



HAL
open science

Explications en aide multicritère à la décision : schémas déductifs, algorithmes et expérimentations

Manuel Amoussou

► **To cite this version:**

Manuel Amoussou. Explications en aide multicritère à la décision : schémas déductifs, algorithmes et expérimentations. Intelligence artificielle [cs.AI]. Université Paris-Saclay, 2023. Français. NNT : 2023UPAST216 . tel-04612786

HAL Id: tel-04612786

<https://theses.hal.science/tel-04612786>

Submitted on 14 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Explications en aide multicritère à la décision : schémas déductifs, algorithmes et expérimentations

*Explanations in multicriteria decision aiding: deductive schemes,
algorithms and experiments.*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 573 : interfaces : matériaux, systèmes, usages (INTERFACES)
Spécialité de doctorat : Informatique
Graduate School : Sciences de l'Ingénierie et des Systèmes. Référent : CentraleSupélec

Thèse préparée dans l'unité de recherche **Mathématiques et Informatique pour la Complexité et les Systèmes (Université Paris-Saclay, CentraleSupélec)**, sous la direction de **Vincent MOUSSEAU**, Professeur, Université Paris-Saclay, CentraleSupélec (MICS), la co-direction de **Wassila OUERDANE**, Maître de conférences HDR, Université Paris-Saclay, CentraleSupélec (MICS), le co-encadrement de **Khaled BELAHCÈNE**, Maître de conférences, Université Paris-Saclay, CentraleSupélec (MICS) et de **Nicolas MAUDET**, Professeur, Sorbonne Université, (LIP6)

Thèse soutenue à Paris-Saclay, le 11 décembre 2023, par

Manuel AMOUSSOU

Composition du Jury

Membres du jury avec voix délibérative

| | |
|--|----------------------------|
| Patrick MEYER Professeur, IMT Atlantique de Brest | Président |
| Meltem ÖZTÜRK Professeure, Université Paris-Dauphine | Rapporteuse & Examinatrice |
| Marc PIRLOT Professeur, Université de Mons | Rapporteur & Examineur |
| Olivier SPANJAARD Maître de conférences HDR, Sorbonne Université | Examineur |

Titre : Explications en aide multicritère à la décision : schémas déductifs, algorithmes et expérimentations

Mots clés : Aide Multicritère à la Décision, Contexte d'explication, Préférences additives, Programmation mathématique, Expérimentations numériques

Résumé : La thèse traite de l'explication en aide multicritère à la décision, pour des problèmes de choix et de rangement. Elle complète l'état de l'art en proposant une nouvelle approche de l'explication d'une comparaison par paire d'alternatives basée sur la définition de ce que nous appelons un contexte : un triplet rassemblant une relation binaire sur les alternatives, un sous-ensemble de cette dernière composé de comparaisons par paire supposées admises pour le décideur, et un schéma déductif. Le schéma déductif explicite les mécanismes en jeu dans la déduction de la comparaison par paire à expliquer à partir des comparaisons par paire admises. Les mécanismes de déduction considérés s'appuient sur les propriétés de la relation binaire à laquelle appartiennent la comparaison par paire à expliquer et l'ensemble des comparaisons par paire admises. Les relations binaires considérées dans le travail ont toutes pour point commun d'être représentables par un modèle additif. Les propriétés du modèle additif retenues sont celles de transitivité et d'indépendance. Le choix porté sur ces propriétés se justifie par le fait qu'elles correspondent à des modes de raisonnement intuitifs pour les décideurs. À partir de ces propriétés, nous avons construit différents schémas déductifs auxquels peut être conforme l'explication d'une comparaison par paire d'alternatives. Ensuite, nous avons décrit et implémenté différents algorithmes de calcul d'explications conformes à ces schémas déductifs. Ces algorithmes ont permis de mesurer l'expressivité des schémas déductifs auxquels ils se rapportent. On a ainsi pu observer expérimentale-

ment que les propriétés considérées suffisent à reconstruire intégralement, et à partir de leurs paires critiques, toutes les relations d'ordre linéaire additives d'ensembles d'items de cardinalité inférieure ou égale à six. Les autres contributions de la thèse introduisent notre approche de l'explication dans les processus d'aide à la décision multicritère. Le cadre de décision traité est celui dans lequel sont en présence deux acteurs : le décideur et un analyste dont le rôle est d'aider le décideur. Le problème considéré concerne le choix d'un sous-ensemble d'alternatives et où chaque critère possède exactement deux niveaux d'évaluation. Nous montrons comment l'explication peut permettre à l'analyste de soutenir sa recommandation auprès du décideur, de justifier les conclusions robustes obtenues, mais aussi d'éliciter les préférences du décideur à partir de l'explication des conclusions robustes. Pour rendre concrète cette fonction d'élicitation, nous ébauchons un protocole interactif qui, à travers les réactions du décideur aux explications qui lui sont présentées, rend possible la collecte d'information préférentielle additionnelle susceptible de faire progresser le processus d'aide vers la recommandation à formuler. En ce qui concerne l'explication de la recommandation finale, nous avons conduit des expérimentations numériques destinées à mesurer la proportion de recommandations explicables à l'aide du schéma de couverture et à partir de sous-comparaisons par paire non décomposables. Nous avons enfin proposé quelques façons de remédier aux situations d'inexplicabilité.

Title : Explanations in multicriteria decision aiding : deductive schemes, algorithms and experiments.

Keywords : Multicriteria Decision Aiding, Explanation context, Additive preferences, Mathematical Programming, Numerical experiments.

Abstract : This thesis deals with explanation in multicriteria decision aiding for choice and ranking problems. It completes the state of the art by proposing a new approach to the explanation of a pairwise comparison of alternatives based on the definition of what we call a context : a triplet gathering a binary relation on the alternatives, a subset of the latter composed of pairwise comparisons assumed to be admitted by the decision-maker, and a deductive scheme. The deductive scheme explains the mechanisms involved in deducing the pairwise comparison to be explained from the admitted pairwise comparisons. The deductive mechanisms considered are based on the properties of the binary relation to which the set of admitted pairwise comparisons and the pairwise comparison to be explained all belong to. The binary relations considered in this work all have in common that they can be represented by an additive model. The properties of the additive model chosen are those of transitivity and independence. The choice of these properties is justified by the fact that they correspond to intuitive modes of reasoning for decision-makers. Based on these properties, we have constructed various deductive schemes to which the explanation of a pairwise comparison of alternatives can be conformed to. We then described and implemented various algorithms for calculating explanations that correspond to these deductive schemes. These algorithms were used to measure the expressiveness of the deductive schemes to which they relate. We

were able to observe experimentally that the properties considered are sufficient to completely reconstruct, from their critical pairs, all additive linear order relations of item sets of cardinality less than or equal to six. The other contributions of the thesis introduce our approach to explanation in multi-criteria decision support processes. The decision framework considered is one in which two actors are present : the decision-maker and an analyst whose role is to help the decision-maker. The problem considered concerns the choice of a subset of alternatives, where each criterion has exactly two levels of evaluation. We show how the explanation can enable the analyst to support his recommendation to the decision-maker, to justify the robust conclusions obtained, but also to elicit the decision-maker's preferences from the explanation of the robust conclusions. To make this elicitation function concrete, we are sketching out an interactive protocol which, through the decision-maker's reactions to the explanations presented to him, makes it possible to gather additional preferential information likely to make the aiding process progress towards the recommendation to be formulated. Regarding the explanation of the final recommendation, we have conducted numerical experiments to measure the proportion of recommendations that can be explained using the covering scheme and non-decomposable pairwise sub-comparisons. Finally, we proposed a number of ways to remedy inexplicability situations.

À Mathieu Jaume

Remerciements

Mes profonds remerciements vont à l'endroit de tous ceux et toutes celles qui ont suscité, encouragé et contribué à la concrétisation de ce projet de thèse :

- Nicolas Maudet, sans qui rien de tout ceci n'aurait été possible,
- Vincent Mousseau et Wassila Ouerdane, mes directeurs de thèse,
- Khaled Belahcène, pour sa Présence,
- Christophe Labreuche, pour les riches interactions,
- Meltem Öztürk et Marc Pirlot, pour avoir accepté de rapporter le manuscrit de thèse,
- Patrick Meyer et Olivier Spanjaard, pour avoir aussi accepté de participer à l'évaluation de ce travail,
- tous les collègues du laboratoire MICS qui n'est petit que de taille,
- Elisabeth Delozanne qui a facilité mon inscription en Licence 2 d'informatique,
- Yann-Thierry Mieg, Bruno Escoffier et Jean-Daniel Kant qui ont soutenu mes premiers pas dans la recherche en Informatique,
- Stéphane Doncieux, Jessica Vala-Arlapen, Olivier Castagna pour leurs encouragements,
- tout le personnel du SHSE et de la PPTI de Sorbonne Université en particulier Sandra Schirmer, Nathalie Picot et Bijan Majeri pour la confiance placée en moi,
- Mes tante et oncle Rollande et Gontran Aganahi, qui ont cru en ce projet et ont créé et entretenu les conditions (familiales) de sa réalisation,
- Lynda Belhia pour son soutien et ses aides précieuses,
- tous les étudiants du groupe 2 du cours d'Algorithmique II pour le sacrifice consenti en vue du succès de la soutenance du 11 décembre : Yann, Adel, Iman, Gautier, Aurélien, Halimatou, Marie, Daba, Iona, David, Massilva, Raphaël, Stella, Emmanuella, Zélie, Guillame, Hoang Thuy, Hoang Nguyen, Zhenming et Éric,
- enfin, tous ceux, nombreux (qui se reconnaîtront), que j'ai côtoyés et qui ont cru en ce projet.

