernières, la première suivant alors facilement;

- l'ensemble  $D^* \supseteq D$  des ensembles de jugements *cohérents et déductivement clos*, tels que définis par les deux premières propriétés.

A partir de là, on peut combiner librement les définitions pour contraindre le domaine de  $F$  et son espace d'arrivée en appliquant l'hypothèse de domaine universel, mais nous considérerons seulement les cas suivants:

- (i)  $F : D^n \rightarrow D$ ,
- (ii)  $F : D^n \rightarrow D^*$ ,
- (iii)  $F : (D^*)^n \rightarrow D^*$ ,
- (iv)  $F : D^n \rightarrow 2^X$  ou  $F : (D^*)^n \rightarrow 2^X$ .

Avec (iv), les ensembles de jugement collectifs échappent à toute contrainte logique et sont ainsi traités différemment des ensembles individuels. Ce cas sert uniquement à faciliter l'exposition des autres, sur lesquels portent les résultats de la théorie. Au départ, elle n'a traité que le cas (i), commode pour les démonstrations, mais douteux au regard de la psychologie cognitive, qui privilégierait des hypothèses logiques plus faibles. On peut aussi - plus subtilement - contester le choix de  $D$  au nom de la conception moderne du jugement qui sous-tend le formalisme. En interdisant l'abstention, il élimine la possibilité qu'elle offre, et que n'offrait pas l'ancienne, *de traiter d'une proposition sans qu'on doive l'asserter ou asserter sa négation*. De même, la théorie perd la distinction de la négation interne et la négation externe: "ne pas juger que" devient équivalent à "juger que ne pas". Il y a donc des raisons diverses à développer les cas fondés sur  $D^*$ , soit (ii) et (iii).

La fonction de jugement collectif  $F$  doit formaliser les règles de décision que se donnent concrètement les groupes. Deux moyens se présentent pour qu'elle joue son rôle: ou bien l'on précisera  $F$  au point de la faire coïncider avec une règle déterminée, ou bien l'on déterminera  $F$  par des conditions axiomatiques, représentatives des principes auxquelles obéissent les règles particulières. La même dualité se retrouve *mutatis mutandis* en théorie du choix social à propos de la fonction de bien-être social, et comme ce précédent l'a montré, on obtiendra le plus si l'on recourt aux deux moyens à la fois.<sup>12</sup> Le *vote majoritaire proposition par proposition (proposition-wise majority voting)*, qui est la règle de décision associée à l'exemple de base, illustrera la démarche. Cette règle se définira ici comme la fonction de jugement collectif  $F_{maj} : D^n \rightarrow 2^X$  telle que, pour tout profil  $(A_1, \dots, A_n)$  du domaine,

$$F_{maj}(A_1, \dots, A_n) = \{\varphi \in X : |\{i : \varphi \in A_i\}| \geq q\} ,$$

avec  $q = \frac{n+1}{2}$  si  $n$  est impair et  $q = \frac{n}{2} + 1$  si  $n$  est pair.

---

<sup>12</sup>Selon Mongin (2003), cette dualité rapproche la méthode axiomatique du choix social de celle de la logique. La définition des règles joue le rôle d'une sémantique par rapport à la syntaxe constituée par les axiomes, et les théorèmes de représentation deviennent ainsi des approximations des théorèmes de complétude prouvés en logique.

L'espace d'arrivée n'est pas  $D$  parce que  $F_{maj}$  ne tranche pas toujours l'égalité entre  $\varphi$  et  $\neg\varphi$  quand  $n$  est pair, et il n'est même pas  $D^*$ , puisque le dilemme discursif correspond à un profil  $(A_1, \dots, A_n)$  qui rend  $F_{maj}(A_1, \dots, A_n)$  incohérent. Après qu'on a formalisé le vote majoritaire proposition par proposition, il est souhaitable d'introduire ses propriétés normatives saillantes. Dans cette section, nous en retenons trois, définies abstraitement pour toute  $F$  concevable, grâce auxquelles un théorème d'impossibilité - le premier apparu par ordre chronologique - se démontre aisément. La liste des propriétés se développera en section 4 avec des résultats plus avancés.

**Systématicité.** Pour tout couple de formules  $\varphi, \psi \in X$ , et pour tout couple de profils  $(A_1, \dots, A_n), (A'_1, \dots, A'_n)$ , si, pour tout  $i = 1, \dots, n$ ,  $\varphi \in A_i \Leftrightarrow \psi \in A'_i$ , alors

$$\varphi \in F(A_1, \dots, A_n) \Leftrightarrow \psi \in F(A'_1, \dots, A'_n).$$

D'après la systémativité, le groupe, devant un profil d'ensembles de jugement individuels, se prononce sur une formule exactement de la même manière que, devant *une autre* profil, il se prononcerait sur *une autre* formule, dès lors que la répartition des jugements individuels sur la formule du premier profil coïnciderait avec la répartition des jugements individuels sur la formule du second profil. La règle  $F_{maj}$  vérifie manifestement la systémativité, dont l'analogue en théorie du choix social, est la neutralité (voir d'Aspremont, 1985).

Nous dirons qu'une fonction de jugement collectif  $F$  est une *dictature* s'il existe un individu  $j$  tel que, pour tout profil  $(A_1, \dots, A_n)$ ,

$$F(A_1, \dots, A_n) = A_j.$$

Compte tenu du domaine de  $F$ , il n'existe qu'un seul  $j$  par dictature, et on le nomme le *dictateur*. A l'évidence,  $F_{maj}$  satisfait la condition:

**Non-dictature.**  $F$  n'est pas une dictature.

La fonction  $F_{maj}$  vérifie en fait une condition beaucoup plus forte que celle-ci:

**Anonymat.** Pour tout profil  $(A_1, \dots, A_n)$ , si  $(A'_1, \dots, A'_n)$  s'obtient de  $(A_1, \dots, A_n)$  en permutant les individus, alors

$$F(A_1, \dots, A_n) = F(A'_1, \dots, A'_n).$$

Le parallélisme avec la théorie du choix social est encore une fois manifeste.

Le premier résultat de la théorie mettait en évidence le conflit de l'anonymat et de la systémativité sous une condition très peu contraignante imposée à  $X$  (List et Pettit, 2002, Theorem 1). En fait, le conflit peut s'exprimer plus fortement comme celui de la non-dictature et de la systémativité (Pauly et van Hees, 2006, Theorem 4) et nous le présentons donc dans cette version améliorée.

**Théorème 1 (Pauly et van Hees, 2006, généralisant List et Pettit, 2002)** Soit un langage propositionnel classique  $\mathcal{L}_{\mathcal{P}}$  avec  $\mathcal{S} = \{\neg, \wedge\}$ ; soit deux variables propositionnelles distinctes  $a, b \in \mathcal{P}$  telles que  $a, b, a \wedge b \in X$ ; alors, il n'existe pas de  $F : D^n \rightarrow D$  vérifiant à la fois la non-dictature et la systémativité.

Comme  $F_{maj}$  vérifie la non-dictature et la systématique sur  $D^n$ , il faut, par contraposition du théorème, que son espace d'arrivée diffère de  $D$ . Avec  $n$  impair, les ensembles de jugement collectifs étant toujours complets, il faut que l'un des  $F_{maj}(A_1, \dots, A_n)$  soit incohérent. Cette déduction s'accorde avec ce que l'exemple de base donnait à connaître empiriquement, sachant qu'il vérifie les conditions demandées à  $\mathcal{L}_{\mathcal{P}}$  et  $X$ . Par delà telle ou telle application, le Théorème 1 approfondit le dilemme discursif: on retrouve la rationalité collective dans l'hypothèse, inscrite dans la définition de  $F$ , que cette fonction va dans  $D$ , et l'attention aux individus dans l'axiome de non-dictature; mais que faire de l'axiome de systématique, qui n'évoque ni un principe, ni l'autre? Quoique List et Pettit préfèrent souligner la continuité des analyses formelle et informelle, il nous semble que le Théorème 1 vaut surtout parce qu'il modifie les intuitions premières. Le problème du jugement collectif apparaît en fait *comme un trilemme dont la systématique forme le terme supplémentaire*. Contrairement aux deux autres, cette propriété n'a pas de valeur normative, et on ne peut la défendre que pour ses avantages techniques éventuels; pourtant, elle participe non moins qu'elles à l'impossibilité du jugement collectif.<sup>13</sup>

A hauteur de sa généralité abstraite, le théorème couvre bien d'autres règles que  $F_{maj}$ . Nous nous arrêterons sur les variantes  $F_{maj}^q$  qui consistent à poser un *quota de majorité qualifiée*  $1 \leq q \leq n$ , c'est-à-dire sur les fonctions de jugement collectif  $F_{maj}^q : D^n \rightarrow 2^X$  ainsi définies: pour tout profil  $(A_1, \dots, A_n)$  du domaine,

$$F_{maj}^q(A_1, \dots, A_n) = \{\varphi \in X : |\{i : \varphi \in A_i\}| \geq q\}.$$

Dans le cas-limite  $q = n$ , une formule est acceptée collectivement si et seulement si tous les individus l'acceptent; on formalise ainsi la *règle de l'unanimité*, qui a son analogue en théorie du choix social (voir Sen, 1970). Manifestement, aucune des ces fonctions ne peut aller dans  $D$ , puisque certains ensembles de jugement collectifs sont incohérents avec  $q$  pris en-dessous des seuils de majorité ordinaires et que d'autres sont incomplets avec  $q$  pris en-dessus. Les  $F_{maj}^q$  vérifiant la non-dictature et la systématique, le Théorème 1 s'impose à elles, et la méthode axiomatique révèle ici le pouvoir unificateur dont on lui fait le mérite.<sup>14</sup>

En résumé, sur l'arrière-plan banal d'un calcul propositionnel classique, des concepts nouveaux prennent forme, l'agenda, les ensembles de jugement individuels et collectifs, et la fonction de jugement collectif pour relier les premiers aux seconds. Le dernier concept donne à la théorie de l'agrégation logique le moyen de traiter à la fois de règles particulières et de conditions axiomatiques. Elle s'attache à l'effet d'ensemble des conditions en précisant les agendas, et c'est ainsi qu'elle transforme le dilemme discursif en théorème d'impossibilité du jugement collectif. Ce premier résultat illustre la théorie en mouvement, mais il ne fait qu'en effleurer les possibilités démonstratives.

## 4 Théorèmes fondés sur l'axiome d'indépendance

L'étude des règles de vote, formalisées comme des fonctions de jugement collectif particulières, n'a pas encore porté tous ses fruits. D'autres propriétés

<sup>13</sup>Selon l'un des premiers critiques de List et Pettit, Chapman (2002), la systématique oblitère la dissymétrie fondamentale des prémisses et des conclusions, les premières servant de raisons aux secondes, mais non l'inverse. La faiblesse normative de l'axiome est, plus généralement, qu'il ignore la différence sémantique des propositions.

<sup>14</sup>Les règles à quotas sont définies comme chez Dietrich et List (2007a), qui, après Nehring et Puppe (2002, 2008), en font l'étude détaillée; voir aussi Dietrich (2010).

attendent d'être explicitées, au premier chef celle qui demande à la collectivité de reproduire les jugements unanimes des individus. Le cadre formel de la théorie l'autorise à faire porter l'unanimité ou bien sur les ensembles de jugement pris en bloc, ou bien - plus fortement - sur leurs formules considérées une à une. On facilite le parallélisme avec la systémativité et les conditions qui vont suivre si l'on opte pour la première des deux variantes; elle est aussi l'analogie des conditions de Pareto en théorie du choix social.

**Préservation de l'unanimité.** Pour toute formule  $\varphi \in X$  et tout profil  $(A_1, \dots, A_n)$ , si pour tout  $i = 1, \dots, n$ ,  $\varphi \in A_i$ , alors  $\varphi \in F(A_1, \dots, A_n)$ .