Table des matières

| | | |
|----------|--|-----------|
| 1 | Introduction générale | 19 |
| 1.1 | Introduction | 21 |
| 1.2 | Explication en décisions multicritère et collective | 25 |
| 1.2.1 | Explication axée modèle d'utilisateur et/ou de décision | 25 |
| 1.2.2 | Représentation de l'explication | 28 |
| 1.2.3 | Autres considérations : Interaction, Inexplicabilité, Complexité | 30 |
| 1.3 | Conclusion et plan du manuscrit | 33 |
| 1.3.1 | De la nécessité d'une nouvelle approche de l'explication en AMCD | 34 |
| 1.3.2 | Plan du manuscrit | 35 |
| | | |
| I | Schémas déductifs pour l'explication de préférences additives | 38 |
| | | |
| 2 | Explication du modèle additif : principes et définitions | 40 |
| 2.1 | Introduction | 42 |
| 2.2 | Principes décisionnels associés au modèle additif | 43 |
| 2.2.1 | Propriétés fondamentales et dérivées | 43 |
| 2.2.2 | Mécanismes de déduction associés | 46 |
| 2.3 | Formalisation de l'explication pour le modèle additif | 51 |
| 2.3.1 | Exemple introductif | 52 |
| 2.3.2 | Schémas déductifs | 54 |
| 2.3.3 | Retour sur l'exemple introductif | 57 |
| 2.4 | Conclusion et perspectives | 62 |
| | | |
| 3 | Pouvoir déductif de schémas d'explication | 63 |
| 3.1 | Introduction | 65 |
| 3.2 | Description générale d'algorithmes de calcul d'explication | 66 |
| 3.2.1 | Le cas des algorithmes de la famille \mathcal{T}_r | 67 |
| 3.2.2 | Le cas des algorithmes de la famille \mathcal{C}_g | 68 |
| 3.2.3 | Le cas des algorithmes de la famille \mathcal{C}_{or} | 69 |

| | | |
|-------|---|----|
| 3.2.4 | Les algorithmes implémentés | 72 |
| 3.3 | Évaluation empirique du pouvoir déductif de schémas d'explication . . . | 73 |
| 3.3.1 | Définitions et rappels | 74 |
| 3.3.2 | Protocole expérimental et résultats | 77 |
| 3.4 | Évaluation du pouvoir déductif des propriétés des ROLA | 83 |
| 3.4.1 | Évolution dynamique du contexte | 84 |
| 3.4.2 | Application | 87 |
| 3.5 | Conclusion et perspectives | 88 |

II Explication de préférences robustes et de recommandations 89

| | | |
|----------|--|------------|
| 4 | Fonctions de l'explication de préférences robustes | 91 |
| 4.1 | Introduction | 93 |
| 4.2 | Brève revue de littérature et définitions | 95 |
| 4.2.1 | Relation nécessaire | 95 |
| 4.2.2 | Conclusions robustes dans le cadre du choix multiple | 97 |
| 4.2.3 | Explications <i>questionnable</i> et à l'aide de <i>preference swaps</i> | 99 |
| 4.3 | Fonction de justification de l'explication | 102 |
| 4.3.1 | Exemple introductif | 102 |
| 4.3.2 | Justification des conclusions robustes de l'exemple introductif . . | 104 |
| 4.3.3 | Expérimentations numériques | 110 |
| 4.4 | Fonction d'élicitation de l'explication | 117 |
| 4.4.1 | Exemple introductif | 118 |
| 4.4.2 | Illustration de la fonction d'élicitation sur l'exemple introductif . | 120 |
| 4.4.3 | Ébauche d'un protocole interactif | 125 |
| 4.4.4 | Formalisation et modélisation | 127 |
| 4.5 | Conclusion et perspectives | 134 |
| 5 | Explications par couverture de recommandations | 136 |
| 5.1 | Introduction | 138 |
| 5.2 | Description du cadre de décision et du contexte d'explication | 139 |
| 5.2.1 | Cadre de décision | 139 |
| 5.2.2 | Contexte d'explication | 141 |
| 5.3 | Explication du choix de la meilleure alternative à l'aide de la PLNE . . . | 143 |
| 5.3.1 | NP-complétude du problème de l'explication par couverture . . . | 143 |
| 5.3.2 | Explication par couverture d'une comparaison par paire | 145 |
| 5.3.3 | Explication par couverture d'une recommandation | 147 |
| 5.4 | Remédier à l'inexplicabilité de la recommandation | 151 |
| 5.4.1 | Recours à la transitivité | 153 |

| | | |
|----------|---|------------|
| 5.4.2 | Relaxation de la recommandation : du 1–best au k –best | 155 |
| 5.4.3 | Translation de la fonction de score : de ω à ω' | 157 |
| 5.5 | Expérimentations numériques | 160 |
| 5.5.1 | Description du protocole expérimental | 161 |
| 5.5.2 | Résultats et discussions | 162 |
| 5.6 | Conclusion et perspectives | 165 |
| 6 | Conclusion générale | 167 |
| 6.1 | Synthèse des contributions et des résultats | 167 |
| 6.2 | Perspectives | 169 |
| A | Génération exhaustive des ROLA | 178 |

Liste des tableaux

| | | |
|------|--|-----|
| 2.1 | Classement des alternatives de $2^{[4]} \setminus \{\emptyset\}$ suivant ω (Équation 2.5) | 48 |
| 2.2 | Niveaux de différents schémas déductifs | 57 |
| 3.1 | Ensemble des algorithmes implémentés | 73 |
| 3.2 | Nombre total de relations d'ordre linéaire additives en fonction de m | 75 |
| 3.3 | Tailles des ensembles \mathbb{T}_m en fonction de m | 76 |
| 3.4 | Fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_4 inexplicables | 79 |
| 4.1 | Description de l'ensemble \mathbf{A} | 103 |
| 4.2 | Description de l'ensemble \mathbf{A}^R | 104 |
| 4.3 | Scores des alternatives de référence (Table 4.2) suivant ω (Équation 4.5) | 104 |
| 4.4 | Justification du choix impossible x^{46} : Table d'ensembles pro et con | 107 |
| 4.5 | Justification du choix nécessaire x^{91} : Table d'ensembles pro et con | 107 |
| 4.6 | Résultats d'analyses descriptives du protocole expérimental | 116 |
| 4.7 | Valeurs numériques de la Figure 4.10. | 116 |
| 4.8 | Description de l'ensemble \mathbf{A} | 119 |
| 4.9 | Description de l'ensemble \mathbf{A}^R | 119 |
| 4.10 | Scores des alternatives de référence (Table 4.9) suivant ω (Équation 4.12) | 119 |
| 5.1 | Exemple de table de performance | 140 |
| 5.2 | Vue ensembliste des alternatives de la Table 5.1 et leurs scores suivant ω | 141 |
| 5.3 | Table de performance \mathbf{A}_2 et scores suivant ω_1 (Équation 5.2) | 148 |
| 5.4 | Éléments de preuve de l'inexplicabilité de $x^{35} \succ_{\omega_1} x^{72}$ | 152 |
| 5.5 | Table de performance \mathbf{A}_3 et scores suivant ω_1 (Équation 5.2) | 152 |
| 5.6 | Valeurs numériques de la Figure 5.11 | 163 |
| 5.7 | Valeurs numériques de la Figure 5.12 | 164 |

Liste des Algorithmes

| | | |
|---|---|-----|
| 1 | Algorithme générique calculant des explications conformes au schéma composé <i>cg/sch</i> | 69 |
| 2 | Explication d'un ensemble \mathbb{E} de comparaisons par paire en appliquant une évolution dynamique du contexte. | 87 |
| 3 | Explication par couverture d'une recommandation | 147 |
| 4 | Inexplicabilité : implémentation de la <i>Relaxation de la recommandation</i> . | 157 |
| 5 | Génération exhaustive de relations d'ordre linéaire additives | 179 |

Table des figures

| | | |
|------|--|-----|
| 2.1 | Explication de $d \succ_{\omega} c$ sous forme d'implications logiques | 58 |
| 2.2 | Explication de $d \succ_{\omega} c$ sous forme de diagramme | 58 |
| 2.3 | Explication de $c \succ_{\omega} a$ sous forme d'implications logiques | 59 |
| 2.4 | Explication de $c \succ_{\omega} a$ sous forme de diagramme | 59 |
| 2.5 | Explication de $bd \succ_{\omega} ac$ sous forme d'implications logiques | 60 |
| 2.6 | Explication de $bd \succ_{\omega} ac$ sous forme de diagramme | 61 |
| 2.7 | Explication (alternative) de $bd \succ_{\omega} ac$ sous forme de diagramme | 61 |
| 3.1 | Graphe d'explication $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ de $d \succ_{\omega} c$ (Paragraphe 2.3.3) | 68 |
| 3.2 | Fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_5 inexplicables | 80 |
| 3.3 | Diagrammes en boîte synthétisant les fonctions de répartition de la Figure 3.2 | 81 |
| 3.4 | Synthèse des fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_6 inexplicables | 82 |
| 3.5 | Explication de $ab \succ_{\omega''} e$ sous forme de diagramme. | 85 |
| 3.6 | Évolution dynamique du contexte : explication de $ab \succ_{\omega''} e$ sous forme de diagramme à partir de $\mathbb{B}_{\succ_{\omega''}}$ | 86 |
| 4.1 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.3.1). | 105 |
| 4.2 | Justification du choix impossible x^{37} | 106 |
| 4.3 | Justification du choix impossible x^{92} | 107 |
| 4.4 | Justification du choix impossible x^{56} | 108 |
| 4.5 | Justification du choix impossible x^{30} | 108 |
| 4.6 | Justification du choix impossible x^{46} | 111 |
| 4.7 | Justification du choix nécessaire x^{91} (partie 1 / 3) | 112 |
| 4.8 | Justification du choix nécessaire x^{91} (partie 2 / 3) | 113 |
| 4.9 | Justification du choix nécessaire x^{91} (partie 3 / 3) | 114 |
| 4.10 | Proportions de conclusions robustes justifiables. | 117 |

| | | |
|------|--|-----|
| 4.11 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P} 1}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1). | 120 |
| 4.12 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P} 1}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1). | 121 |
| 4.13 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P} 2}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1). | 123 |
| 4.14 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P} 3}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1). | 124 |
| 4.15 | Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P} 4}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1). | 125 |
| 4.16 | Graphe d'explication mixte questionnable $\mathcal{G}_{\mathbb{P} 0, \omega}^{(x^{42}, x^7)}$ | 130 |
| 4.17 | Explications questionnables de la Figure 4.16 | 130 |
| 5.1 | Exemple 1 d'explication de la recommandation x^{100} | 148 |
| 5.2 | Exemple 2 d'explication de la recommandation x^{100} | 149 |
| 5.3 | Recommandation de x^{35} : explications de $x^{35} \underset{\omega_1}{\succ} x^{28}$ et $x^{35} \underset{\omega_1}{\succ} x^{21}$. . . | 151 |
| 5.4 | Recommandation de x^{74} : explications de $x^{74} \underset{\omega_1}{\succ} x^{21}$ et $x^{74} \underset{\omega_1}{\succ} x^{22}$. . . | 152 |
| 5.5 | Inexplicabilité : illustration du <i>Recours à la transitivité</i> | 154 |
| 5.6 | Explication de la recommandation de x^{35} via un recours à la transitivité . | 156 |
| 5.7 | Inexplicabilité : illustration de la <i>Relaxation de la recommandation</i> | 156 |
| 5.8 | Inexplicabilité : application de la relaxation de la recommandation | 158 |
| 5.9 | Incomplétude : illustration de la <i>Translation</i> de ω | 158 |
| 5.10 | Recommandation de x^{74} : explication après translation de la fonction de score | 161 |
| 5.11 | Pourcentage d' <i>instances</i> explicables par couverture en fonction du type de comparaisons par paire admises utilisées | 163 |
| 5.12 | Pourcentage d' <i>instances</i> explicables avec mise en œuvre des mesures de traitement de l'inexplicabilité | 164 |

Chapitre 1

Introduction générale

Résumé

Dans ce chapitre introductif, nous rappelons brièvement en quoi consiste l'aide multicritère à la décision et tentons d'entrée de jeu de montrer comment l'introduction de l'"Explication" dans un processus d'aide peut être profitable au décideur comme à l'analyste. Nous dressons ensuite un bref état de l'art de la problématique de l'explication en décisions multicritère et collective. Nous organisons cette présentation sous forme de thématiques pertinentes pour notre travail. Sont entre autres abordées les questions de la nature (preuve ou non) de l'explication, de sa représentation (langage naturel, tables, graphiques ...) et de son calcul (algorithmes, classes de complexité). Aussi, faisons-nous remarquer que l'inexplicabilité d'une décision est relative aux spécificités de la méthode d'explication employée (schémas d'argument, patterns d'explication ...). Ce constat nous conduit à introduire le concept de "*contexte*" : une même décision pouvant être explicable dans un "*contexte*" et pas dans un autre.

C'est autour de cette notion de "*contexte*" que nous proposons une nouvelle approche de l'explication en aide multicritère à la décision. Les problématiques traitées sont essentiellement celles reposant sur une évaluation relative des alternatives (choix et rangement). Le *contexte*, c'est le triplet composé d'une relation binaire de préférences à laquelle appartient la comparaison par paire à expliquer, un sous-ensemble de cette relation composé de comparaisons par paire admises (c'est-à-dire dispensables d'explication) et d'un schéma déductif qui explicite les mécanismes qui, à partir de ces dernières, permettent de déduire la comparaison par paire à expliquer. Les schémas déductifs s'appuient sur les axiomes de la relation de préférences considérée. Dans tout notre travail, nous avons fait l'hypothèse de préférences représentables par un modèle additif. Le chapitre s'achève en dressant le plan du manuscrit.

Sommaire

| | | |
|------------|---|-----------|
| 1.1 | Introduction | 21 |
| 1.2 | Explication en décisions multicritère et collective | 25 |
| 1.2.1 | Explication axée modèle d'utilisateur et/ou de décision | 25 |
| 1.2.2 | Représentation de l'explication | 28 |
| 1.2.3 | Autres considérations : Interaction, Inexplicabilité, Complexité | 30 |
| 1.3 | Conclusion et plan du manuscrit | 33 |
| 1.3.1 | De la nécessité d'une nouvelle approche de l'explication en AMCD | 34 |
| 1.3.2 | Plan du manuscrit | 35 |

1.1 Introduction

Décider, c'est complexe et difficile [Bouyssou et al., 2006b].

Ceci justifie le recours, dans certaines situations de décision, critiques ou non, à une "aide" portée au "décideur" par un oracle, un géomancien Fâ, une prêtresse vaudoue ou une experte en épidémiologie Avec les progrès de l'informatique en général et de l'intelligence artificielle en particulier, ont massivement été conçus et commercialisés des systèmes (automatiques) d'aide à la décision – *Decision Support System (DSS)* en anglais – destinés à accompagner des organisations de différents domaines de compétences, dans leurs processus de prise de décision. Les DSS sont des programmes informatiques qui encodent des connaissances d'experts et qui aident leurs utilisateurs à prendre des décisions de gestion [Greer et al., 1994]. Constatant que nombre de DSS ne parviennent pas à expliquer et justifier convenablement leurs recommandations, ces auteurs décrivent EXPLAIN, un système hybride d'aide à la décision qui interprète les données quantitatives de simulation de productions agricoles et construit, en fonction du niveau d'expertise de l'utilisateur, des explications personnalisées présentées sous la forme de textes écrits. Le système VOTE de [Slade, 1992] produit également, et dans différentes langues, des explications du comportement (prédit) de vote des membres de la Chambre des Représentants des États-Unis. Un tel outil doterait ainsi certains lobbyistes d'arguments persuasifs à faire valoir auprès de sénateurs américains. Dans le domaine (encore plus sensible) de la gestion des urgences occasionnées par un accident radiologique en Europe, [Papamichail and French, 2003] ont développé un module, intégré au DSS RODOS qui, construit à partir d'une bibliothèque de plans de textes et d'un générateur de phrases, l'explication de la comparaison de différentes stratégies d'action. Les concepteurs de ce module appelé FES (*Fine Expert System*) précisent que ce dernier n'a pas vocation à remplacer les analystes ; il est une sorte d'assistant de l'analyste qui l'aide à mieux comprendre l'interaction entre les différents attributs et alternatives (stratégies).

Il en va de même des outils et méthodes utilisés en Aide MultiCritère à la Décision (AMCD). Ceux-ci ne sont pas à considérer comme des algorithmes et modèles indépendants qui, appliqués à une certaine information, permettent d'aboutir à une conclusion [Bouyssou et al., 2006b]. Ces derniers auteurs poursuivent en rappelant qu'ils (les outils et méthodes) sont utilisés dans un flux d'interactions structuré autour d'un processus de décision dans lequel un acteur – le client – demande des conseils et de l'aide à un autre acteur – l'analyste – qui a, de ces outils et méthodes, une bonne connaissance et qui en maîtrise l'usage. Le client n'étant pas nécessairement le décideur, on commence à percevoir le rôle que pourrait jouer l'"Explication" en AMCD : celui de faciliter la reddition de comptes et de garantir la transparence. Aussi, l'intégration de l'"Explication" au processus d'aide à la décision multicritère peut être souhaitable en raison (justement) de la multiplicité des critères (souvent conflictuels [Roy, 1996]) qui caractérise les objets (alternatives) qu'on y manipule et qui rend la prise de décision complexe. Enfin, puis-

qu'en AMCD on traite de *préférences* qui, par nature, sont *subjectives* [Bouyssou et al., 2006b, Cailloux, 2012], on peut espérer que l'«Explication» pourrait d'une certaine façon contribuer à les *objectiver*.

Dans ce travail, nous nous intéressons à répondre à la question : «*Pourquoi et Comment Expliquer en AMCD?*» – plus au *Comment?* qu'au *Pourquoi?* auquel nous avons déjà essayé de fournir une brève de réponse. Ici, nous tenterons de répondre factuellement au *Pourquoi?*. Ceci va nous permettre d'identifier le ou les *Quoi?* au(x)quel(s) va s'appliquer l'explication. Mais avant, il nous paraît utile de rappeler brièvement en quoi consiste l'Aide MultiCritère à la Décision. Ce bref rappel s'appuiera principalement sur l'ouvrage intitulé *Evaluation and decision models with multiple criteria : Stepping stones for the analyst* de [Bouyssou et al., 2006b].

Dans sa version simplifiée, un processus d'aide multicritère à la décision met en présence deux acteurs : un décideur (\mathcal{DM}^1) et un analyste (\mathcal{DA}^2). Le \mathcal{DM} a un problème de décision et sollicite le \mathcal{DA} avec qui, il va s'engager dans un processus interactif allant de la représentation de la situation du problème puis sa formulation à la construction d'un modèle de préférences et la formulation d'une recommandation. La représentation du problème et sa formulation vont permettre d'identifier sa structure (combinatoire ou non), de retenir les différents critères, de définir l'ensemble des alternatives et de déterminer son type (un choix, un rangement ou un tri). Quant à la construction du modèle de préférences, elle est la composante du processus d'aide dans laquelle le \mathcal{DA} applique ses connaissances méthodologiques aux informations préférentielles fournies par le \mathcal{DM} . L'objectif de la méthode est d'aboutir à une façon de comparer qui convienne au décideur, dans le sens où elle lui donne des arguments qu'il trouve suffisamment convaincants pour valider les recommandations qui en découlent [Cailloux, 2012]. Le choix de la méthode n'est nullement neutre car différentes méthodes peuvent conduire à des conclusions complètement différentes [Ishizaka and Nemery, 2013] : il est tributaire de l'ensemble des propriétés qu'elle vérifie. Il appartient donc au \mathcal{DA} de fixer les propriétés que la méthode de son choix doit remplir et d'informer (à partir de la connaissance axiomatique qu'il en a) le \mathcal{DM} des implications de celui-ci. Dans cette perspective, le \mathcal{DA} peut être amené à retenir une classe donnée de modèles de préférence : le modèle additif [Keeney and Raiffa, 1976] par exemple. [Belton and Stewart, 2002] indiquent que déterminer une bonne classe de modèles est un des problèmes difficiles auxquels l'analyste en AMCD fait face. Une fois la classe de modèles retenue, le \mathcal{DA} devra déterminer les valeurs des paramètres du modèle qui servira à la formulation de sa recommandation via un processus d'élicitation [Fischhoff, 1991] des préférences du \mathcal{DM} . L'élicitation peut être directe ou indirecte suivant que ce dernier fournit directement ces valeurs de paramètres ou qu'ils sont obtenus à partir d'affirmations de ses préférences sur un ensemble d'alternatives servant d'exemples (le paradigme de désagrégation des préférences [Sis-

1. Decision Maker.
2. Decision Analyst.

kos et al., 2016]). Quelle que soit sa nature, l'élicitation requiert de la part du \mathcal{DM} un effort cognitif plus ou moins important; ce qui expose l'ensemble du processus d'aide à la décision à des erreurs ou des inconsistances qu'il faudra traiter [Mousseau et al., 2003]. En ce qui concerne les procédures d'élicitation indirecte, [Bouyssou et al., 2006b] signalent que bien souvent, le \mathcal{DM} n'est capable que de comparer un petit nombre d'alternatives et est indécis sur un grand nombre d'autres paires d'alternatives. Ses préférences sont généralement incomplètes et certaines, instables, peuvent changer au cours du processus d'aide à la décision. Ce point de vue est également partagé par [Belton and Stewart, 2002] qui précisent que les préférences (du décideur) tendent à évoluer au fur et à mesure qu'il parvient à une meilleure compréhension du problème. Tous ces facteurs (choix d'une classe de modèles de préférence, erreurs, indécisions, hésitations...) influencent la formulation de la recommandation, l'aboutissement du processus d'aide à la décision.