Par ailleurs, la systémativité donne lieu à une variante affaiblie, donc aussi vérifiée par les règles de vote de la section précédente:

**Indépendance.** Pour toute formule  $\varphi \in X$  et tout couple de profils  $(A_1, \dots, A_n)$ ,  $(A'_1, \dots, A'_n)$ , si, pour tout  $i = 1, \dots, n$ ,  $\varphi \in A_i \Leftrightarrow \varphi \in A'_i$ , alors

$$\varphi \in F(A_1, \dots, A_n) \Leftrightarrow \varphi \in F(A'_1, \dots, A'_n).$$

L'indépendance revient à limiter l'axiome de la section précédente au cas où  $\varphi = \psi$ . Elle fait disparaître l'élément conceptuel de la *neutralité*, c'est-à-dire l'indifférence au contenu sémantique des propositions, tout en préservant un autre élément conceptuel qui s'imbriquait dans celui-là: le jugement collectif sur  $\varphi$  dépend des seuls jugements individuels sur  $\varphi$ .<sup>15</sup> En d'autres termes, l'ensemble  $A$  se définit à partir des ensembles  $A_1, \dots, A_n$  *formule par formule* (*formula-wise*). La théorie ne peut exprimer cette idée que par la comparaison d'un profil donné avec des profils hypothétiques, ceux où les jugements individuels sur les  $\psi \neq \varphi$  varient alors que les jugements individuels sur  $\varphi$  restent identiques. L'axiome, aussi bien que sa formulation multi-profil, suivent de près l'indépendance des options non pertinentes, qui joue le rôle essentiel qu'on sait dans la théorie arrovienne.

Les règles de vote satisfont une propriété classique de renforcement: lorsqu'un résultat collectif reflète le jugement d'un groupe d'électeurs, il se maintient si ce groupe rallie de nouveaux électeurs au même jugement. Comme la condition de sensibilité positive (*positive responsiveness*) qui lui est apparentée en théorie du choix social, celle-ci appelle une formulation multi-profil.

**Monotonie.** Pour toute formule  $\varphi \in X$  et tout couple de profils  $(A_1, \dots, A_n)$ ,  $(A'_1, \dots, A'_n)$ , si, pour tout  $i = 1, \dots, n$ ,  $\varphi \in A_i \implies \varphi \in A'_i$ , avec un  $j$  au moins tel que  $\varphi \notin A_j$  et  $\varphi \in A'_j$ , alors

$$\varphi \in F(A_1, \dots, A_n) \implies \varphi \in F(A'_1, \dots, A'_n).$$

Nous parlerons d'*indépendance*, ou de *systémativité*, *monotone* pour désigner la conjonction de l'axiome de monotonie avec ceux d'indépendance ou de systémativité. Ils ne l'impliquent pas, comme il est évident pour l'indépendance et comme on peut le vérifier maintenant pour la systémativité. Soit la règle d'*antidictature* ainsi définie: il existe  $j$  tel que pour tout  $(A_1, \dots, A_n)$  et tout  $\varphi \in X$ ,

<sup>15</sup>De ce fait, l'indépendance améliore le statut normatif de la systémativité, mais des objections demeurent cependant (voir Mongin, 2008).

$$\varphi \in A_j \Leftrightarrow \varphi \in F(A_1, \dots, A_n).$$

Sous la restriction d'agenda convenable, l'antidictature est une  $F : D^n \rightarrow D$ .<sup>16</sup> Elle est systématique, mais non monotone, comme on le voit en rapprochant deux profils  $(A_1, \dots, A_n)$  et  $(A'_1, \dots, A'_n)$  tels que  $\neg\varphi \in A_j$  et  $\varphi \in A'_j$ , et  $A_i = A'_i, i \neq j$ .

On approfondirait l'explication du dilemme discursif si l'on parvenait à reproduire la conclusion d'impossibilité du Théorème 1 sous l'indépendance en lieu et place de la systématisme. La théorie ferait alors remonter le problème fondamental du jugement collectif à l'un des deux éléments conceptuels, l'agrégation formule par formule, plutôt qu'à l'autre, la neutralité. Sans doute la systématisme demeurerait-elle dans la conclusion - les fonctions dictatoriales, seules existantes d'après le Théorème 1, vérifient cette propriété - et donc aussi dans les hypothèses, mais il serait préférable de l'expliciter seulement dans la démonstration. Par là, on restaurerait le parallèle avec la théorie du choix social, dont les résultats les plus forts traitent l'indépendance des options non pertinentes comme une hypothèse et la neutralité comme un intermédiaire.<sup>17</sup>

Ce programme a été mené à bien par Pauly et van Hees (2006, Theorem 4) et Dietrich (2006, Theorem 1, Corollary 2), qui mettent l'indépendance au point de départ. Pour démontrer leurs résultats, ces auteurs se servent de la préservation de l'unanimité, mais ils n'ont pas besoin de l'expliciter dans leurs hypothèses, ce qui fait une grande différence avec les démonstrations correspondantes en théorie du choix social. En effet, sous leurs conditions d'agenda respectives, cet axiome découle de l'indépendance. L'économie de moyens est remarquable, mais on peut vouloir faire apparaître au grand jour les principes distincts qu'elle assimile. Mongin (2008, Theorem 2) propose alors d'expliciter la préservation de l'unanimité dans les hypothèses, et pour cela, il affaiblit l'indépendance de manière que celle-ci ne l'implique plus. Sous les restrictions d'agenda qui sont les siennes, ce théorème d'impossibilité restaure le parallélisme avec la théorie du choix social.

Les trois théorèmes ont en commun de renforcer les conditions d'agenda peu exigeantes du Théorème 1, ce qui est la contrepartie des affaiblissements obtenus du côté de la systématisme. Nous dirons qu'un agenda  $X$  est *fermé par variables propositionnelles* si, pour tout  $\varphi \in X$  et tout  $a \in \mathcal{P}$  qui figure dans  $\varphi$ ,  $a \in X$ . Par exemple, si  $\varphi = (d \leftrightarrow v \wedge r) \in X$ , on aura  $v, r, d \in X$ ; ainsi l'agenda  $\overline{X}$  la vérifie. Un *littéral* se définit comme une variable propositionnelle  $a$  ou sa négation  $\neg a$ ; il est noté par  $\pm a$ . Compte tenu de la fermeture de  $X$  par négations, la condition demande en fait que, pour tout  $\varphi \in X$  et tout  $a \in \mathcal{P}$  qui figure dans  $\varphi$ ,  $\pm a \in X$ .

**Théorème 2 (Pauly et van Hees, 2006).** Soit un langage propositionnel classique  $\mathcal{L}_{\mathcal{P}}$  avec  $\mathcal{S} = \{\neg, \wedge\}$ ; soit  $X$  fermé par variables propositionnelles, contenant deux variables propositionnelles distinctes, et tel que, si  $\pm a, \pm b \in X$ ,

<sup>16</sup>Pour tout sous-ensemble cohérent  $B \subseteq X$ , le sous-ensemble nié  $\{\neg\varphi : \varphi \in B\}$  doit être aussi cohérent.

<sup>17</sup>Curieusement, la théorie de l'agrégation logique et celle du choix social ont suivi des progressions opposées. Le théorème d'Arrow en 1951 procède de l'indépendance, et ce n'est qu'ensuite et pour des contextes spéciaux que certains théorèmes partiront de la neutralité. Fleurbaey et Mongin (2005) réexaminent l'enchaînement historique.

$\pm a \wedge \pm b \in X$ ; alors toute  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature et l'indépendance est une fonction constante.<sup>18</sup>

Une fonction de jugement collectif  $F$  est *constante* s'il existe un ensemble de jugement  $A$  tel que, pour tout profil  $(A_1, \dots, A_n)$ ,  $F(A_1, \dots, A_n) = A$ . Ce cas dégénéré survient à propos des fonctions de bien-être social, et certaines axiomatiques du choix social choisissent de l'exclure d'emblée. En théorie de l'agrégation logique, il se présente avec l'affaiblissement de la condition la plus caractéristique: une fonction de jugement collectif dans  $D$  ou  $D^*$  qui est constante n'est pas systématique, mais elle est indépendante.<sup>19</sup>

Nous n'avons fait jusqu'à présent, et ne ferons en général, aucune hypothèse sur la cardinalité  $|\mathcal{P}|$  de l'ensemble des variables propositionnelles. Mais le théorème suivant s'énonce mieux avec  $\mathcal{P}$  fini (et donc, à l'équivalence logique près,  $\mathcal{L}_{\mathcal{P}}$  également fini). On peut alors définir les *atomes* de  $\mathcal{L}_{\mathcal{P}}$ , qui sont les formules du type  $\pm a_1 \wedge \dots \wedge \pm a_k$ , dans lesquelles figurent les  $k$  variables propositionnelles distinctes de  $\mathcal{P}$ . Soit  $\mathcal{AT}$  l'ensemble des atomes de  $\mathcal{L}_{\mathcal{P}}$ ; il constitue la partition logique - c'est-à-dire la classe de formules logiquement exclusives et logiquement exhaustives - la plus fine concevable pour ce langage. La proposition qui correspond sémantiquement à un atome décrit les états du monde possibles avec une précision maximale. Dietrich montre que, si l'agenda contient les atomes, la conclusion de Pauly et van Hees se retrouve à l'identique; il suffit même que l'indépendance porte seulement sur les atomes au lieu de l'agenda entier.

**Théorème 3 (Dietrich, 2006).** Soit un langage propositionnel classique  $\mathcal{L}_{\mathcal{P}}$  avec  $\mathcal{S} = \{\neg, \wedge\}$  et  $\mathcal{P}$  fini contenant deux variables propositionnelles au moins; soit  $\mathcal{AT} \subset X$ ; alors toute  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature et l'indépendance restreinte à  $\mathcal{AT}$  est une fonction constante.

Il ne s'agit en fait que d'un cas particulier du théorème d'origine. Pour illustrer plus complètement celui-ci, considérons l'exemple où  $\mathcal{P} = \{a_1, a_2\}$  et  $X = \{a_1, \neg a_1 \wedge a_2, \neg a_1 \wedge \neg a_2\}^{\pm}$ . Cet agenda ne contient pas tous les atomes de  $\mathcal{L}_{\mathcal{P}}$ , mais satisfait néanmoins une propriété approchante: pour tout  $B \in D$ , il existe dans  $X$  une formule équivalente à la conjonction des éléments de  $B$ . En effet,  $D$  contient trois ensembles de jugement,

$$\{a, \neg(\neg a \wedge b), \neg(\neg a \wedge \neg b)\}, \{\neg a, \neg a \wedge b, \neg(\neg a \wedge \neg b)\}, \{\neg a, \neg(\neg a \wedge b), \neg a \wedge \neg b\},$$

et la conjonction de chacun d'eux est représentée par une formule équivalente:

$$a, \neg a \wedge b, \neg a \wedge \neg b.$$

Parce qu'elles résument les ensembles cohérents et complets de formules qui sont permis par  $X$ , ces trois formules jouent le rôle d'*atomes relatifs de X*. Le résultat de Dietrich porte sur cette notion plus générale que celle d'atome, ce qui en accroît la portée relativement au Théorème 3.

<sup>18</sup>Tout en étant classique au sens de la section 3, le calcul propositionnel de Pauly et van Hees autorise des valeurs de vérité en nombre fini quelconque. Van Hees (2007) et Duddy et Piggins (2009) se placent aussi en dehors de la sémantique bivalente ordinaire, à laquelle notre formalisme syntaxique se conforme implicitement.

<sup>19</sup>Sans la restriction à  $D$  ou  $D^*$ , il existerait une fonction de jugement social constante et systématique: pour tout  $(A_1, \dots, A_n)$ ,  $F(A_1, \dots, A_n) = X$ .

Dans cet énoncé, l'indépendance porte sur un sous-ensemble particulier de l'agenda. Le résultat qui suit restreint de même l'axiome, mais à un sous-ensemble différent, celui des variables propositionnelles qui apparaissent dans les formules de  $X$ .

**Théorème 4 (Mongin, 2008).** Soit un langage propositionnel classique  $\mathcal{L}_{\mathcal{P}}$  quelconque; soit  $X$  fermé par variables propositionnelles et contenant deux variables propositionnelles au moins; de plus,  $X$  satisfait les conditions de la section 7. Alors, il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature, la préservation de l'unanimité et l'indépendance restreinte à l'ensemble  $VP_X$  des variables propositionnelles de  $X$ .

Les trois théorèmes de cette section concordent pour désigner l'indépendance comme la source principale du dilemme discursif. Les deux premiers disent en substance qu'une fonction de jugement collectif dégénère en dictature ou en fonction constante si elle procède formule par formule *sur un agenda dont les formules entretiennent des relations logiques déterminées*. Le dernier théorème ajoute une considération nouvelle à ce diagnostic: l'indépendance devrait porter sur les seules formules qui, dans un calcul propositionnel classique, n'entretiennent pas de relations logiques préexistantes, c'est-à-dire sur  $VP_X$ . La conclusion d'impossibilité demande alors qu'on ajoute la préservation de l'unanimité, qui devient la cible des objections normatives. Mongin rattache le dilemme discursif au *problème de l'unanimité factice* (*spurious unanimity*) qu'il a mis en évidence dans le cadre du bayésianisme collectif.<sup>20</sup>

L'exemple de base met en scène les deux interprétations données tour à tour. Suivant la première, le tribunal s'expose à des difficultés parce qu'il fait voter les juges sur chaque proposition considérée isolément, alors qu'elles sont logiquement reliées par la doctrine juridique. Suivant la seconde, même si le tribunal fait voter les juges sur des propositions indépendantes, il doit encore veiller à soustraire la doctrine juridique à la préservation de l'unanimité; en l'occurrence, les juges ne sont pas d'accord sur la manière de l'appliquer, ce qui prive cet axiome de sa force normative supposée.