[Bouyssou et al., 2006b] insistent sur le fait qu'en plus d'être techniquement fondée et opérationnalisable, la recommandation finale doit pouvoir être légitimée dans le cadre du processus de décision pour lequel elle a été conçue. À ce processus de décision, peuvent être associées d'autres parties prenantes auprès desquelles la recommandation doit pouvoir se justifier. C'est l'un des rôles que peut jouer l'"Explication" en AMCD : rendre la recommandation accessible et comprise des différentes parties prenantes au processus de décision. Comme autres services que peut rendre l'"Explication" en AMCD, nous pouvons, au vu de la description qu'on en a faite précédemment, affirmer qu'en fournissant des éléments d'explication d'une comparaison par paire d'alternatives par rapport à laquelle le \mathcal{DM} est indécis, le \mathcal{DA} **soutient directement ce dernier dans l'effort cognitif que requiert l'expression de ses préférences sur des alternatives qu'il ne connaît pas et contribue indirectement à limiter les erreurs de jugements pouvant négativement affecter la décision finale**. Par ailleurs, **à travers l'explication de comparaisons par paire d'alternatives, le \mathcal{DA} peut s'assurer du choix de la bonne classe de modèles** représentant les préférences du \mathcal{DM} . C'est ce que proposent [Labreuche et al., 2015] à travers la description d'un protocole de dialogue dans lequel un analyste artificiel utilise une variété de classes de modèles de préférence pour construire sa recommandation; la justification qu'il en donne s'appuyant sur les propriétés des dites classes et les feed-backs du décideur à cette justification lui permettant d'écarter certaines classes incompatibles. Ce faire, en AMCD (où l'analyste est généralement un humain), est également envisageable (et même souhaitable) car comme le rappellent [Bouyssou et al., 2006a], certaines classes de modèles ont été axiomatisées. Ces axiomes prennent généralement la forme de conditions nécessaires et (parfois suffisantes) de représentabilité des préférences du décideur par un modèle de la classe retenue. C'est le cas par exemple du système de propriétés de *cancellation* [Fishburn, 1997, Fishburn, 2001] caractérisant une relation de préférences représentable par un modèle additif dans le cas "fini" c'est-à-dire une relation définie sur un ensemble fini d'alternatives. Déterminer une classe de modèles, c'est également déterminer ce qui est considéré comme un

mode de raisonnement a priori acceptable [Cailloux, 2012]. Ce mode de raisonnement rend possible l'inférence de préférences mais permet également la détection de situation d'inconsistance. En ce qui concerne l'inférence, **présenter au \mathcal{DM} et (surtout) lui expliquer les déductions faites à partir ses préférences incomplètes, l'éclaire sur le mode de raisonnement qu'utilise le \mathcal{DA} , un mode de raisonnement qu'il peut s'approprier ou rejeter parce que non conforme au sien.** De même, nous pensons que l'“Explication” peut servir dans la résolution des situations d'inconsistance liée à des erreurs de jugements du \mathcal{DM} . En effet, cette inconsistance est due à la violation d'un ou plusieurs axiomes de la classe de modèles retenue par le \mathcal{DA} . Identifier cette ou ces propriété(s) et **exhiber des explications de deux propositions contradictoires déduites à partir des préférences exprimées par le \mathcal{DM} nous paraît être une bonne façon de l'impliquer dans l'activité de résolution de l'inconsistance.** Enfin, la méthode adoptée par le \mathcal{DA} pour déterminer les paramètres du modèle de préférences du \mathcal{DM} peut nécessiter le recours à des outils techniques de Recherche Opérationnelle tels que la Programmation Linéaire. C'est le cas par exemple de l'élicitation indirecte du modèle additif par regression ordinale [Jacquet-Lagrezze and Siskos, 1982] où la fonction de valeur “optimale” retournée par le programme linéaire dépend du choix (arbitraire) de la fonction objectif (minimisation de la somme des variables d'écarts ...) et de l'implémentation de l'algorithme de programmation linéaire utilisé. D'autres procédures comme celle de [Kadziński et al., 2012] (problématiques de choix et de rangement) et de [Greco et al., 2011] (problématique de tri) reposent sur la définition de “critères de représentabilité” que pourraient remplir certaines fonctions de valeur représentant les préférences du décideur. Ces “critères de représentabilité”, que le \mathcal{DM} peut néanmoins sélectionner, sont aussi à optimiser en recourant à la Programmation Linéaire. Dans l'un ou l'autre des cas, chacune de ces procédures, du fait de sa technicité (optimisation sous contraintes linéaires), présente le risque de paraître abstraite au \mathcal{DM} et donc de compromettre la bonne conduite du processus d'aide à la décision. En la matière, **nous pensons que l'Explication peut permettre de supprimer la part d'arbitraire inhérente à ces procédures en retenant une fonction de valeur dont les implications sur l'ensemble des alternatives en jeu sont (facilement) explicables.**

Dans le (long) paragraphe précédent, nous avons tenté de montrer comment l'introduction de l'“Explication” dans un processus d'aide multicritère à la décision peut être profitable tant au décideur qu'à l'analyste, qu'il s'agisse d'expliquer la recommandation finale, une déduction de comparaison par paire ou un simple élément de la relation de préférence induite par une fonction de valeur compatible avec les préférences du décideur. Nous nous sommes limités dans ce travail à étudier la question de l'explication en rapport aux problématiques de choix et de rangement en AMCD où l'on évalue la performance d'une alternative relativement à une autre. Ceci a pour conséquence que l'objet de base auquel va s'appliquer l'explication est la comparaison par paire d'alternatives. Comme indiqué précédemment, nos contributions porteront principalement sur les façons dont l'explication (de la préférence d'une alternative sur une autre) peut être conçue

en AMCD. Comme nous le préciserons dans la Section 1.3, elles (nos contributions) se distinguent de la variété de celles proposées dans la littérature de par la nécessité de définition du *contexte* dans lequel l'explication sera produite ; le *contexte* précisant entre autres les prémisses et la structure du raisonnement qui appliqué aux prémisses permet de déduire la comparaison par paire à expliquer. De ce fait, de telles explications peuvent être considérées comme des « explications scientifiques »³ [Kass and Finin, 1988]. Mais avant, nous réaliserons une brève revue de littérature (Section 1.2) de la thématique de l'explication dans le domaine de la décision en général et de la décision multicritère et collective en particulier ; chaque sous-section aborde une facette de cette question qui facilite le positionnement de nos contributions par rapport aux travaux qui y sont mentionnés. Enfin, la Section 1.3 conclura le chapitre et dressera le plan du manuscrit.

1.2 Explication en décisions multicritère et collective : une revue de littérature

Cette section a pour vocation de dresser un état de l'art de la question de l'explication dans les situations de prise de décision en général et particulièrement, lorsque plusieurs points de vue (critères) ou agents sont à prendre en compte. Les aspects de la question qui vont nous intéresser seront relatifs à la façon dont l'explication est construite. Ceux-ci englobent le choix de la structure de représentation de l'explication (langage naturel, autres langages, graphiques, patterns, schémas d'argument . . .), les éléments (objectifs et subjectifs) qui déterminent ce choix de même que la complexité algorithmique de son calcul. Remontant aux premiers systèmes experts (ES⁴) dotés de module d'explication, notre travail tentera de couvrir une large palette de situations de décision dans lesquelles la production d'une explication est requise : justification des recommandations de ES [Kass and Finin, 1988], DSS [Papamichail and French, 2003], RS⁵ [Daher et al., 2017], d'une simple comparaison par paire d'alternatives [Nunes et al., 2014], de la déduction d'un modèle de préférences [Kadzinski et al., 2014], du résultat de l'application d'une règle de vote [Boixel et al., 2022] ou d'une procédure d'agrégation de préférences [Labreuche et al., 2011].

1.2.1 Explication axée modèle d'utilisateur et/ou de décision

Le terme *Explication* recouvre une large diversité de constructions retraçant le raisonnement interne d'un système ou les étapes d'une méthode ayant permis d'aboutir à une conclusion (à expliquer). L'objectif de cette sous-section est de rendre compte

3. Traduction personnelle de *scientific explanations*.

4. *ES* : *Expert System*.

5. *RS* : *Recommender System*.

de cette diversité en distinguant les travaux portant sur la production d'explications conçues comme des preuves (axées modèle de décision) de ceux qui mettent l'accent sur la nécessité de produire des explications mimant le mode de raisonnement des humains et/ou tenant compte de leur niveau de connaissances, de leurs croyances, préférences et de leurs besoins en termes d'explication (axées modèle d'utilisateur). Nous verrons que la frontière entre ces deux catégories est loin d'être claire : certaines explications axées modèle d'utilisateur présentent des spécificités relatives au modèle de décision duquel découle la conclusion à expliquer et on note, au niveau des propositions d'explications-preuves, le souci de les structurer de sorte qu'elles soient intelligibles pour un décideur humain et tiennent compte de ses potentielles (ou réelles) limites cognitives.