Un autre exemple, formel celui-là, précisera les deux raisonnements. Soit un langage propositionnel  $\mathcal{L}_{\mathcal{P}}$  avec  $\mathcal{S} = \{\neg, \vee\}$  et  $\mathcal{P} = \{a, b, c\}$ ; soit l'agenda  $X = \{a, b, c, a \vee b \vee c\}^{\pm}$ , qui entre dans les conditions du Théorème 4; soit enfin  $n = 3$  et le profil  $(A_1, A_2, A_3) \in D^3$  ainsi donné:

$$a, \neg b, \neg c \in A_1; \neg a, b, \neg c \in A_2, j \in J; \neg a, \neg b, c \in A_3.$$

Par clôture déductive,  $a \vee b \vee c \in A_i, i = 1, 2, 3$ . Si l'on applique  $F_{maj}$ , l'ensemble de jugement collectif  $A$  est contradictoire, puisque  $\neg a, \neg b, \neg c, a \vee b \vee c \in A$ . Cette observation illustre les difficultés de l'agrégation formule par formule, compte tenu de la liaison logique entre  $a, b, c$  instaurée par  $a \vee b \vee c$ , et elle reflète l'esprit des Théorèmes 2 et 3 (l'esprit mais non la lettre, car le  $X$  choisi échappe à leurs conditions d'agenda). Mais loin d'être limitée à  $F_{maj}$ , la contradiction persiste avec toute  $F$  qui coïncide avec elle sur  $a, b, c$  et se contente d'enregistrer

<sup>20</sup>Des individus peuvent faire les mêmes comparaisons d'utilité espérée alors qu'ils s'opposent à la fois sur les valeurs d'utilité et de probabilité. Mongin (1995) explique par là l'impossibilité du bayésianisme collectif. Généralisant le propos, Mongin (1997) parle d'*unanimité factice* lorsque l'accord sur un jugement collectif s'accompagne de désaccords sur les raisons de le porter. Nehring (2005) retrouve l'idée dans son formalisme abstrait d'agrégation parétienne.

l'unanimité sur  $a \vee b \vee c$ . Le Théorème 4 laissait attendre cette amplification, et l'exemple en fait comprend le lien avec l'unanimité factice: les trois individus ont des raisons incompatibles d'accepter la même formule  $a \vee b \vee c$ .

Un théorème de Nehring et Puppe (2002), fondé sur l'indépendance monotone, précède les Théorèmes 2-4. Il ne se place pas encore dans le cadre formel de la théorie de l'agrégation logique, mais il est possible d'en faire la traduction rétrospective, ainsi que les deux auteurs l'ont montré (voir Nehring et Puppe, 2010). Ce résultat se rattache en fait à la section 6, et nous tenterons d'illustrer autrement la force de la condition d'indépendance monotone. L'énoncé qui suit s'exprime à l'aide d'une notion affaiblie de dictature. Une fonction de jugement collectif  $F$  comporte un *pouvoir de veto local* s'il existe un individu  $j$  et une formule  $\varphi \in X$  tels que, pour tout profil  $(A_1, \dots, A_n)$ ,

$$\varphi \notin A_j \implies \varphi \notin F(A_1, \dots, A_n).$$

Pour une  $F$  donnée, il peut exister de multiples détenteurs de veto  $j$ , chacun étant relatif à une  $\varphi$  particulière. En rapport avec les préférences, la théorie du choix social définit des notions de veto dont la plus faible - pour deux options données - s'apparente à celle-ci (voir Gibbard, 1969).

**Absence de veto.**  $F$  ne comporte pas de pouvoir de veto local.

**Théorème 5** (Nehring et Puppe, 2008). Soit un langage propositionnel classique  $\mathcal{L}_{\mathcal{P}}$  quelconque; soit  $X$  fermé par variables propositionnelles et contenant au moins une formule contingente qui n'équivaut pas logiquement à un littéral; alors, il n'existe pas de  $F$  surjective, satisfaisant l'indépendance monotone et l'absence de veto.

Par rapport aux Théorèmes 2, 3 et 4, les axiomes sur  $F$  se renforcent - l'indépendance s'est accrue de la monotonie, qui, en présence de la surjectivité, implique la préservation de l'unanimité, et l'absence de veto est nettement plus exigeante que celle de la dictature -, mais les contraintes sur  $X$  s'allègent dans le même temps. Il se confirme ainsi que de nombreux arbitrages sont possibles entre conditions d'agenda et axiomes; la section 6 développera cette observation au point de préciser les équivalences métathéoriques.

Le Théorème 5 affaiblit deux résultats de Nehring et Puppe (2008, Theorems 1 et 2) qui reposent sur une notion particulière aux deux auteurs. Ils définissent  $F$  comme une *oligarchie avec solution par défaut* (*oligarchy with default*) s'il existe deux sous-ensembles non vides  $J \subset X$  et  $M \subset \{1, \dots, n\}$  tels que pour tout  $(A_1, \dots, A_n)$  et tout  $\varphi \in X$ ,

$$\begin{aligned} \varphi \in F(A_1, \dots, A_n) \text{ ssi ou bien } \varphi \in A_j \text{ pour tout } j \in M, \\ \text{ou bien } \varphi \in J \text{ et } \varphi \in A_j \text{ pour un } j \in M. \end{aligned}$$

Les  $j \in M$  sont alors désignés comme *les oligarques*, et l'ensemble  $J$  comme la *solution par défaut* (*default*). En substance, si les oligarques s'accordent sur la formule  $\varphi$ , elle s'impose au jugement collectif, et si elle les divise, on décide entre  $\varphi$  et  $\neg\varphi$  d'après ce qu'indique la solution par défaut. Pour des agendas dont nous ne reproduisons pas les conditions, Nehring et Puppe montrent que les seules  $F$  satisfaisant l'indépendance monotone et la surjectivité sont les oligarchies avec solution par défaut.<sup>21</sup>

---

<sup>21</sup>Nehring et Puppe (2010) abandonnent la condition - dite de *vérifonctionnalité* (*truth-functionality*) - qui correspond ici à la fermeture de l'agenda par variables propositionnelles. Il existe alors des  $F$  qui échappent à la dégénérescence oligarchique.

En résumé, avec les Théorèmes 2-5, la théorie l'agrégation logique approfondit encore le dilemme discursif. La systématisme a fait place à l'indépendance, tantôt posée isolément, chez Pauly et van Hees ou Dietrich, tantôt renforcée de la préservation de l'unanimité, chez Mongin, tantôt renforcée de la monotonie, chez Nehring et Puppe. La théorie ultérieure a privilégié l'axiomatique intermédiaire. Avant d'introduire son théorème d'impossibilité canonique, nous reviendrons sur le cadre formel de la théorie pour en faire saisir mieux l'extrême généralité.

## 5 Une logique générale pour la théorie

Les théorèmes des deux sections précédentes sont formulés à propos des calculs propositionnels classiques, ce qui est restrictif. La question se pose de les étendre à des calculs propositionnels *non classiques*, c'est-à-dire comportant des connecteurs non booléens et des relations d'inférence plus fortes que celles des calculs précédents. De même, la question se pose de les étendre aux *calculs des prédicats*, qui s'en tiennent aux connecteurs booléens, mais perfectionnent l'analyse des propositions à l'aide de symboles de prédicats et de quantificateurs. Ceux-là renforcent dans une autre direction l'inférence des calculs précédents; ils sont les calculs immédiatement pertinents pour obtenir les théorèmes du choix social à partir de ceux de l'agrégation logique. Mais au lieu de démontrer un résultat de la théorie en deux temps, d'abord pour des calculs élémentaires, puis en la vérifiant sur d'autres plus avancés, il vaut mieux le démontrer en une seule fois *dans une logique générale qui englobe tous les calculs auxquels on s'intéresse*. Cette exigence métathéorique remonte à Dietrich (2007a), qui l'a réalisée en axiomatisant la relation d'inférence  $\vdash$  sans faire référence à une logique particulière. Nous la reprendrons dans l'axiomatique perfectionnée de Dietrich et Mongin (2010). Dès lors qu'un théorème est démontré avec la logique générale, il suffira, pour qu'il s'applique à un calcul dont le langage est du type  $\mathcal{L}$  défini à la section 4, que sa relation d'inférence obéisse à l'axiomatique en question. Compte tenu de l'importance reconnue au théorème canonique de la section 6, il importera de l'énoncer dans ce nouveau cadre formel.

La logique générale comporte six axiomes qui déterminent *la relation d'inférence*  $S \vdash \psi$  entre un ensemble de prémisses  $S \subset \mathcal{L}$  et une conclusion  $\psi \in \mathcal{L}$ . Les écritures  $S \not\vdash \psi$  et  $\varphi \vdash \psi$ , signifient, respectivement, que  $S \vdash \psi$  n'a pas lieu et que  $S \vdash \psi$  a lieu avec  $S = \{\varphi\}$ .

(E1) Il n'existe pas de  $\psi \in \mathcal{L}$  tel que  $\emptyset \vdash \psi$  et  $\emptyset \vdash \neg\psi$  (non-trivialité).

(E2) Pour tout  $\varphi \in \mathcal{L}$ ,  $\varphi \vdash \varphi$  (réflexivité).

(E3) Pour tout  $S \subset \mathcal{L}$  et tout  $\varphi, \psi \in \mathcal{L}$ , si  $S \cup \{\varphi\} \vdash \psi$  ou  $S \cup \{\neg\varphi\} \vdash \psi$ ,  $S \vdash \psi$  (complétion à une étape).

(E4) Pour tout  $S \subset S' \subset \mathcal{L}$  et tout  $\psi \in \mathcal{L}$ , si  $S \vdash \psi$ , alors  $S' \vdash \psi$  (monotonie).

(E5) Pour tout  $S \subset \mathcal{L}$  et tout  $\psi \in \mathcal{L}$ , si  $S \vdash \psi$ , il existe un sous-ensemble fini  $S_0 \subset S$  tel que  $S_0 \vdash \psi$  (compacité).

(E6) Pour tout  $S \subset \mathcal{L}$ , s'il existe  $\psi \in \mathcal{L}$  tel que  $S \vdash \psi$  et  $S \vdash \neg\psi$ , alors pour tout  $\psi \in \mathcal{L}$ ,  $S \vdash \psi$  (non-paraconsistance).

Une nouvelle propriété découle des précédentes:

(E7) Pour tous  $S, T \subset \mathcal{L}$  et tout  $\psi \in \mathcal{L}$ , si  $T \vdash \psi$  et  $S \vdash \varphi$  pour tout  $\varphi \in T$ , alors  $S \vdash \psi$  (transitivité).

De toute cette liste, (E4) est sans doute la condition la plus importante. Elle exprime la monotonie caractéristique de l'inférence *déductive* par opposition à la non-monotonie caractéristique des inférences *inductives*, comme dans l'exemple

suivant:  $S$  dit que tous les corbeaux examinés jusqu'à l'étape  $t$  sont noirs et  $\psi$  que tous les corbeaux sont noirs; l'inférence inductive de  $S$  vers  $\psi$  n'a plus lieu si  $S$  est accru d'un  $\varphi$  disant que le corbeau examiné à l'étape  $t + 1$  n'est pas noir. Or le concept ordinaire ou philosophique du jugement s'articule sur celui du raisonnement pris en général, et non pas seulement de la déduction. Il faut donc voir (E4) comme *une restriction apportée aux jugements dont s'occupe la théorie*; c'est une raison supplémentaire pour la désigner comme celle de l'agrégation logique.

La condition (E1) est essentielle, non seulement parce qu'elle définit une inférence qui est non-triviale, mais parce que, si elle était violée, (E4) ferait que n'importe quelle inférence est triviale. (E2) énonce une propriété attendue de l'inférence, déductive ou inductive. (E3) autorise l'abréviation des prémisses, ce qui est convenable pour l'inférence déductive: si l'on infère  $\psi$  de  $\varphi$  aussi bien que de  $\neg\varphi$ , on peut inférer  $\psi$  en se passant de l'une et de l'autre formule. D'ordinaire, (E3) se dissimule sous d'autres conditions plus courantes.<sup>22</sup> (E5) permet de supposer que l'ensemble de prémisses est fini, ce que les logiciens justifient d'habitude par leur préoccupations constructives; le théorème canonique aura besoin de cet axiome. (E6) vise à exclure un groupe de calculs déductifs qui les embarrassent depuis longtemps, celui des *logiques paraconsistantes*. Leur singularité se comprendra mieux une fois définies les notions de cohérence et d'incohérence.

Soit  $\mathcal{I}$  l'ensemble des ensembles incohérents de formules de  $\mathcal{L}$ ; par définition, les ensembles *cohérents* sont ceux qui n'appartiennent pas à  $\mathcal{I}$ . D'après la définition la plus courante en logique,

(Déf\*)  $S \in \mathcal{I}$  si et seulement si pour tout  $\psi \in \mathcal{L}$ ,  $S \vdash \psi$ .

On remarquera que notre définition de la cohérence, à la section 3, se conformait à (Déf\*). Les logiciens de la paraconsistance choisissent une définition plus faible:

(Déf\*\*)  $S \in \mathcal{I}$  si et seulement s'il existe  $\psi \in \mathcal{L}$  tel que  $S \vdash \psi$  et  $S \vdash \neg\psi$ .

Chaque définition peut contester l'autre au motif qu'elle donne lieu à trop - pour Déf\*\* - ou trop peu - pour Déf\* - d'ensembles incohérents. Les mathématiciens ont implicitement écarté ce débat, et (E6), qui rend les définitions identiques, reflète leur position. En reprenant cet axiome, la théorie de l'agrégation logique se conforme aux intuitions démonstratives ordinaires et elle ne laisse de côté qu'une famille assez peu répandue de calculs logiques.<sup>23</sup>

Sous l'une ou l'autre définition, l'axiomatique (E1)-(E6) implique les propriétés suivantes de  $\mathcal{I}$ :

(I1)  $\emptyset \notin \mathcal{I}$  (non-trivialité).

(I2) Pour tout  $\varphi \in \mathcal{L}$ ,  $\{\varphi, \neg\varphi\} \in \mathcal{I}$  (réflexivité).

(I3) Pour tout  $S \subset \mathcal{L}$  et tout  $\varphi \in \mathcal{L}$ , si  $S \notin \mathcal{I}$ , ou bien  $S \cup \{\varphi\} \notin \mathcal{I}$ , ou bien  $S \cup \{\neg\varphi\} \notin \mathcal{I}$  (complétion à une étape).

(I4) Pour tout  $S \subset S' \subset \mathcal{L}$ , si  $S \in \mathcal{I}$ , alors  $S' \in \mathcal{I}$  (monotonie).

(I5) Pour tout  $S \subset \mathcal{L}$ , si  $S \in \mathcal{I}$ , il existe un sous-ensemble fini  $S_0 \subset S$  tel que  $S_0 \in \mathcal{I}$  (compacité).

La monotonie se retrouve en (I4). Comme son homologue, (I1) évite de trivialisier la notion qu'on veut définir. (I2) s'impose une fois qu'on a écarté l'option

<sup>22</sup>Dietrich (2008) posait directement que chaque ensemble cohérent de formules s'étend à un ensemble maximal cohérent, ce qui correspond au *lemme de Lindenbaum* bien connu en logique. Cette propriété importante est ici déduite à partir de (E3).

<sup>23</sup>Price (2002) expose les calculs paraconsistants et leurs motivations.

paraconsistante. (I3) justifie rétrospectivement l'appellation donnée à son homologue, puisqu'elle autorise à compléter un ensemble cohérent par une formule ou sa négation. En présence de la compacité, ici (I5), la propriété d'extension cohérente finie implique la propriété infinie correspondante:

(I3<sup>+</sup>) Pour tout  $S \subset \mathcal{L}$ , si  $S \notin \mathcal{I}$ , il existe  $T \subset \mathcal{L}$  tel que (i)  $S \subset T$ , (ii)  $T \notin \mathcal{I}$ , et (iii), pour tout  $\varphi, \neg\varphi \in \mathcal{L}$ , soit  $\varphi \in T$ , soit  $\neg\varphi \in T$  (complétion entière). La définition de  $D$  à la section 3 est rétrospectivement justifiée: il existe des ensembles de jugement cohérents et complets quelle que soit la cardinalité de  $\mathcal{L}$ ; le cas infini posait un problème désormais résolu.

La logique générale peut s'exposer dans l'ordre inverse, c'est-à-dire en partant de l'ensemble  $\mathcal{I}$  axiomatisé par (I1)-(I5), et en traitant la relation  $\vdash$  comme dérivée. On vérifie alors qu'elle obéit à (E1)-(E6), qui perdent leur statut d'axiomes pour devenir des propriétés. Une définition nouvelle raccorde les registres si l'on prend cet ordre de parcours:

(Déf<sup>\*\*\*</sup>)  $S \vdash \varphi$  si et seulement si  $S \cup \{\neg\varphi\} \in \mathcal{I}$ .

La réduction de l'inférence à l'incohérence dans (Déf<sup>\*\*\*</sup>) n'est pas moins commune que la réduction inverse dans (Déf<sup>\*</sup>). Que la logique générale ait deux axiomatisations, cela en facilite beaucoup l'emploi par la théorie agrégative. En fait, comme les axiomes ou propriétés choisis reflètent la pratique immédiate de la déduction, sauf pour la compacité, qui est facultative, les démonstrations n'appellent de précautions particulières que lorsque cette dernière propriété entre en jeu; il en va ainsi dans la démonstration du théorème canonique.

Suivant l'une ou l'autre liste prise comme critérium, les calculs propositionnels classiques tombent sous la logique générale. Les non-classiques, en revanche, doivent être examinés un à un: ceux qui relèvent de la logique déductive, par opposition à la logique inductive ou non-monotone, sont nombreux à remplir les exigences, mais la compacité manque à certains, en particulier dans la classe importante des logiques épistémiques.<sup>24</sup>

Les calculs des prédicats entrent dans la logique générale une fois qu'on met en relation leurs langages logiques avec ceux de calculs propositionnels introduits pour la circonstance. Dans leur cas, il faut en effet tenir compte de la spécificité morphologique des formules, construites à partir de symboles de prédicats et de quantificateurs, donc à première vue étrangères au cadre formel de la théorie. La méthode consiste à identifier un sous-ensemble pertinent de formules avec l'ensemble des variables propositionnelles  $\mathcal{P}$  d'un langage  $\mathcal{L}$  au sens de la section 3; on vérifie qu'une fois restreinte à  $\mathcal{L}$ , la relation d'inférence de ces calculs satisfait (E1)-(E6).<sup>25</sup>

Outre les notions d'inférence et d'incohérence, que nous emploierons simultanément par commodité, deux notions dérivées vont intervenir dans l'énoncé du théorème canonique. En premier lieu, un ensemble de formules  $S \subset \mathcal{L}$  est dit *minimalement incohérent* s'il est incohérent et que tous ses sous-ensembles stricts soient cohérents; avec un calcul propositionnel classique, c'est le cas de

$$\{v, r, d \longleftrightarrow v \wedge r, \neg d\},$$

<sup>24</sup>C'est le cas des logiques épistémiques probabilistes (voir Heifetz et Mongin, 2001) et de certaines logiques de la connaissance commune (mais non pas de toutes, voir Lismont et Mongin, 2002).

<sup>25</sup>Techniquement parlant, on exploite l'homomorphisme entre le calcul des prédicats et son calcul propositionnel associé; voir par exemple Barwise (1977).

mais non pas de

$$\{\neg v, \neg r, d \longleftrightarrow v \wedge r, d\}.$$

En second lieu, on dit, pour  $\varphi, \psi \in X$ , que  $\psi$  s'infère conditionnellement de  $\varphi$  - ce qui se note  $\varphi \vdash^* \psi$  - s'il existe un ensemble de prémisses auxiliaires  $Y \subset X$  tel que (i)  $Y \cup \{\varphi\} \vdash \psi$  et (ii)  $Y \cup \{\varphi\}$  et  $Y \cup \{\neg\psi\}$  soient cohérents ( $Y = \emptyset$  est autorisé). L'inférence conditionnelle peut se reformuler comme une propriété d'ensembles minimalement incohérents, et c'est d'abord ainsi qu'elle a été comprise. En effet, sous la compacité (I5), l'inférence conditionnelle  $\varphi \vdash^* \psi$  équivaut à la propriété que  $\varphi \neq \neg\psi$  et qu'il existe un  $Y' \subset X$  minimalement incohérent avec  $\varphi, \neg\psi \in Y'$ .

Il est utile d'enregistrer deux propriétés élémentaires de la relation  $\vdash^*$ . D'une part, elle vérifie la contraposition, c'est-à-dire que  $\varphi \vdash^* \psi \iff \neg\psi \vdash^* \neg\varphi$ ; d'autre part,  $\varphi \not\vdash^* \neg\varphi$ . Pour clore les préparatifs, nous fixons un raccourci de notation. Il nous arrivera de transformer un ensemble  $Y \subset X$  en un autre  $Y'$  qui remplace chaque formule d'un  $Z \subset Y$  par sa négation; l'ensemble transformé

$$Y' = (Y \setminus Z) \cup \{\neg\varphi : \varphi \in Z\}$$

s'écrira  $Y_{-Z}$ .

## 6 Le théorème d'impossibilité canonique

Outre le fait qu'ils restreignent la logique choisie, les Théorèmes 1 à 5 souffrent d'une certaine imprécision. Tels qu'ils sont formulés, ils se contentent d'énoncer des conditions suffisantes sur l'agenda pour qu'il n'existe pas de fonction de jugement collectif - au moins non-dégénérée - qui vérifie une ou plusieurs propriétés axiomatiques choisies. Ces hypothèses peuvent être trop fortes pour la conclusion obtenue, et si elles ne le sont pas, une démonstration supplémentaire de nécessité devrait l'établir. Sous l'influence de Nehring et Puppe (2002, 2010) et de Dokow et Holzman (2009, 2010a et b), la théorie de l'agrégation logique s'est imposé l'exigence de *caractériser*, au sens des conditions nécessaires et suffisantes, les agendas qui transforment une liste donnée d'axiomes en impossibilité d'un certain type. Si les résultats de ces auteurs méritent, selon nous, d'être appelés *canoniques*, ce n'est pas tant pour leur généralité supérieure, car ils sont loin d'unifier la théorie entière, que parce qu'ils ont établi un standard de la démonstration qui est devenu maintenant le partage de tous les autres. Nous suivrons l'analyse de Dokow et Holzman, parce qu'elle est la plus explicite des deux et qu'elle unifie un plus grand nombre de travaux existants. Nous ne la reproduirons pas telle quelle, mais plongée dans la logique générale, qui n'est pas le formalisme voulu par Dokow et Holzman; la différence, qui n'est pas négligeable, sera expliquée à la fin de la section.

Dokow et Holzman ont posé et résolu le problème agrégatif suivant: comment caractériser - au sens mathématique de conditions nécessaires et suffisantes - les agendas  $X$  tels que, si l'on définit  $D$  par rapport à  $X$ , il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie simultanément la non-dictature, l'indépendance et la préservation de l'unanimité? La réponse à ce problème est le théorème canonique annoncé. Elle met en relief les conditions d'agenda suivantes:

(a) *Il existe un ensemble minimal incohérent de formules  $Y \subset X$  et un choix de  $Z \subset Y$  tels que  $|Z| = 2$  et  $Y_{-Z}$  soit cohérent.*<sup>26</sup>

<sup>26</sup>La condition est équivalente si, plus généralement,  $Z$  est de cardinalité paire (voir Dokow et Holzman, 2010, et Dietrich et Mongin, 2010).

(b) Pour tout couple de formules  $\varphi, \psi \in X$ , il existe des formules  $\varphi_1, \dots, \varphi_k \in X$  telles que

$$\varphi = \varphi_1 \vdash^* \varphi_2 \vdash^* \dots \vdash^* \varphi_k = \psi.$$

**Théorème 6 (Dokow et Holzman, 2010a; pour la suffisance, voir aussi Dietrich et List, 2007b).** Sous les conditions (a) et (b), il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature, la préservation de l'unanimité et l'indépendance. Pour un nombre d'individus  $n \geq 3$ , (a) et (b) sont aussi nécessaires à cette conclusion.