Dans des situations de décision critiques ou à fort enjeu, il semble légitime qu'un décideur exige une explication qui rende fidèlement compte du raisonnement ou de la méthode ayant permis d'aboutir à la recommandation qui lui est faite. En effet, la décision à prendre engage sa responsabilité et il est impératif qu'il en comprenne tous les ressorts et confronte ceux-ci à son propre système de valeurs et de connaissances. Une méthode standard permettant d'appréhender le raisonnement interne d'un système automatique de simulation ou de recommandation, consiste à analyser les différences entre les sorties du système à partir d'au moins deux jeux de données contrôlés fournis en entrées. C'est ce que permet de faire le module d'explication EXPLAIN qui génère (sous la forme de tableaux et de données textuelles) une interprétation qualitative des résultats de simulation de productions agricoles contenant toutes les informations nécessaires à la génération d'une explication "scientifique" [Greer et al., 1994]. En effet, cet outil permet d'analyser les facteurs responsables de la différence des rendements en grains suivant deux scénarios de simulation et de restituer, sous une forme personnalisée, les résultats d'une telle analyse en langage naturel. De même, le *Fine Expert System* (FES) développé par [Papamichail and French, 2003] génère, sur commande de l'utilisateur et à partir de techniques de génération de langage naturel⁶, différents types d'explication sous la forme d'un rapport expliquant la comparaison de paire de stratégies (ensemble de contre-mesures) à mettre en place en cas d'accident radiologique en Europe ; chaque alternative étant évaluée suivant une fonction linéaire multivariée. [Klein, 1994, Carenini and Moore, 2006] se sont également intéressés à la génération d'explication de comparaison par paire d'alternatives évaluées suivant une fonction de score multivariée. À la différence de l'approche précédente, le processus de génération d'explication est divisé en deux parties distinctes : la sélection des arguments (*deep generation*) à présenter et leur expression en langage naturel (*surface generation*). Au sujet de la phase de sélection des arguments, [Labreuche, 2011] fait remarquer qu'en plus de ne pas mentionner les poids des critères – essentiels dans ce type de modèle –, les deux précédentes approches ne fournissent aucune justification formelle du choix de leurs arguments. Sa proposition, qui ne traite pas de l'expression de l'explication en langage naturel, tente de remédier à

6. En anglais, *Natural Language Generation (NLG) techniques*.

ces “inconvenients” en évoquant la nécessité de réaliser ce choix vis-à-vis d’un éventail de quatre *ancres* qui sont des types de raisonnement adaptables à la complexité de la situation de décision à expliquer. Le choix des arguments y est réalisé en tenant compte des spécificités du modèle de décision à partir duquel les alternatives sont comparées. Aux *ancres* de [Labreuche, 2011], [Nunes et al., 2014] opposent des *patterns d’explication* qu’ils estiment plus en phase avec « les formes d’explication dont les utilisateurs ont besoin ». En effet, affirment-ils que le choix (respectivement le rejet) d’une alternative par rapport (respectivement au profit) d’une autre ne dépend pas uniquement d’une fonction de score multivariée mais aussi de l’application de deux principes de psychologie relatifs à la façon dont les humains prennent leurs décisions : *l’aversion pour les extrêmes et le contraste entre les compromis* [Simonson and Tversky, 1992]. S’appuyant sur les résultats d’une étude (cf. [Nunes et al., 2012]) réalisée auprès de 100 enquêtés et destinée à comprendre comment les utilisateurs de systèmes de recommandation justifient leur choix (d’hôtel en l’occurrence), [Nunes et al., 2014] ont défini sept *patterns d’explication* ordonnés suivant leur niveau de complexité et auxquels sont associés des modèles de textes (à trous). Ces *patterns d’explication* sont conformes à des directives formulées par [Nunes et al., 2012] et qui sont censées servir de base au développement de nouvelles approches d’explication.

Dans la quasi-totalité des travaux mentionnés précédemment, on remarque que l’Explication n’intervient (après-coup) que pour justifier une recommandation faite à partir d’un modèle de décision. Il existe cependant des situations de décision où l’on peut considérer que l’Explication est “consubstantielle” à la prise de décision. C’est le cas par exemple en argumentation [Amgoud and Prade, 2009] ou lorsque le raisonnement est conduit à l’aide d’outils graphiques de représentation (qualitative) de préférences tels que les CP-nets [Boutilier et al., 2004]. C’est aussi le cas lorsque la décision est résultat de l’application d’une procédure d’agrégation de vote [Boixel et al., 2022] en Théorie du Choix Social [Zwicker and Moulin, 2016] ou d’une procédure d’agrégation de préférences multicritères (e.g. la méthode de Condorcet pondérée [Labreuche et al., 2011]). Dans ces cas, les efforts se concentrent plutôt sur la représentation de l’explication : calcul d’une explication minimale (au sens de l’inclusion ou suivant une fonction de coût) et “lisible” pour un décideur humain. Ça pourrait être aussi le cas du processus interactif d’éllicitation *even swaps Method* de [Hammond et al., 1998] ; l’explication consisterait alors à retracer l’ensemble des compromis réalisés tout au long du processus ; le choix de l’alternative retenue à son issue se déduisant de cet ensemble de préférences (indifférences) exprimées, à chaque fois, sur exactement deux critères. La *even swaps Method* de [Hammond et al., 1998] a inspiré les travaux de [Belahcène et al., 2017] qui portent sur l’explication des déductions du calcul de la relation nécessaire [Greco et al., 2008] à l’aide de *preference swaps*. Cette proposition repose sur l’hypothèse de la simplicité cognitive (pour tout décideur) des *preference swaps* et de l’intelligibilité des propriétés du modèle additif (la transitivité et l’indépendance d’ordre 1 en l’occurrence) qui, appliquées à ceux-ci conduisent à la déduction de la comparaison par paire à expliquer.

Les travaux de [Belahcène et al., 2017] ont été d'un grand intérêt pour nous car ils assimilent l'explication à une opération de déduction de la préférence à expliquer à partir d'autres préférences "atomiques" mettant en évidence les propriétés/axiomes du modèle de décision sous-jacent qui entrent en jeu dans la déduction. Dans notre travail, l'explication s'effectue conformément à des schémas déductifs s'appliquant à des comparaisons par paire considérées comme admises. Une autre proposition d'explication de Belahcène *et al.* fondée sur les propriétés du modèle additif est [Belahcène et al., 2019]. À la différence de l'approche basée sur les *preference swaps*, le "pouvoir explicatif" de cette proposition est total car les formes d'explication qui y sont décrites s'appuient sur la caractérisation des relations de préférences représentables par un modèle additif. Dans une optique de facilitation de l'interaction entre un analyste et un décideur et pour permettre la révision des préférences de ce dernier, [Kadzinski et al., 2014] se sont aussi intéressés à l'explication des déductions du calcul de la relation nécessaire [Greco et al., 2008]. Toujours dans le cadre de l'AMCD et dans une perspective d'élicitation, [Greco et al., 2013] proposent, en s'appuyant sur les outils de la Théorie des ensembles approximatifs de [Pawlak, 1991] (l'induction de règles à partir d'exemples), une explication de comparaisons par paire d'alternatives à base de *règles de décision* de la forme (*si ... alors ...*).

1.2.2 Représentation de l'explication

L'objectif de cette sous-section est de tenter de faire l'inventaire des différentes représentations d'explication que l'on retrouve dans la littérature et de mettre en évidence, le cas échéant, les similarités en ces dernières et les structures d'explication que nous adopterons dans ce travail.

Le destinataire de l'explication étant un décideur ou un analyste humain, son support de délivrance est généralement le langage naturel. Les modules d'explication équipant les systèmes experts et d'aide à la décision produisent ainsi des explications textuelles qui sont transmises à leurs utilisateurs sous des formats divers et par le biais d'interfaces telles que les navigateurs Web. En fonction du domaine d'application, certaines explications, tenant compte du profil de l'utilisateur (expert, novice ...), sont "personnalisées". C'est le cas par exemple des modules EXPLAIN [Greer et al., 1994] et VOTE [Slade, 1991] que nous avons présentés précédemment. Ces travaux sont en accord avec ceux de [Kass and Finin, 1988] qui insistent sur la nécessité de disposer d'un modèle d'utilisateur pour pouvoir produire de "bonnes" explications. Ils décrivent à cet effet, le système GUMAC qui est un modèle d'acquisition des croyances et connaissances des utilisateurs de systèmes experts. En ce qui concerne les explications du système FES (présenté précédemment) [Papamichail and French, 2003], elles sont produites, sous la forme de rapports écrits, à l'aide de techniques de génération de langage naturel (NLG) suivant trois étapes distinctes : (i) la détermination du contenu, (ii) la planification du discours et (iii) la génération des phrases. Lesdits rapports peuvent comporter des gra-

phiques et des tables accompagnés de leur interprétation. C'est typiquement le cas du rapport d'analyse de sensibilité sur la valeur d'un critère. [Wulf and Bertsch, 2017] ont développé un système similaire au FES, qui génère, suivant la même approche, des rapports d'analyse de sensibilité portant sur plusieurs critères. Il faut dire que certaines explications produites par ces systèmes sont des modèles de textes à trous qui sont remplis par des chaînes de caractères (quantificateurs sémantiques, noms de critère ...) ou des valeurs numériques calculées par ces mêmes systèmes et non importées des DSS qu'ils équipent en qualité de modules d'explication. De même, [Nunes et al., 2014] associent à chaque *pattern d'explication* un modèle de texte à trous à remplir à l'aide d'un unique attribut ou d'un ensemble d'attributs. Au sujet des formes de représentation des explications produites par des systèmes automatiques, [Cawsey, 1995] considère que certains utilisateurs pourraient préférer aux explications textuelles (conformes aux conventions de dialogue entre humains) des représentations graphiques. **Les explications que nous proposons dans ce travail, construites conformément à des schémas déductifs peuvent être aussi vues comme des structures (graphiques) à trous à remplir avec des comparaisons par paire d'alternatives ; lesdites structures explicitant les "mécanismes" en jeu dans la déduction de la comparaison par paire à expliquer.** À l'inverse, les deux types d'explication que proposent [Kadzinski et al., 2014] sont tout simplement un ensemble (minimal ou maximal au sens de l'inclusion) d'affirmations de préférences exprimées par le décideur. Les explications de [Belahcène et al., 2017] sont une séquence de *preference swaps* ; une représentation qui se prête bien à une expression en langage naturel [Belahcène, 2018] contrairement aux *syntactic and elliptic cancellative explanations* de [Belahcène et al., 2019] représentées sous la forme d'une "somme" de plusieurs comparaisons par paire. Comme indiqué dans la précédente sous-section, [Greco et al., 2013] proposent des explications de type *règles de décision* se présentant sous la forme « Si E alors pwc » où E est un ensemble de conditions sur les critères et pwc la comparaison par paire à expliquer. **Les explications que nous proposons dans ce travail, de par leur nature déductive, peuvent également être vues sous la forme de règles de décision du type « Par application du schéma sch , si E alors pwc » qui précisent par ailleurs le schéma de déduction utilisé.** Seulement, à la différence des explications de [Greco et al., 2013], nos explications sont des preuves de la déduction de la comparaison par paire à expliquer. En effet, les schémas de déduction que nous considérerons sont des instanciations des propriétés/axiomes du modèle de décision considéré. D'autres travaux portant sur la production d'explications-preuves ont défini des *langages* "techniques" devant servir de passerelle à leur expression en langage naturel. À titre d'exemples, nous pouvons mentionner les *simple* et *factored languages* de [Labreuche et al., 2011] composés de comparaisons par paire d'options sur un ou plusieurs critères ou encore les différents langages "imbriqués" de comparaisons par paire atomiques décrits par [Amoussou et al., 2022] dans des situations de décision où l'information préférentielle à disposition est complète. Lorsque ce n'est pas le cas, le calcul de l'explication peut nécessiter le recours à des outils d'optimisation de la Recherche Opé-