Comme après le Théorème 1, nous réexaminerons en détail l'exemple de base. Désignons par (Inc) la propriété, manifestée par le dilemme discursif, qu'un  $(A_1, \dots, A_n)$  au moins rend  $F_{maj}(A_1, \dots, A_n)$  incohérent; elle équivaut à ce que  $F$  ne soit pas  $D^n \rightarrow D$  dans le cas où  $n$  est impair, le seul que nous considérerons. Etant donné que  $F_{maj}$  satisfait les axiomes, le Théorème 6 donne l'implication (a),(b)  $\Rightarrow$  (Inc). Vérifions-la sur l'agenda

$$\overline{X} = \{v, r, d, d \longleftrightarrow v \wedge r\}^\pm.$$

On s'assure de (a) en prenant

$$\begin{aligned} Y &= \{\neg v, d, d \longleftrightarrow v \wedge r\} \text{ et } Z = \{\neg v, d\}, \text{ ou} \\ Y &= \{v, r, d, \neg(d \longleftrightarrow v \wedge r)\} \text{ et } Z = \{v, r\}, \end{aligned}$$

ou d'autres choix encore, ce qui donne à penser que cette condition est facile à remplir malgré son écriture complexe. Quant à (b), elle est satisfaite aussi, comme on le voit sur la Figure 1, dans laquelle nous notons  $q = d \longleftrightarrow v \wedge r$  (les flèches indiquent les inférences conditionnelles et les formules en petits caractères sont des prémisses auxiliaires convenables pour ces inférences).

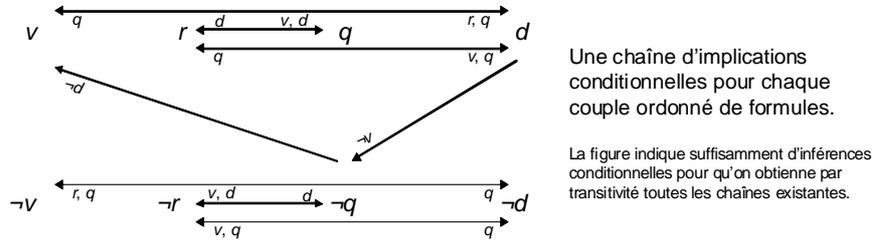


Figure 1: L'agenda  $\overline{X}$  vérifie (b).

Nous venons d'illustrer (a) et (b) dans leur rôle de conditions *suffisantes* de l'impossibilité; il reste à faire comprendre en quoi elles sont *nécessaires*. On sait par la section 2 que le choix de  $\overline{X}$  accompagne l'abandon du paradoxe doctrinal au profit du dilemme discursif et que l'on peut se refuser à cet abandon. La section 7 remplacera la biimplication classique par une biimplication non-classique dans la formule de la doctrine juridique, ce qui fera échouer la condition (b). A moindres frais, celle-ci peut être mise en défaut par *une réinterprétation de la doctrine qui rejette la formule  $d \longleftrightarrow v \wedge r$  de l'agenda vers la relation d'inférence*. On formalise ainsi le cas, envisagé par Kornhauser et Sager, où le tribunal détermine sa décision à partir d'une doctrine commune qu'il n'a pas mise au vote. Soit, donc, la relation d'inférence  $\vdash_{d \longleftrightarrow v \wedge r}$  définie par:

$$S \vdash_{d \longleftrightarrow v \wedge r} \varphi \text{ ssi } S \cup \{d \longleftrightarrow v \wedge r\} \vdash \varphi,$$

et l'agenda concordant:

$$\overline{\overline{X}} = \{v, r, d\}^\pm.$$

En étudiant les inférences conditionnelles engendrées par  $\vdash_{d \leftrightarrow v \wedge r}$ , on trouve que  $\overline{\overline{X}}$  vérifie (a), mais non pas (b), ainsi que le fait voir le graphe suivant:

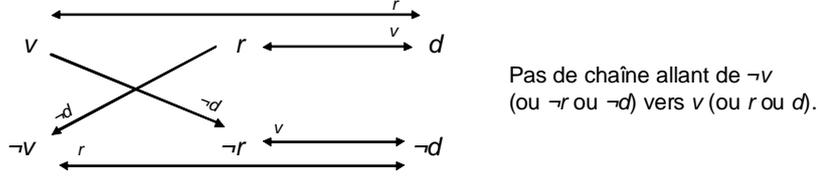


Figure 2: L'agenda  $\overline{\overline{X}}$  viole (b).

D'après le Théorème 6, quand l'agenda est  $\overline{\overline{X}}$ , il existe une  $F : D^n \rightarrow D$  non-dictatoriale, indépendante et qui préserve l'unanimité. En voici un exemple, la fonction  $\tilde{F}_{maj}^n$ , ainsi définie sur  $D^n$ : pour toute formule *positive*  $\varphi \in X$ , elle respecte l'unanimité en faveur de  $\varphi$  ou de  $\neg\varphi$ , et elle tranche les divisions en sélectionnant toujours  $\neg\varphi$ . La fonction va bien dans  $D$ : d'une part, elle modifie  $F_{maj}^n$ , définie à la section 3, pour que les ensembles de jugement collectif soient *complets*; d'autre part, avec  $\overline{\overline{X}}$  donné, ces ensembles sont aussi *cohérents*.<sup>27</sup>

Les conditions nécessaires et suffisantes d'impossibilité qu'énonce le Théorème 6 se simplifient souvent lorsqu'on s'intéresse à une fonction de jugement collectif en particulier. Soit encore  $F_{maj}$ ; nous savons que (a),(b)  $\Rightarrow$  (Inc), mais ne peut-on améliorer l'implication dans le cas d'espèce? De fait, la condition qui suit, plus faible que (b), est à elle seule nécessaire et suffisante à (Inc):

(c) *Il existe un ensemble minimal incohérent de formules  $Y \subset X$  tel que  $|Y| \geq 3$ .*

Il est facile de prouver la suffisance en construisant un profil  $(A_1, \dots, A_n)$  tel que  $Y \subset F_{maj}(A_1, \dots, A_n)$ . Quant à la nécessité, elle découle du Théorème 6 une fois montré que (b) implique (c). Cette dernière observation découle à son tour facilement. D'après (b), il existe pour tout  $\varphi \in X$  une chaîne d'inférences conditionnelles

$$\varphi = \varphi_1 \vdash^* \varphi_2 \vdash^* \dots \vdash^* \varphi_k = \neg\varphi.$$

En l'absence de (c), elle serait faite seulement d'inférences logiques:

$$\varphi = \varphi_1 \vdash \varphi_2 \vdash \dots \vdash \varphi_k = \neg\varphi,$$

ce qui est impossible en logique générale, compte tenu de la transitivité de l'inférence.

La disparition de (a) et l'amenuisement de (b) en (c) que l'on constate sur  $F_{maj}$  pourraient avoir partie liée avec deux propriétés saillantes de cette fonction que le Théorème 6 ne prend pas en compte, la monotonie et la systémativité. Le résultat qui suit, dans sa partie (iii), cautionne formellement la conjecture.

**Théorème 6'.** (i) Sous les conditions (a) et (c), il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature, la préservation de l'unanimité

<sup>27</sup>Pour qu'ils ne le soient pas, il faudrait qu'ils contiennent soit  $\{v, r, \neg d\}$ , soit  $\{-v, d\}$ , soit  $\{-r, d\}$ ; or les trois cas sont exclus par définition de  $\tilde{F}_{maj}^n$ .

et la systématique. Pour un nombre d'individus  $n \geq 3$ , (a) et (c) sont aussi nécessaires à cette conclusion.

(ii) Sous la condition (b), il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature, la préservation de l'unanimité et l'indépendance monotone, et (b) est aussi nécessaire à cette conclusion.

(iii) Sous la condition (c), il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois la non-dictature, la préservation de l'unanimité, la systématique et la monotonie. Pour un nombre d'individus  $n \geq 3$ , (c) est aussi nécessaire à cette conclusion.

Chacun des trois énoncés a fait l'objet de démonstrations indépendantes, et en particulier, (ii) est la version du théorème canonique établie par Nehring et Puppe. Aujourd'hui, il vaut mieux appréhender (i), (ii) et (iii) en bloc, comme autant de résultats partiels tendant vers le Théorème 6. La comparaison permet de situer chaque fois la contrainte sur  $X$  qui équivaut à un axiome sur  $F$ ; ainsi, l'arbitrage si caractéristique de la théorie nouvelle trouve son expression la plus rigoureuse.

Tous les résultats qui précèdent admettent des variantes fondées sur  $D^*$  au lieu de  $D$ . Elles supposent d'introduire un nouvel affaiblissement de la dictature, qui est l'oligarchie, comprise dans un sens qui affaiblit également celui de la section 4. On dira que  $F$  est une *oligarchie* s'il existe un sous-ensemble non vide  $M \subset \{1, \dots, n\}$  tel que, pour tout  $(A_1, \dots, A_n)$ ,

$$F(A_1, \dots, A_n) = \bigcap_{j \in M} A_j,$$

et pour  $F$  donnée, on nomme  $M$ , qui est nécessairement unique, *l'ensemble des oligarques*. D'après une observation de logique élémentaire que récupère la logique générale, l'intersection d'une famille d'ensembles cohérents et déductivement clos préserve ces deux propriétés; il en résulte que, si  $F$  est définie sur  $D^n$  ou  $(D^*)^n$  et qu'elle soit une oligarchie, elle va dans  $D^*$ . Comme rien n'est stipulé pour trancher les désaccords entre les oligarques,  $F$  produira souvent des ensembles de jugement collectifs qui ne sont pas complets. Le phénomène s'illustre avec la règle à quota  $F_{maj}^n$ , qui correspond au groupe maximal d'oligarques  $M = N$ .

**Non-oligarchie.**  $F$  n'est pas une oligarchie.

**Théorèmes 7 et 7'.** Les énoncés reproduisent ceux des Théorèmes 6 et 6' en remplaçant  $F : D^n \rightarrow D$  par  $F : D^n \rightarrow D^*$  ou par  $F : (D^*)^n \rightarrow D^*$ , et la condition de non-dictature par celle de non-oligarchie.

Ces différentes extensions peuvent se lire chez Dietrich et List (2008) et Dokow et Holzman (2010b).<sup>28</sup> L'ensemble de la section doit être comparé aux résultats de dictature ou d'oligarchie en théorie du choix social. Sommairement parlant, chaque théorème d'agrégation logique induit un théorème de choix social par la transformation suivante: *on spécifie le langage logique  $\mathcal{L}$  de manière que les ensembles de jugement expriment des relations de préférence, et l'on spécifie la relation d'inférence  $\vdash$  de manière que les ensembles de jugement respectent les propriétés choisies pour ces relations*. C'est ainsi que Dietrich et List (2007b) dérivent du Théorème 6 une première version du théorème d'Arrow,

<sup>28</sup>Un résultat oligarchique de Gärdenfors (2006) précède le Théorème 7, mais il n'impose que des conditions suffisantes, et d'ailleurs inutilement fortes, à l'agenda d'impossibilité.

dans laquelle ne figurent que des préférences strictes. Ils prennent comme langage logique  $\mathcal{L}_{\succ}$  un fragment de calcul des prédicats construit sur des formules élémentaires  $x \succ y$ , interprétables comme "x est strictement préféré à y", et ils incorporent dans une relation d'inférence  $\vdash_{\succ}$  les trois propriétés d'asymétrie, de transitivité et de totalité retenues pour leur variante arrowienne. Une fois défini l'agenda  $X \subset \mathcal{L}_{\succ}$ , il ne reste qu'à s'assurer des conditions (a) et (b), puis de la correspondance terme à terme entre les deux classes d'axiomes, ceux de la fonction de jugement collectif, d'une part, et ceux de la fonction de bien-être social, de l'autre.

La dérivation n'a pas toujours cette clarté didactique. Pour obtenir le théorème d'Arrow dans son intégralité, c'est-à-dire avec des préférences faibles, Dokow et Holzman (2010b) font un détour par le Théorème 7, qui ne semblait pas destiné à ce but.<sup>29</sup> Le même théorème, en revanche, implique assez directement celui de Gibbard (1969) sur les oligarchies. Le champ des applications vient à peine d'être ouvert. Jusqu'à présent, les Théorèmes 6 et 7 ont surtout servi en prenant (a) et (b) comme conditions suffisantes et pour retrouver des résultats, comme ceux d'Arrow ou de Gibbard, qui portent sur des domaines non spécialisés d'options et de préférences. Les relations paraissent plus faciles à établir avec les domaines politiques de la théorie du choix social qu'avec ses domaines économiques (décrits par Le Breton et Weymark, 2003).

Nous avons énoncé les résultats de cette section dans la logique générale, ce qui leur donne la plus grande applicabilité possible, mais cette présentation ne reflète pas le processus historique de la découverte, qui a transité par des hypothèses techniques diverses, toutes plus restrictives que les nôtres. Dokow et Holzman, pour leur part, utilisent le formalisme dit *de l'agrégation abstraite*, qu'on peut faire remonter à Fishburn et Rubinstein (1986), Wilson (1975) et Guilbaud (1952). Partant d'un nombre fini  $k$  de propositions qui correspondent aux formules positives de nos agendas, les deux auteurs décrivent les jugements individuels et collectifs sur ces propositions par les valeurs 0 ou 1 qui leur sont attribuées. Après avoir fixé un ordre quelconque sur les propositions, ils ramènent le problème agrégatif à l'étude des sous-ensembles de  $\{0, 1\}^k$  et des fonctions définies à partir de ces sous-ensembles. Plus précisément, si  $\mathcal{E} \subset \{0, 1\}^k$  représente l'ensemble des ensembles de jugement autorisés,  $G : \mathcal{E}^n \rightarrow \mathcal{E}$  représente une fonction de jugement collectif, analogue de notre  $F : D^n \rightarrow D$ . Les conditions définies sur  $F$ , à commencer par l'indépendance et la préservation de l'unanimité, s'énoncent facilement sur  $G$ .

Une présentation aussi ramassée du problème agrégatif débouche sur des démonstrations rapides et élégantes, comme il s'en trouve déjà chez les prédécesseurs de Dokow et Holzman, mais elle tend à effacer le caractère *logique et linguistique* de ce problème et certaines distinctions qui l'accompagnent naturellement. L'étape de définition de l'agenda est absorbée par celle, normalement ultérieure, de la définition des ensembles de jugement autorisés. Or le même  $\mathcal{E}$  peut correspondre à des agendas très différents. Voici un exemple représentatif: soit  $k = 2$  et l'ensemble  $\mathcal{E} = \{(1, 1), (0, 1), (0, 0)\}$ ; en logique propositionnelle classique, il lui correspond au moins deux agendas,  $X = \{a, a \vee b\}^{\pm}$  et  $X' = \{a \wedge b, a \rightarrow b\}^{\pm}$ ,

<sup>29</sup>Dietrich (2007b) obtient le théorème d'Arrow intégral par un autre détour. Il se place dans un cadre plus complexe que celui du théorème canonique, où les formules entretiennent des relations de pertinence (*relevance*) en sus des relations logiques. La dérivation de Nehring (2003) doit être signalée, quoiqu'elle ne relève pas formellement de l'agrégation logique.

qu'il serait conceptuellement abusif d'identifier.<sup>30</sup> Une autre distinction importante, celle des ensembles de jugement de  $D$  et  $D^*$ , devient lourde à traduire dans le cadre de l'agrégation abstraite. A titre d'objection secondaire, l'hypothèse d'un nombre fini  $k$  de propositions empêche de localiser le besoin de finitude là où il se manifeste dans les démonstrations. Grâce au recours modulable à la compacité, la logique générale marque ici un avantage technique, payé d'une certaine lourdeur de maniement.

Le formalisme ensembliste de Dokow et Holzman ne se confond pas avec les présentations, également différentes de la nôtre, qui *optent pour la logique en accentuant la sémantique par rapport à la syntaxe*. Lorsque, par exemple, Pauly et van Hees (2006) se servent de valuations booléennes, plutôt que d'ensembles de formules, pour représenter les jugements d'un individu ou de la collectivité, ils font un simple choix technique à l'intérieur du cadre offert par les calculs propositionnels classiques. La théorie se prête au travail du logicien de bien des manières encore, que nous ne pouvons développer ici.<sup>31</sup>

Le dilemme discursif a guidé les développements techniques des sections qui se sont succédé jusqu'à celle-ci. Le théorème canonique et ses variantes représentent l'approfondissement ultime de ce problème: dès la section 4, il était apparu qu'il était un trilemme, avec la condition d'indépendance comme branche supplémentaire, et il apparaît maintenant qu'il est un *quadrilemme*, avec la définition de l'agenda comme dernière branche, puisqu'il se résout tout aussi bien par la violation des conditions (a) ou (b). Comme nous faisons une distinction tranchée entre le dilemme discursif et le paradoxe doctrinal, il reste à voir si les résultats obtenus pour le premier s'étendent au second.

## 7 Retour au paradoxe doctrinal

Le paradoxe doctrinal se distingue du dilemme discursif, en premier lieu, parce qu'il suppose une distinction que celui-ci efface entre les prémisses et les conclusions de la décision collective, en second lieu, parce qu'il met au centre la doctrine juridique dont celui-ci néglige la spécificité conceptuelle. Nous réexaminerons les deux aspects tour à tour et montrerons - c'est le but dernier de l'article - qu'il est possible de rendre justice au paradoxe doctrinal grâce à la théorie des sections précédentes, à condition de lui faire accomplir un pas de généralité supplémentaire.

Dans une théorie axiomatisée comme celle de l'agrégation logique, la distinction des prémisses et des conclusions ne peut évidemment se manifester que si un axiome au moins la prend en charge. A contrario, supposons que toute condition qu'on voudrait imposer à  $F$  s'applique identiquement aux deux groupes de formules. Alors, puisque la théorie n'a pas les moyens expressifs de les séparer, il faut opter soit contre elle, soit contre la distinction recherchée. L'opposition latente entre Kornhauser et Sager, d'une part, List et Pettit, de l'autre, semble refléter cette alternative. Nous y échappons en sélectionnant un axiome qui

<sup>30</sup>L'exemple provient de List et Puppe (2009).

<sup>31</sup>Voici quelques exemples non exhaustifs. Pauly (2007 et 2009) propose de représenter l'acceptation des formules à l'aide d'un opérateur propositionnel non classique plutôt que par l'appartenance ensembliste. Cariani, Pauly et Snyder (2008) appliquent aux fonctions de jugement social une propriété métathéorique d'invariance qui débouche sur de nouveaux résultats d'impossibilité. Représentatifs d'un courant plus vaste, Konieczny et Pino-Perez (2002) et Pigozzi (2006) étudient des règles de fusion (*merging*) pour les ensembles de jugement. Dans cette dernière option, le dilemme discursif est surmonté par l'abandon de l'indépendance.

portera la distinction à exprimer. Le théorème canonique, dont, pour des raisons de méthode, nous voudrions nous éloigner le moins possible, laisse le choix entre l'indépendance et de la préservation de l'unanimité, la non-dictature ne lui servant qu'à énoncer l'impossibilité. Soit  $P$  et  $C$  les sous-ensembles de  $X$  qui représentent les prémisses et les conclusions; nous supposons pour simplifier qu'ils forment une partition exacte de  $X$ . Parmi les différentes combinaisons axiomatiques, nous retenons *celle qui fait porter, sur  $P$ , l'indépendance et la préservation de l'unanimité, et sur  $C$ , la préservation de l'unanimité seulement*; c'est donc l'indépendance qui joue le rôle discriminant.

Le dispositif choisi détermine mieux la méthode des prémisses que celle des conclusions, et nous justifions cette dissymétrie par la théorie judiciaire elle-même. Chez Kornhauser et Sager (1993), les décisions sur les aspects se prennent à la suite d'un vote majoritaire sur chacun d'entre eux, et nous ne faisons qu'étendre leur idée en imposant à  $P$  les deux conditions fondamentales que satisfait cette procédure. En revanche, si nous lisons bien les mêmes auteurs, *une décision sur l'affaire peut se prendre autrement que par un vote formel*, qu'il soit à la majorité simple ou autrement, et c'est par une simple facilité d'exposition que, dans l'exemple de base, ils font voter les juges sur l'affaire aussi. Le respect du consensus est la seule norme que la procédure doit alors garantir, ce que notre unique condition posée sur  $C$  vient retraduire. Il s'ajoute une raison de sous-déterminer la méthode des conclusions si, dans la partition de  $X$ , on choisit de rattacher la doctrine juridique à  $C$  plutôt qu'à  $P$ ; car il n'y a pas de doute que Kornhauser et Sager, au contraire de Pettit dans sa reconstruction, ne soumettent pas toujours la doctrine juridique au vote formel des juges. L'accord résulte, ici encore, d'une procédure qui n'est pas entièrement décrite et dont on peut seulement affirmer qu'elle respecte la préservation de l'unanimité.

Le Théorème 8 montre que, sous des conditions d'agenda proches de celles du théorème canonique, l'ensemble des conditions proposées sur  $F$  la font basculer dans la dictature. Le Théorème 9 énonce le résultat oligarchique correspondant, et des variantes que nous n'exposons pas ici reproduisent les deux conclusions en renforçant l'indépendance (par la systématité, l'indépendance monotone ou la systématité monotone). L'analyse, due à Dietrich et Mongin (2010), retrouve la conclusion du théorème canonique en partant d'hypothèses plus faibles, puisqu'elle ne requiert l'indépendance ou les conditions apparentées que sur la partie  $P$  de l'agenda. Au plan conceptuel, elle met à l'épreuve le paradoxe doctrinal, qui en ressort consolidé: il n'est pas besoin du vote majoritaire pour que la méthode des prémisses et celle des conclusions s'opposent, il suffit d'énoncer quelques-unes des conditions que les deux méthodes vérifient.

Définissons  $P = \{p, \neg p, q, \neg q, \dots\} \subset X$  comme un sous-ensemble non vide de  $X$  qui est fermé par négation, et posons  $C = X \setminus P$ . Les nouveaux axiomes sur  $F$  s'obtiennent par restriction de ceux du théorème canonique à l'un des deux éléments de la partition.

**Systématité (Indépendance) sur les prémisses:** seulement pour toute paire de formules  $p, q \in P$  (toute formule  $p \in P$ ).

**Préservation de l'unanimité sur les prémisses (sur les conclusions):** seulement pour toute formule  $p \in P$  (toute formule  $p \in X \setminus P$ ).

**Non-dictature sur les prémisses:** il n'existe pas de  $i \in \{1, \dots, n\}$  tel que  $F(A_1, \dots, A_n) \cap P = A_i \cap P$  pour tout  $(A_1, \dots, A_n) \in D^n$ .

**Non-oligarchie sur les prémisses:** il n'existe pas de sous-ensemble non vide  $M \subset \{1, \dots, n\}$  tel que  $F(A_1, \dots, A_n) \cap P = (\bigcap_{j \in M} A_j) \cap P$  pour tout  $(A_1, \dots, A_n) \in D^n$ .

De nouvelles conditions d'agenda tiennent compte de la partition de  $X$ :

(a<sub>P</sub>) Il existe un ensemble minimal incohérent de formules  $Y \subset X$  et un choix de  $Z \subset Y \cap P$  tels que  $|Z| = 2$  et  $Y_{-Z}$  soit cohérent.

(b<sub>P</sub>) Pour tout couple de formules  $p, q \in P$ , il existe des formules  $p_1, \dots, p_k \in P$  telles que

$$p = p_1 \vdash^* p_2 \vdash^* \dots \vdash^* p_k = q.$$

Par définition de  $p \vdash^* q$ ,  $Y \cup \{p\} \vdash^* q$  pour un choix de  $Y \subset X$ ; ainsi, les formules de  $P$  peuvent être logiquement liées par des formules de  $C$ .

(c<sub>P</sub>) Il existe un ensemble minimal incohérent de formules  $Y \subset X$  tel que  $|Y \cap P| \geq 3$ .

**Théorème 8.** Sous les conditions (a<sub>P</sub>) et (b<sub>P</sub>), il n'existe pas de  $F : D^n \rightarrow D$  qui vérifie à la fois:

- sur les prémisses, la non-dictature, l'indépendance et la préservation de l'unanimité,
- sur les conclusions, la préservation de l'unanimité.

Pour un nombre d'individus  $n \geq 3$ , (a<sub>P</sub>) et (b<sub>P</sub>) sont aussi nécessaires à cette conclusion.

**Théorème 9.** L'énoncé reproduit celui du Théorème 8, avec  $F : D^n \rightarrow D^*$  au lieu de  $F : D^n \rightarrow D$ , et la non-oligarchie sur les prémisses au lieu de la non-dictature.

Comme le sous-ensemble  $P$  est quelconque, il est permis de le faire varier librement, et l'on obtient les Théorèmes 6 et 7 en posant  $P = X$ . Ainsi, le théorème canonique et sa version oligarchique sont englobés. De même, les Théorèmes 6' et 7' sont des cas particuliers de ceux que nous n'énonçons pas. La caractérisation d'agenda propre à  $F_{maj}$  se retrouve aussi englobée. On vérifie en effet que (b<sub>P</sub>) implique (c<sub>P</sub>) et que cette dernière condition est nécessaire et suffisante pour qu'il existe un  $F_{maj}(A_1, \dots, A_n) \cap P$  incohérent.