rationnelle conduisant, par exemple, à représenter l'explication sous la forme de l'ensemble des coefficients significatifs obtenus après exécution de programmes linéaires en nombres entiers (PLNE) [Labreuche et al., 2012]. **La mise en œuvre de nombre de schémas déductifs que nous considérerons dans ce travail requièrent la résolution de PLNE.** Cependant, comme indiqué précédemment, les explications correspondantes seront une traduction des résultats numériques obtenus après résolution des PLNE en structure d'explication conforme au schéma déductif considéré et explicitant les propriétés/axiomes du modèle de décision sous-jacent. Dans le même ordre d'idée, [Boixel and Endriss, 2020] décrivent des algorithmes de Programmation par contraintes calculant une explication de l'issue d'une procédure de vote. Ces algorithmes calculant des MUS (Minimal Unsatisfiable Subsets) [Liffiton and Sakallah, 2008] correspondant à des instances d'axiomes justifiant l'élection d'un ou plusieurs candidats. Enfin, il convient de signaler que les approches d'explication de préférences fondées sur des axiomes ne sont pas récentes. En effet, [Kim and Kim, 1991], rappelant un ensemble de cinq axiomes de la théorie de l'utilité espérée en décision dans l'incertain [Raiffa, 1970], démontrent qu'il est possible de fournir des explications personnalisées basées sur un modèle de décision théorique sans requérir de l'utilisateur qu'il ait une compréhension explicite des aspects mathématiques et techniques de la Théorie de la Décision. Ces cinq axiomes (*la comparabilité, la solvabilité, la substituabilité, la monotonie et la décomposabilité*) sont utilisés, au sein du système DETES, pour justifier une recommandation déduite d'un arbre de décision. **Dans notre travail qui porte sur l'explication en décision multicritère, nous avons construit différents schémas déductifs autour de trois propriétés/axiomes (*transitivité, congruence et couverture*) et pour lesquels nous avons proposé quelques algorithmes de calcul.** Dans la prochaine section, nous détaillerons plus amplement notre approche. Mais avant, achevons notre brève revue de littérature en évoquant d'autres considérations pertinentes en lien avec l'explication en décisions multicritère et collective.

1.2.3 Autres considérations : Interaction, Inexplicabilité, Complexité

Dans cette dernière sous-section de revue de littérature, nous nous proposons d'examiner comment la question de l'explication a été traitée au regard des trois thèmes que sont : l'interaction, l'inexplicabilité et la complexité (algorithmique). Le choix de ces trois thèmes se justifie par le fait qu'ils constituent des considérations pertinentes pour notre travail. À travers le thème de l'interaction, nous allons mettre en lumière les formes d'explication qui se prêtent à l'expression de réactions (de la part du destinataire) pouvant être traitées sans ambiguïté. Autour du thème de l'inexplicabilité, nous tenterons d'attirer l'attention sur le fait que l'explication d'une recommandation est construite relativement à un ensemble de présupposés et/ou à partir de données précises et sous une ou plusieurs hypothèses. C'est cet ensemble d'éléments que nous désignerons sous le vocable de *contexte* (Section 1.3) : contexte dans lequel l'explication est générée ; ceci

permettant de remarquer qu'en fonction du *contexte*, une décision peut être explicable ou non. Enfin, puisque le calcul de l'explication est délégué à des algorithmes, nous ne pouvons terminer sans évoquer la question de la complexité computationnelle des problèmes sous-jacents.

Justifier les recommandations automatiques (de systèmes de recommandation (RS)) à l'utilisateur n'est que la moitié de la solution, la seconde moitié étant de rendre le système contrôlable en permettant à l'utilisateur d'y apporter des modifications [Tintarev and Masthoff, 2007]. Ce "contrôle" du RS peut être rendu effectif par des *feed-backs* (spécification des besoins, modification d'anciennes notes...) de l'utilisateur. Ces *feed-backs* ont pour effet d'actualiser la connaissance dont dispose le RS des préférences du décideur et peuvent permettre, comme le démontrent [Labreuche et al., 2015], d'écarter des modèles de décision incompatibles. Dans les situations où le type de modèle de décision est "fixé", certains éléments d'explication peuvent amener l'utilisateur à en modifier les valeurs. C'est par exemple le cas des explications produites par le système FES [Papa-michail and French, 2003] décrivant les paramètres de la fonction linéaire multivariée d'évaluation des alternatives. En Aide Multicritère à la Décision, les *preferential reducts* et *preferential constructs* de [Kadzinski et al., 2014] sont deux types d'explication reliant un ensemble de déclarations de préférences du décideur à une conclusion et qui, de ce fait, se prêtent bien à un traitement sans ambiguïté des réactions de celui-ci. En effet, le décideur peut confirmer ses précédentes déclarations de préférences ou au contraire les réviser lorsqu'elles ne lui semblent pas en adéquation avec la conclusion à laquelle elles sont associées. Les explications proposées par [Greco et al., 2013] sous la forme de règles de décision, quand bien même elles ne sont pas exprimées dans "les mêmes termes" que les données de préférence collectées auprès du décideur (des comparaisons par paire d'alternatives de référence), sont, du fait de leur simplicité, facilitatrices d'interaction. Le décideur peut réagir à ces explications en validant ou modifiant les prémisses desdites règles qui sont des conditions sur les critères; ce qui contribuera indirectement à l'élicitation de ses préférences. **Dans notre travail, nous nous sommes également intéressé à la fonction d'élicitation que peut remplir l'explication lorsque celle-ci fait intervenir, outre les déclarations de préférences du décideur, des comparaisons par paire de critères appelées *preference swaps*.** A contrario, l'utilisation de langage mentionnant explicitement les coefficients significatifs résultant de la résolution de programme linéaire en nombres entiers, pour expliquer pourquoi une coalition de critères est gagnante dans [Labreuche et al., 2012], nous semble réduire considérablement, sinon totalement les possibilités d'interaction du fait de la difficulté d'interprétation des valeurs de tels coefficients, pour un décideur quelconque.

Nous traiterons du thème de l'inexplicabilité en nous référant principalement aux travaux de [Labreuche, 2011], [Labreuche et al., 2012] et [Nunes et al., 2014] qui tous, d'une manière ou d'une autre, font état de différents niveaux de complexité des situations de décision à expliquer et que peuvent "capturer" certaines formes d'explication et pas d'autres. [Labreuche, 2011] définit quatre *ancres* qui correspondent à différents

modes de raisonnement dont la prise en compte est nécessaire pour pouvoir couvrir l'ensemble de ces différents niveaux de complexité. Ces *ancres* sont ordonnées suivant leur « complexité intrinsèque »⁷ allant des situations où la décision est triviale (dominance de Pareto) à celles où elle est plus serrée (inextricable). Le choix du nom “**remaining**” de la dernière *ancree* dans cet ordre, est à ce titre, assez éloquent. Suivant la même logique et rappelant la directive « *keep it simple* » formulée dans [Nunes et al., 2012], [Nunes et al., 2014] décrit de façon informelle l'ordre dans lequel il convient de considérer leurs sept *patterns d'explication*. Enfin, [Labreuche et al., 2012] distingue, suivant le niveau de détails que peuvent nécessiter les explications correspondantes, trois sous-ensembles d'options dominées par une option donnée du *Smith set*[Fishburn, 1977] dont celui des options dominées à une *faible majorité* (*weak majority*) et pour lesquelles la décision n'est pas évidente et peut requérir une explication détaillée. Suivant l'option dominée de ce sous-ensemble, l'explication peut être plus ou moins complexe. Quatre niveaux de complexité ont été retenus ; ce qui sous-entend qu'en s'imposant un niveau de complexité donné, une explication de la dominance pourrait ne pas exister. En Aide MultiCritère à la Décision, les *explications questionnables* de [Amoussou et al., 2023a] sont des explications à base de *preference-swaps* qui ne sont “valides” que pour un sous-ensemble de fonctions de score compatibles avec les préférences du décideur. En Choix Social Computationnel, en fonction de la base normative d'axiomes considérée [Boixel and Endriss, 2020], la justification de l'élection d'un ou plusieurs candidats peut ne pas être possible. **Toutes ces considérations nous ont conduit dans notre travail à la définition du “contexte” dans lequel l'explication sera produite et à proposer des algorithmes de calcul d'explication en fonction de la spécificité de chaque contexte ; une décision pouvant dès lors admettre une explication dans un contexte et pas dans un autre.**

Sans risque de se tromper, on peut affirmer que les problèmes consistant à vérifier l'existence de l'explication d'une décision (collective ou multicritère) sont en général des problèmes (algorithmiquement) difficiles, notamment en raison des outils que mobilise leur résolution. Nous avons précédemment évoqué les algorithmes de calcul de MUS [Liffiton and Sakallah, 2008] dans le cadre de la justification de l'élection d'un ou plusieurs candidats suivant l'approche proposée par [Boixel and Endriss, 2020]. Au sujet de cette contribution, [Boixel and de Haan, 2021] réalisent deux études de complexité : celle du problème consistant à vérifier si une justification donnée est correcte et du problème consistant à vérifier si une justification existe. En décision multicritère, les explications sous forme de règles de décision de [Greco et al., 2013] s'obtiennent à partir d'algorithmes d'induction de règles qui, « ...du point de vue calcul sont complexes »⁸. La plupart des propositions d'explications “minimales” précédemment mentionnées sont accompagnées d'une analyse de complexité. Les problèmes de décision (algorithmique)

7. Traduction personnelle.

8. Traduction personnelle.

associés sont généralement NP-complets : [Kadzinski et al., 2014] (en particulier les *preferential constructs*), les *elliptic cancellative explanations* de [Belahcène et al., 2019], les explications composées d’« éléments factorisés » de [Labreuche et al., 2011] (lorsque la fonction de coût considérée n’est pas super-additive) et celles proposées par [Labreuche et al., 2012] justifiant une dominance à faible et large majorités. Il convient d’indiquer que l’outil utilisé pour calculer ces différentes formes d’explication est la programmation linéaire en nombres entiers (PLNE). **Nombre des algorithmes de calcul d’explication que nous présenterons dans ce travail reposent également sur la résolution de PLNE. Nous avons pu montrer que certains problèmes de décision (algorithmique) associés sont NP-complets et que d’autres appartiennent à la classe P.** Contrairement aux *preferential constructs*, les *preferential reducts* de [Kadzinski et al., 2014] peuvent être calculées en temps polynomial. Il en est de même des explications à base de *preference swaps* de [Belahcène et al., 2017] et des explications “minimales” de [Labreuche et al., 2011] composées uniquement de comparaisons par paire d’options « non factorisées » (*simple language*).