Si l'on renforce la définition des prémisses, on renforce du même coup les résultats. Il serait conforme à la notion ordinaire que les formules de  $P$  aient une liaison logique avec celles de  $C$ , et plus précisément, qu'elles donnent le moyen d'axiomatiser chaque ensemble de jugement  $B \in D$ . On envisagera donc la condition:

(d<sub>P</sub>) Pour tout  $B \in D$ ,

$$B = \{\varphi \in X \mid B \cap P \vdash \varphi\}.$$

Avec l'adjonction de (d<sub>P</sub>), on conclut du Théorème 8 à la dictature, et non plus seulement à la dictature sur les prémisses. Ce corollaire permet de retrouver le Théorème 4 si l'on spécialise la logique générale sous la forme d'un calcul propositionnel classique. On se rappelle que ce résultat limite l'indépendance à  $PV_X \subset X$  et que la section 4 n'énonçait pas toutes les conditions qu'il demande sur  $X$ ; il se trouve qu'elles coïncident avec (a<sub>P</sub>), (b<sub>P</sub>) et (d<sub>P</sub>) dans le cas où  $P = PV_X$ . La généralité de l'analyse se vérifie de cette manière aussi.

Les deux conditions fondamentales  $(a_P)$  et  $(b_P)$  peuvent s'illustrer par de nouvelles variantes de l'exemple de base. Soit l'agenda  $\overline{X}$ , qui inclut la formule  $q = d \iff v \wedge r$  pour la doctrine juridique. Si, comme il est envisageable, on pose  $P = \{v, r\}^\pm$ , alors  $(b_P)$  est violée, et la décision judiciaire échappe à l'impossibilité, au contraire de ce que prédisait le théorème canonique. Comme  $(c_P)$  est aussi violée,  $F_{maj}$  n'a plus d'inconvénient, du moins si  $n$  est impair.

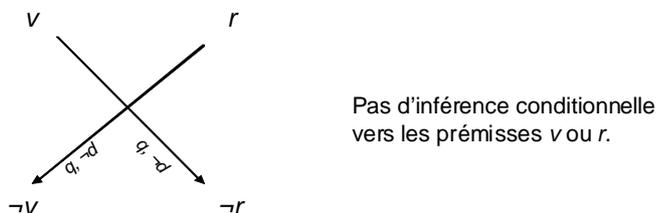


Figure 3: L'agenda  $\overline{X}$  avec  $P = \{v, r\}^\pm$  viole  $(b_P)$ .

En revanche, si l'on prend  $P = \{v, r, q\}^\pm$ , en traitant donc la doctrine comme une prémisses supplémentaire, ce qui est un choix naturel, alors  $(b_P)$  est satisfaite.

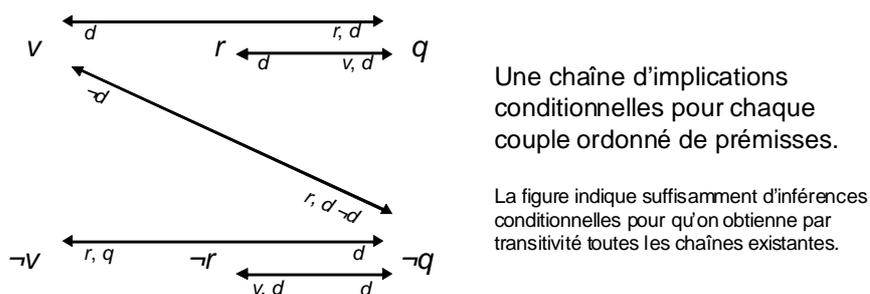


Figure 4: L'agenda  $\overline{X}$  avec  $P = \{v, r, q\}^\pm$  vérifie  $(b_P)$ .

On constate de même que  $(a_P)$  est vérifiée. Ainsi, pour cet ensemble de prémisses, différent de l'ensemble  $P = X$  supposé dans le théorème canonique, la décision judiciaire retombe dans l'impossibilité prédite par ce théorème.

Soit maintenant l'autre agenda considéré,  $\overline{\overline{X}} = \{v, r, d\}^\pm$ , qui s'étudie avec la relation d'inférence  $\vdash_q$  définie à la section précédente. Si l'on prend  $P = \{v, r\}^\pm$ , comme il semble aller de soi,  $(b_P)$  est violée, car aucune inférence conditionnelle n'atteint jamais une formule positive.

Même  $(c_P)$  est violée, de sorte que  $F_{maj}$  peut être à nouveau convoquée. De fait, cette résolution illustre *l'application exclusive de la méthode des prémisses*, qui est la voie de sortie la plus courante du paradoxe doctrinal.<sup>32</sup>

Les agendas qui précèdent illustrent mathématiquement le Théorème 8, mais leur pertinence est à débattre: ils représentent la doctrine juridique à la manière du dilemme discursif, et la fin de la section 2 a soulevé au moins deux objections

<sup>32</sup>Voir Nash (2003). Pour leur part, Kornhauser et Sager vont varier la résolution avec les cas, en recommandant au tribunal de procéder à un "méta-vote" sur la méthode de décision, par les prémisses ou bien par la conclusion.

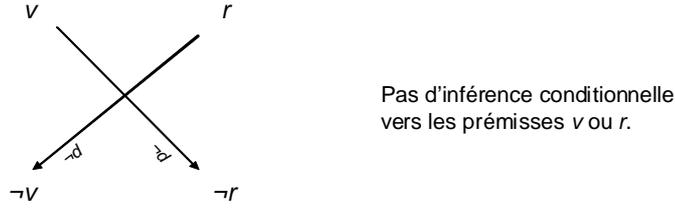


Figure 5: L'agenda  $\overline{\overline{X}}$  avec  $P = \{v, r\}^\pm$  viole (b<sub>P</sub>).

contre cette manière. Suivant la première, notée (i), *il ne va pas de soi que la doctrine se représente par une formule de la logique ordinaire*. Nous renforcerons la critique en montrant que, dans la formulation de la doctrine par équivalence, la biimplication booléenne  $\longleftrightarrow$  a des effets indésirables que l'opérateur non-booléen  $\leftrightarrow$  des logiques conditionnelles évite. L'argumentation s'inspire de celle que Dietrich (2010) déploie en leur faveur plus généralement.<sup>33</sup>

L'agenda  $\overline{\overline{X}}$  engendre la liste suivante de sous-ensembles minimalement incohérents:

$$\begin{aligned} Y_1 &= \{-r, d, q\}, Y_2 = \{-v, d, q\}, \\ Y_3 &= \{v, r, \neg d, q\}, Y_4 = \{q, \neg q\}, \\ Y_5 &= \{v, r, d, \neg q\}, Y_6 = \{-v, \neg d, \neg q\}, Y_7 = \{-r, \neg d, \neg q\}. \end{aligned}$$

A considérer  $Y_5, Y_6, Y_7$ , on voit que la formule  $\neg q = \neg(d \longleftrightarrow v \wedge r)$  a des propriétés contre-intuitives. En effet, un juge pour lequel les deux clauses mentionnées, la validité du contrat et sa rupture, ne sont pas équivalentes à l'attribution de dommages-intérêts, ne devrait pas pour autant se voir interdire l'une de ces possibilités:

- il accepte la validité du contrat, sa rupture et les dommages-intérêts (c'est qu'il a en vue une troisième condition d'aspect, non mentionnée ici, et qu'il la tient pour satisfaite),
- il refuse la validité du contrat et les dommages-intérêts (c'est qu'il conditionne l'attribution des dommages-intérêts à l'absence de validité seulement),
- il constate l'absence de rupture et n'accorde pas les dommages-intérêts (même raisonnement avec l'absence de rupture).

Pour le dire autrement, le théorème de logique propositionnelle classique

$$(*) \quad \neg(d \longleftrightarrow v \wedge r) \dashv\vdash \neg d \longleftrightarrow v \wedge r$$

contredit les intuitions normales sur la délibération judiciaire. *On s'attendrait que le refus de la doctrine juridique n'entraîne pour le juge aucune conséquence déterminée, mais ce n'est pas le cas*. Il n'y a rien à redire, en revanche, aux ensembles  $Y_1$  à  $Y_5$  ou aux théorèmes correspondants:

$$(**) \quad \{d \longleftrightarrow v \wedge r, v, r\} \vdash d, \{d \longleftrightarrow v \wedge r, d\} \vdash v, \{d \longleftrightarrow v \wedge r, d\} \vdash r.$$

<sup>33</sup>Très développées aujourd'hui, les logiques conditionnelles procèdent des travaux de Stalnaker (1968) et Lewis (1973). Elles visent à dépasser les paradoxes de l'implication "matérielle" (booléenne), dont le paragraphe suivant livre un nouvel échantillon.

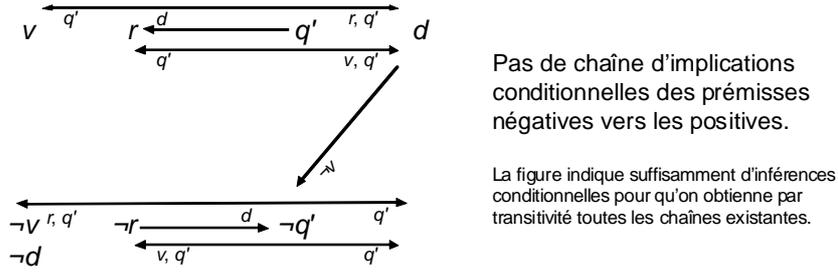


Figure 6: L'agenda  $\overline{X}'$  viole (b).

On s'attend certainement que l'acceptation de la doctrine, par opposition à son refus, ait des conséquences, et celles-là mêmes qui sont indiquées.

Les logiques conditionnelles axiomatisent l'implication,  $\hookrightarrow$ , et donc la bi-implication,  $\leftrightarrow$ , d'une manière qui répond exactement aux desiderata: elles donnent lieu à une liste d'ensembles minimalement incohérents réduite à  $Y_1$  à  $Y_5$ , ou de manière équivalente, aux théorèmes (\*\*\*) à l'exclusion de (\*). C'est pourquoi elles nous semblent préférables aux logiques propositionnelles classiques si l'on veut approfondir la nature du paradoxe doctrinal. Au niveau de la logique générale, il n'est pas nécessaire de trancher entre les systèmes concurrents, car n'importe lequel remplit les besoins.<sup>34</sup>

Il y aurait donc lieu de remplacer  $\overline{X}$  par  $\overline{X}' = \{v, r, d, q'\}^\pm$  avec  $q' = d \leftrightarrow v \wedge r$ . Comme cet agenda est nouveau, il demande à être confronté au théorème canonique. Il apparaît que  $\overline{X}'$  vérifie encore (a), mais non (b), de sorte que la conclusion négative obtenue pour  $\overline{X}$  à la section précédente ne vaut plus maintenant. Il est instructif de comparer la Figure 1 à celle-ci:

La question se pose ensuite de savoir comment les théorèmes de cette section s'appliquent à  $\overline{X}'$ . Nous laissons le lecteur constater que (b<sub>P</sub>) n'est satisfaite pour aucun des choix de  $P$  que nous avons envisagés. La conclusion négative est de nouveau battue en brèche.

Qu'en est-il maintenant de l'agenda  $\overline{\overline{X}} = \{v, r, d\}^\pm$  lorsque la formule placée dans la règle d'inférence est non-classique? La nouvelle règle  $\vdash_{q'}$  se définit par:

$$S \vdash_{q'} \varphi \text{ ssi } S \cup \{q'\} \vdash \varphi,$$

et elle conduit aux mêmes violations de (b) et de (b<sub>P</sub>) que  $\vdash_q$  précédemment. En effet, avec ces deux règles, seule l'acceptation de la doctrine modifie le jeu normal de l'inférence, et dans ce cas, l'équivalence classique et l'équivalence non-classique reviennent au même. Pour le dire autrement, avec  $\vdash_{q'}$ , les sous-ensembles minimalement incohérents sont les mêmes qu'avec  $\vdash_q$ , au remplacement près de  $q$  par  $q'$ .

En résumé, quand la doctrine juridique est *interne* à l'agenda, l'abandon de la logique classique ouvre des possibilités supplémentaires, mais elle n'apporte rien de plus quand la doctrine est *externe*. Y a-t-il moyen de trancher entre les représentations interne et externe, comme nous venons de le faire entre les

<sup>34</sup>Le lecteur peut se reporter à Nuete et Cross (2001).

logiques classique et non-classique? Nous ne le croyons pas. Le dilemme discursif impose automatiquement la représentation interne, et la deuxième objection de la section 2, notée (ii), rejette cette liaison, mais sans imposer l'automatisme de la représentation externe. L'analyse de Kornhauser et Sager et de leurs successeurs n'interdit pas que les juges votent sur la doctrine, même s'ils ne le font normalement pas. Ainsi, notre conclusion dernière se fait prudente: la modélisation préférable est  $\overline{X} = \{v, r, d\}^\pm$  avec pour règle  $\vdash_{q'}$  ou  $\vdash_{q''}$ , puisque cela ne fait pas de différence pratique, mais  $\overline{X}'$  conviendra tout de même dans certaines circonstances. Le seul agenda que nous excluons est  $\overline{X}$ , celui du dilemme discursif.