1.3 Conclusion et plan du manuscrit

Dans l’introduction de ce chapitre, nous avons décrit, dans ses grandes lignes, ce qu’est qu’un processus d’aide à la décision multicritère. Nous avons montré qu’il s’agit d’un processus interactif entre au moins deux acteurs (un décideur et un analyste) dans lequel l’explication pourrait jouer plusieurs rôles bénéfiques; bénéfiques au décideur en lui permettant de comprendre “ce que fait l’analyste”, bénéfiques à l’analyste en lui permettant de saisir “à quelle forme de rationalité” obéissent les préférences du décideur et donc bénéfiques au processus tout entier.

Dans la Section 1.2, nous avons présenté, sous divers angles de vue, les différentes approches de l’explication de décision proposées dans la littérature. Les domaines de la Théorie de la Décision [Luce and Raiffa, 1957] couverts ont principalement été ceux de la décision multicritère et de la décision collective. En fin de sous-section 2.3.1, nous avons fait mention d’un article ([Kim and Kim, 1991]) proposant une approche axiomatique de l’explication en décision dans l’incertain. En décision collective, les travaux de [Boixel and Endriss, 2020] s’inscrivent dans la même logique : justifier l’issue d’une procédure de vote à partir d’axiomes. **La motivation de notre travail est de tenter d’adopter une approche similaire en décision multicritère adaptée aux problématiques de choix et de rangement.**

La recommandation de type choix ou rangement résulte de la construction d’une relation (binaire) de préférences sur un ensemble (généralement fini) d’alternatives [Labreuche, 2011]. Il existe une large variété de modèles de préférences permettant de représenter de manière plus ou moins compacte de telles relations de préférences. **Nous développerons notre approche de l’explication en décision multicritère autour du**

modèle additif [Fishburn, 1967] qui est l'un des modèles de préférences les plus compacts. Ce faisant, nos contributions complètent les travaux de [Belahcène et al., 2017] et [Belahcène et al., 2019]. Il convient cependant d'indiquer que notre approche se veut générique dans le sens où elle pourrait s'appliquer à toute classe de modèles de préférence disposant d'un système d'axiomes.

Dans la sous-section 1.3.1, nous précisons les grandes lignes de notre approche de l'explication construite autour du concept de *contexte* et tenterons d'en justifier l'intérêt. Nous clôturerons ensuite le chapitre en détaillant le plan du manuscrit (sous-section 1.3.2).

1.3.1 De la nécessité d'une nouvelle approche de l'explication en AMCD

Au regard des problématiques de décision multicritère (choix et rangement) auxquelles nous nous intéressons dans ce travail et qui reposent sur une évaluation relative des alternatives (les unes par rapport aux autres), nous pouvons dire que celui-ci se ramène à tenter d'apporter des éléments de réponse à la question :

Comment expliquer que x est préférée à y ?

En introduction de ce chapitre, nous avons inventorié un petit nombre de "moments" du processus d'aide à la décision où la production d'une explication de la préférence de x sur y est souhaitable quelle que soit de la classe de modèles utilisée pour représenter les préférences du décideur. En introduction de cette section, nous avons posé l'hypothèse que celles-ci sont représentables par un modèle additif, ce qui signifie qu'il existe une fonction de valeur multivariée qui associe à x un score supérieur à celui de y et qui se calcule en sommant les évaluations marginales de ces deux alternatives sur l'ensemble des critères considérés. Faire mention de cette fonction de valeur dans l'explication de la préférence (x, y) (comme le suggère [Labreuche, 2011]) nous semble inadaptée aux contextes de délivrance de l'explication que nous considérons ici. En effet, les explications basées sur les *ancres* de [Labreuche, 2011] sont destinées à des « acteurs peu intéressés de la technicité des modèles de décision utilisés et désireux d'avoir une explication synthétiques de la recommandation ». Par ailleurs, *une telle fonction de valeur n'est que représentation*, « un thermomètre de préférence » selon les mots de [Kim and Kim, 1991] qui rappellent que les valeurs numériques d'utilité n'ont aucune signification par elles-mêmes ; ce que ne contrediraient pas [Bouyssou et al., 2006b, Chapitre 3]. Cette fonction qui peut ne pas être unique, représente une relation binaire de préférences c'est-à-dire un ensemble de comparaisons par paire *consistant* du point de vue des axiomes du modèle additif.

Dans notre approche, une préférence s'expliquera (par déduction) à partir d'une ou plusieurs préférence(s), une comparaison par paire à partir d'une ou plusieurs comparaison(s) par paire comme le proposent [Kadzinski et al., 2014]. Ce faire semble cohérent

avec la *rationalité a posteriori* de [March, 1978] et justifier l'intérêt sans cesse croissant accordé au paradigme de désagrégation de préférences où le décideur est invité à fournir ses préférences sous la forme de comparaisons par paire d'alternatives de référence. Comme annoncé précédemment, notre approche sera construite autour du concept de *contexte* qui rassemble :

- Une relation \mathbb{R} de préférences représentée ou non par une fonction de valeur multivariée et à laquelle appartient (x, y) ;
- Un sous-ensemble \mathbb{B} de comparaisons par paire de la relation \mathbb{R} ($\mathbb{B} \subset \mathbb{R}$) ;
- Un schéma déductif *sch* décrivant précisément une règle de déduction s'appuyant sur les propriétés/axiomes du modèle additif.

L'explication de (x, y) dans un *contexte* $(\mathbb{R}, \mathbb{B}, sch)$ donné reviendra donc à déterminer les éléments de \mathbb{B} auxquels l'application de la règle décrite par *sch* conduit à la déduction de (x, y) .

Suivant le paradigme de désagrégation de préférences, l'ensemble \mathbb{B} peut être l'ensemble des comparaisons par paire fournies par le décideur [Kadzinski et al., 2014, Belahcène et al., 2019]. Il peut s'agir aussi de comparaisons par paire supposément "élémentaires" ou "atomiques" comme les *preference swaps* [Belahcène et al., 2017]. Le schéma déductif *sch* peut être tout simplement l'application de la propriété de transitivité ou d'indépendance ou encore une combinaison plus ou moins complexe de ces deux propriétés bien connues.

À l'aide de la définition du *contexte*, on peut attribuer à l'explication en AMCD trois fonctions distinctes à savoir :

- Une fonction de justification d'une recommandation ou d'une conclusion robuste [Roy, 2009],
- Une fonction (liée à la première) de résolution d'inconsistance revenant à justifier deux décisions contradictoires,
- Une fonction d'élicitation de préférences.

Ce sont ces différentes fonctions que nous tenterons d'illustrer tout au long de ce document de travail dont les détails du plan sont précisés dans la prochaine sous-section.

1.3.2 Plan du manuscrit

La suite de ce document est composée de deux parties.

La première partie, s'étalant sur deux chapitres, servira à présenter notre approche d'explication-preuve en utilisant comme modèles de préférences les Relations d'Ordre Linéaire Additives (ROLA) [Fishburn, 1996, Fishburn et al., 2002]. Elle sera centrée autour des schémas déductifs. Son premier chapitre (Chapitre 2) déroulera l'ensemble des étapes conduisant à leur définition. À cet effet, il introduira différents concepts qui seront

définis et illustrés. Seront aussi présentées les différentes structures de représentation de l'explication.

Le Chapitre 3 va, quant à lui, proposer quelques algorithmes de calcul d'explication conformes aux schémas déductifs décrits au Chapitre 2. Leur expressivité en termes de taux d'inexplicabilité global sera expérimentalement mesurée et de façon exhaustive. Les analyses des dites expérimentations numériques permettront d'une certaine façon d'appréhender la diversité des raisonnements cohérents avec la classe de modèles de préférences représentables par des ROLA. Ensuite, à travers la procédure d'*évolution dynamique du contexte*, il nous sera possible, comme l'a proposé [Boixel and Endriss, 2020, Chapitre 4] en décision collective, de mesurer le pouvoir explicatif de propriétés/axiomes des ROLA.

La deuxième partie sera composée de deux chapitres. Le premier (Chapitre 4) va tenter d'illustrer les fonctions de justification et d'élicitation que peut remplir l'explication en Aide MultiCritère à la Décision lorsque l'on tient compte de l'ensemble de toutes les fonctions de valeur compatibles avec les préférences du décideur [Greco et al., 2008, Kadzinski et al., 2014, Greco et al., 2013, Belahcène et al., 2017, Belahcène et al., 2019, Labreuche et al., 2012]. Nous nous intéresserons dans un premier temps à la fonction de justification de l'explication de conclusions robustes [Roy, 2009]. Le modèle de préférences considéré est le modèle additif défini sur une échelle binaire de critères (comme dans [Belahcène et al., 2017]). Les expérimentations numériques qui seront conduites dans ce cadre, en plus de compléter les résultats d'expressivité du chapitre précédent, vont tenter de les comparer à ceux de l'état de l'art : les explications basées sur les *preference swaps* de [Belahcène et al., 2017]. Dans la deuxième partie du Chapitre 4, notre principale contribution sera la proposition d'une ébauche de protocole interactif à travers lequel l'explication peut pleinement remplir sa fonction d'élicitation. Cette proposition sera illustrée en détails et les outils algorithmiques sur lesquels elle repose seront présentés.