## 8 Références

- Arrow, K.J. (1951), *Social Choice and Individual Values*, New York, Cowles Foundation and Wiley, 2nd ed., 1963.
- d'Aspremont, C. (1985), *Axioms for Social Welfare Orderings*, dans *Social Goals and Social Organization*, publié par Hurwicz, L., Schmeidler, D., Sonnenschein, H., Cambridge, Cambridge University Press.
- Barwise, J. (1977), An Introduction to First-Order Logic, dans *Handbook of Mathematical Logic*, publié par Barwise, J., Amsterdam, North Holland.
- Cariani, F., Pauly, M., Snyder, J. (2008), Decision Framing in Judgment Aggregation, *Synthese*, 163, 1-24.
- Chang, C.C., Keisler, H.J. (1973), *Model Theory*, Amsterdam, Elsevier.
- Chapman, B. (2002), Rational Aggregation, *Politics, Philosophy and Economics*, 1, 337-354.
- Condorcet (Nicolas de Caritat, Marquis de) (1785), *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris Imprimerie Royale. Repris dans les *Oeuvres*, publiées par Condorcet-O'Connor, A., Arago, M.F., Paris, Firmin Didot, 1847-1849. Sur le site des bibliothèques canadiennes: <http://www.archive.org/details/essaisurlapplica00cond>
- Dietrich, F. (2006), Judgment Aggregation: (Im)Possibility Theorems, *Journal of Economic Theory*, 126, 286-298.
- Dietrich, F. (2007a), A Generalized Model of Judgment Aggregation, *Social Choice and Welfare*, 28, 529-565.
- Dietrich, F. (2007b), Aggregation and the Relevance of Some Issues for Others, Research Memoranda 002, METEOR, Maastricht Research School of Economics of Technology and Organization.
- Dietrich, F. (2010), The Possibility of Judgment Aggregation on Agendas with Subjunctive Implications, *Journal of Economic Theory*, 145, Symposium on Judgment Aggregation, 603-638.
- Dietrich, F., List, C. (2007a), Judgment Aggregation by Quota Rules: Majority Voting Generalized, *Journal of Theoretical Politics*, 19, 391-424.
- Dietrich, F., List, C. (2007b), Arrow's Theorem in Judgment Aggregation, *Social Choice and Welfare*, 29, 19-33.
- Dietrich, F., List, C. (2008), Judgment Aggregation Without Full Rationality, *Social Choice and Welfare*, 31, 15-39.
- Dietrich, F. and List, C. (2010), Majority Voting on Restricted Domains, *Journal of Economic Theory*, 145, Symposium on Judgment Aggregation, 512-543.
- Dietrich, F., Mongin, P. (2007), The Premiss-Based Approach to Logical Aggregation, CR 886/2008, Groupe HEC, Paris.

- Dietrich, F., Mongin, P. (2010), The Premiss-Based Approach to Judgment Aggregation, *Journal of Economic Theory*, 145, Symposium on Judgment Aggregation, 562-582.
- Dokow, E., Holzman, R. (2009), Aggregation of Binary Evaluations for Truth-functional Agendas, *Social Choice and Welfare*, 32, 221-241.
- Dokow, E., Holzman, R. (2010a), Aggregation of Binary Evaluations, *Journal of Economic Theory*, Symposium on Judgment Aggregation, 145, 495-511.
- Dokow, E., Holzman, R. (2010b), Aggregation of Binary Evaluations with Abstentions, *Journal of Economic Theory*, Symposium on Judgment Aggregation, 544-561
- Duddy, C., Piggins, A. (2009), Many-valued Judgment Aggregation: Characterizing the Possibility/impossibility Boundary for an Important Class of Agendas, Working Paper, J.E. Cairnes School of Business and Economics, University of Galway.
- Eckert, D., Monjardet, B. (2009), Guilbaud's Theorem : An Early Contribution to Judgment Aggregation, Document de travail 2009.47, Centre d'Economie de la Sorbonne, Paris.
- Fishburn, P. C., Rubinstein, A. (1986), Algebraic Aggregation Theory, *Journal of Economic Theory*, 38, 63-77.
- Fleurbaey, M., Mongin, P. (2005), The News of the Death of Welfare Economics is Greatly Exaggerated, *Social Choice and Welfare*, 25, 381-418.
- Frege, G. (1918-1919-1923), Logische Untersuchungen, *Beiträge zur Philosophie der deutschen Idealismus*, Hefte I/2, II/3-4, III/1. Trad. fr.: Recherches logiques, dans *Ecrits logiques et philosophiques*, Paris, Le Seuil, 170-234.
- Gärdenfors, P. (2006), An Arrow-like Theorem for Voting with Logical Consequences, *Economics and Philosophy*, 22, 181-190.
- Genest, C. and J. V. Zidek (1986), Combining Probability Distributions: a Critique and an Annotated Bibliography, *Statistical Science*, 1, 114-148.
- Gibbard, A. (1969), Social Choice and the Arrow Conditions, Working Paper, Department of Philosophy, University of Michigan.
- Granger, G.G. (1956), *La mathématique sociale du marquis de Condorcet*, Paris, Presses Universitaires de France (2e éd., Paris, Odile Jacob, 1989).
- Guilbaud, G.T. (1952), Les théories de l'intérêt général et le problème logique de l'agrégation, *Economie appliquée*, 5, 501-551.
- Herzberg, F.S. (2010), Judgment Aggregators and Boolean Algebra Homomorphisms, *Journal of Mathematical Economics*, 2010, 46, 132-140.
- Herzberg, F.S., Eckert, D. (2010), Impossibility Results for Infinite-electorate Abstract Aggregation Rules, IMW Working Paper 427, Universität Bielefeld.
- Heifetz, A., Mongin P. (2001), Probability Logic for Type Spaces, *Games and Economic Behavior*, 35, 31-53.
- Kant, E. (1800), *Logik, ein Handbuch für Vorlesungen*, herausgegeben by Jäsche, G.B., Königsberg. Trad. fr.: *Logique*, Paris, Vrin, 1970.
- Kirman, A.P., Sondermann, D. (1972), Arrow's Theorem, Many agents, and Invisible Dictators, *Journal of Economic Theory*, 267-277.
- Konieczny, S, Pino-Perez, R (2002), Merging Information Under Constraints: a Logical Framework, *Journal of Logic and Computation*, 12, 773-808.
- Kornhauser, L.A. (1992), Modelling Collegial Courts II: Legal Doctrine, *Journal of Law, Economics and Organization*, 8, 441-470.
- Kornhauser, L.A., Sager L.G. (1986), Unpacking the Court, *The Yale Law Journal*, 96, 82-117.
- Kornhauser, L.A., Sager L.G. (1993), The One and the Many: Adjudication in Collegial Courts, *California Law Review*, 81, 1-59.

- Kornhauser, L.A, Sager, L.G. (2004), The Many as One: Integrity and Group Choice in Paradoxical Cases, *Philosophy and Public Affairs*, 32, 249-276.
- Le Breton, M., Weymark, J. (2003), Arrowian Social Choice Theory on Economic Domains, Working Paper, GREMAQ, Toulouse, et Dept of Economics, Vanderbilt University.
- Lehrer, K., Wagner C. (1981), *Rational Consensus in Science and Society*, Dordrecht, Reidel.
- Lewis, D. (1973), *Counterfactuals*, Oxford, Blackwell
- Lismont, L., Mongin P. (2003), Strong Completeness for Weak Logics of Common Belief, *Journal of Philosophical Logic*, 32, 115-137.
- List, C. (2003), A Possibility Theorem on Aggregation over Multiple Interconnected Propositions, *Mathematical Social Sciences*, 45, 1-13. (Corrigendum in *Mathematical Social Sciences* 52:109-110).
- List, C. (2004), A Model of Path-Dependence in Decisions Over Multiple Propositions, *American Political Science Review*, 98, 495-513.
- List, C., Pettit, P. (2002), Aggregating Sets of Judgments: An Impossibility Result, *Economics and Philosophy*, 18, 89-110.
- List, C., Petit, P. (2005), On the Many as One: A Reply to Kornhauser and Sager, *Philosophy and Public Affairs*, 33, 377-390.
- List, C., Puppe, C. (2009), Judgment Aggregation: A Survey, dans *Oxford Handbook of Rational and Social Choice*, publié par Anand, P., Puppe, C., Pattanaik, P., Oxford, Oxford University Press.
- McConway, K.J. (1981), Marginalization and Linear Opinion Pools, *Journal of the American Statistical Association*, 76, 410-414.
- Mongin, P. (1995), Consistent Bayesian aggregation, *Journal of Economic Theory*, 66, 313-351.
- Mongin, P. (1997), Spurious Unanimity and the Pareto Principle, Document de travail, THEMA, Université de Cergy-Pontoise.
- Mongin, P. (2003), L'axiomatisation et les théories économiques, *Revue économique*, 54, 99-138.
- Mongin, P. (2008), Factoring Out the Impossibility of Logical Aggregation, *Journal of Economic Theory*, 141, 100-113.
- Monjardet, B. (2003), De Condorcet à Arrow via Guilbaud, Nakamura et les "jeux simples", *Mathématiques et sciences humaines / Mathematics and Social Sciences*, n° 163, 5-32.
- Nash, J.R. (2003), A Context-Sensitive Voting Protocol Paradigm for Multimembers Courts, *Stanford Law Review*, 56, 75-159.
- Nehring, K. (2003), Arrow's Theorem as a Corollary, *Economics Letters*, 80, 379-382.
- Nehring, K. (2005), The (Im)Possibility of a Paretian Rational, Working Paper, Department of Economics, University of California, Davis.
- Nehring, K., Puppe, C. (2002), Strategy-Proof Social Choice on Single-Peaked Domains: Possibility, Impossibility and the Space Between, Working Paper, Department of Economics, University of California, Davies.
- Nehring, K., Puppe, C. (2008), Consistent Judgement Aggregation: The Truth-Functional Case, *Social Choice and Welfare*, 31, 41-57.
- Nehring, K., Puppe, C. (2010), Abstract Arrowian Aggregation, *Journal of Economic Theory*, 145, Symposium on Judgment Aggregation, 467-494.
- Nehring, K., Puppe, C. (2010b), Justifiable Group Choice, *Journal of Economic Theory*, 145, Symposium on Judgment Aggregation, 583-602.
- Nute, D., Cross, C.B. (2001), Conditional Logic, dans *Handbook of Philosophical Logic*, 4, publié par Gabbay, D., Guenther, F., Dordrecht, Kluwer.

- Pauly, M. (2007), Axiomatizing Collective Judgment Sets in a Minimal Logical Language, *Synthese*, 158, 233-250.
- Pauly, M. (2008), On the Role of Language in Social Choice Theory, *Synthese*, 163, 227-243.
- Pauly, M., van Hees, M. (2006), Logical Constraints on Judgment Aggregation, *Journal of Philosophical Logic*, 35, 569-585.
- Pettit, P. (2001), Deliberative Democracy and the Discursive Dilemma, *Philosophical Issues*, 11, 268-299.
- Piggozzi, G. (2006), Belief Merging and the Discursive Dilemma: An Argument-based Account to Paradoxes of Judgment Aggregation, *Synthese*, 152, 285-298.
- Priest, G. (2002), Paraconsistent Logic, dans *Handbook of Philosophical Logic*, vol. 6 (2ème éd.) publié par Gabbay, D., Guenther, F., Dordrecht, Kluwer Academic Publishers, 287-393.
- Priest, G., Routley, R., Norman, J. (publié par) (1989), *Paraconsistent Logic: Essays on the Inconsistent*, München, Philosophia Verlag.
- Rashed, R. (1974), *Condorcet, Mathématique et société, Choix de textes et commentaires*, Paris, Hermann.
- Sen, A. (1970), *Collective Choice and Social Welfare*, Amsterdam, North Holland.
- Stalnaker, R. (1968), A Theory of Conditionals, dans *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, 2, 98-112.
- van Hees, M. (2007), The Limits of Epistemic Democracy, *Social Choice and Welfare*, 28, 649-666.
- Wilson, R. (1975), On the Theory of Aggregation, *Journal of Economic Theory*, 10, 89-99.