Dans le deuxième chapitre de la Partie II (Chapitre 5), nous nous intéresserons à l'explication d'une recommandation dans le cas "précis" : en ne considérant qu'une unique fonction de valeur compatible avec les préférences du décideur. La multiplicité des critères (conflictuels) étant source de complexité en décision multicritère, les formes d'explication que nous considérerons dans cette partie consisteront à décomposer suivant différents *patterns*, les comparaisons par paire justifiant la recommandation en sous-comparaisons "élémentaires". Une analyse de complexité du problème de décision associé à de telles décompositions est proposée de même que des mesures de traitement de l'inexplicabilité de la recommandation. Ce chapitre comporte également des expérimentations numériques qui, du point de vue des *patterns de décomposition* considérés, nous offrent d'une certaine façon une description des caractéristiques de l'alternative faisant l'objet de la recommandation. Enfin, à l'image des *factored language* de [Labreuche et al., 2011], des tentatives de factorisation de l'explication de la recommandation sont envisagées ; ceci, dans une optique de minimisation de l'effort cognitif [Miller, 1956] requis par

le traitement des potentiels multiples *chunks* d'explication.

Première partie

Schémas déductifs pour l'explication de préférences additives

Chapitre 2

Explication dans le cadre du modèle additif : principes et définitions

Résumé

L'objet du chapitre est de montrer comment expliquer l'appartenance d'une comparaison par paire à une relation binaire à partir d'autres comparaisons par paire. Suivant cette approche, l'explication est une preuve de la déduction de la comparaison par paire expliquée. Cette preuve s'appuie sur certaines propriétés de la relation binaire considérée pouvant correspondre à des modes de raisonnement d'un décideur humain. Les propriétés retenues dans ce chapitre sont celles de transitivité, d'indépendance et de couverture des relations d'ordre linéaire additives.

Ce chapitre présente la construction de schémas déductifs sur la base des précédentes propriétés. Deux grandes catégories de schémas déductifs ont été définies de façon formelle : les schémas de base qui mettent en œuvre un unique mécanisme de déduction et les schémas déductifs composés qui combinent plusieurs mécanismes de déduction. Le chapitre comporte en outre plusieurs exemples et illustrations d'explication qui sont représentées de façon équivalente sous trois différentes formes.

Sommaire

| | | |
|------------|--|-----------|
| 2.1 | Introduction | 42 |
| 2.2 | Principes décisionnels associés au modèle additif | 43 |
| 2.2.1 | Propriétés fondamentales et dérivées | 43 |
| 2.2.2 | Mécanismes de déduction associés | 46 |
| 2.3 | Formalisation de l'explication pour le modèle additif | 51 |
| 2.3.1 | Exemple introductif | 52 |
| 2.3.2 | Schémas déductifs | 54 |
| 2.3.3 | Retour sur l'exemple introductif | 57 |
| 2.4 | Conclusion et perspectives | 62 |

2.1 Introduction

Ce chapitre a pour objectif de poser les bases d'une nouvelle approche de génération de l'explication d'une comparaison par paire d'alternatives décrites sur plusieurs critères. Les explications de préférences que nous proposerons dans ce cadre sont des *preuves* de la déduction de la comparaison par paire expliquée. Leur production se prête bien aux situations de décision dans lesquelles on dispose d'un historique ou d'une base de préférences telle que l'ensemble des affirmations de préférences d'un décideur sur des exemples d'alternatives qu'il connaît bien [Greco et al., 2008]. L'objet de l'explication, c'est la déduction de nouvelles comparaisons par paire à partir de cette base de préférences. L'explication, en tant que preuve de cette déduction, vient donc mettre en lumière les caractéristiques du modèle rationnel de préférences censé représenter les préférences globales du décideur. Ces caractéristiques prennent généralement la forme d'un système d'axiomes édictant un jeu de conditions nécessaires et suffisantes sous lesquelles se justifie cette représentabilité.

Dans ce chapitre, nous considérerons des préférences exprimées sur des ensembles d'items. Nous supposerons qu'elles sont représentables par un modèle additif. Plus précisément, nous considérerons que les préférences du décideur forment une relation d'ordre linéaire additive [Fishburn et al., 2002]. L'objectif du chapitre est de montrer **comment** expliquer l'appartenance d'une comparaison par paire à une telle relation à partir d'autres comparaisons par paire de la même relation. Il est organisé de façon à permettre au lecteur d'appréhender tous les ressorts de notre proposition d'approche de l'explication. En effet, dans la Section 2.2, partant de la définition axiomatique des relations d'ordre linéaire additives, nous dégagerons un ensemble de propriétés qui serviront de base à la définition de *mécanismes de déduction*. Dans la Section 2.3, nous montrerons comment l'explication peut être assimilée à un processus récursif mettant en œuvre plusieurs mécanismes de déduction à partir de comparaisons par paire jugées *admises* pour le décideur. Pour ce faire, nous définirons formellement (sous-section 2.3.2) différents *schémas déductifs* suivant lesquels seront construites les explications de comparaisons par paire déduites. L'ensemble composé de la relation d'ordre linéaire additive représentant les préférences globales du décideur, du sous-ensemble des comparaisons par paire admises et d'un schéma déductif constituera le *contexte* dans lequel sera délivrée l'explication et dont le support pourra indifféremment être le langage naturel, une représentation sous forme de diagramme ou encore un enchaînement d'implications logiques.

De par son contenu, ce chapitre de contribution constitue une version étendue d'un travail ([Amoussou et al., 2022]) soumis et présenté à l'occasion des 16^{èmes} Journées d'Intelligence Artificielle Fondamentale (JIAF 2022).

2.2 Principes décisionnels associés au modèle additif

Sous le terme de *Coherent Boolean Term Order*, [Maclagan, 1998] définit la relation d'ordre linéaire additive \succ sur $2^{[m]}$ comme celle vérifiant :

$$x \succ y \iff \sum_{i \in x} \omega_i > \sum_{j \in y} \omega_j \quad (2.1)$$

où ω est un vecteur de m composantes entières : $\omega = (\omega_i)_{i \in [m]} \in \mathbb{N}^m$ et où $[m]$ désigne un ensemble de m items, individus, critères ...

Une relation d'ordre linéaire additive (ROLA) est une relation d'ordre strict et total représentée par une séquence d'entiers naturels dont chaque sous-séquence est, de façon unique, déterminée par la somme de ses éléments (*Subset-Sum-Distinct Sequence*) [Lunnon, 1988, Bohman, 1997]. Elle peut servir à la modélisation des préférences d'un décideur sur des ensembles d'items [Fishburn et al., 2002]¹ comme dans le problème de *subset selection* [Rooij et al., 2005]. Elle peut également servir à la modélisation des préférences d'agents dans des problèmes de formation de coalitions [Bogomolnaia and Jackson, 2001].

Comme nous nous intéressons dans ce chapitre à l'explication de préférences à partir d'autres préférences sans en référer à leur représentation globale – la fonction de score ω en l'occurrence – nous privilégierons à la précédente définition, une caractérisation axiomatique des relations d'ordre linéaire additives. De cette caractérisation, dériveront un ensemble de propriétés. Ces propriétés nous permettront de mettre en lumière des *mécanismes de déduction* à partir desquels nous définirons des *schémas d'explication* (Section 2.3).

2.2.1 Propriétés fondamentales et dérivées

Des travaux de [Scott, 1964] et [Fishburn, 1996], il résulte qu'une relation binaire \succ sur $2^{[m]}$ est une relation d'ordre linéaire additive si et seulement si elle vérifie les trois axiomes suivants :

Axiome 1. \succ sur $2^{[m]}$ est une relation d'ordre linéaire.

Axiome 2. $x \succ \emptyset$ pour tout $x \in 2^{[m]} \setminus \{\emptyset\}$.

1. Le terme *relation d'ordre linéaire additive* employé est la traduction littérale du terme anglais *additive linear order* utilisé dans ledit article qui, en étudie les propriétés algébriques et combinatoires dans le but de contribuer à l'implémentation de stratégies efficaces d'élicitation incrémentale de préférences exprimées sur des ensembles d'items. À ce titre, l'article fournit un certain nombre d'informations d'intérêt pour notre travail comme par exemple, la définition des *paires critiques* qui sont caractéristiques de chaque relation d'ordre linéaire additive. Il s'agit en effet, des éléments de la relation à partir desquelles se déduisent l'ensemble de tous ses autres éléments.

Axiome 3. Pour tout $k \geq 2$ et tout $u_1, \dots, u_k, v_1, \dots, v_k$, si

$$|j : \mathbf{i} \in u_j| = |j : \mathbf{i} \in v_j| \text{ pour tout } \mathbf{i} \in [m] \quad (2.2)$$

alors

$$\bigwedge_{j \in [1; k-1]} u_j \succ v_j \implies v_k \succ u_k \quad (2.3)$$

où \bigwedge symbolise la conjonction (logique).

Tel que présenté, l'**Axiome 3** (connu sous le nom de *cancellation axiom*) décrit un mécanisme de déduction de l'appartenance de la comparaison par paire (v_k, u_k) à la relation \succ à partir de l'ensemble $\{(u_1, v_1), \dots, (u_{k-1}, v_{k-1})\}$ de comparaisons par paire appartenant aussi à \succ :

- Les alternatives $u_1, \dots, u_k, v_1, \dots, v_k$ sont liées par une condition : l'alternative u_k d'une part et l'alternative v_k d'autre part sont les seules qui, à partir des autres alternatives en jeu, permettent la vérification de l'Équation (2.2).
- La conclusion (unique donc) $v_k \succ u_k$ s'impose dès lors que les "assertions" $u_j \succ v_j$, $j \in [1; k-1]$ sont toutes vérifiées.

Arrêtons-nous un moment sur la preuve de la validité de l'Équation (2.3). En admettant l'équivalence entre la définition axiomatique des ROLA et leur définition (de représentation) caractérisée par l'Équation 2.1, en procédant par l'absurde (posant $u_k \succ v_k$, la relation \succ étant totale (**Axiome 1**)) et en supposant que la condition (2.2) – qui stipule que chaque item \mathbf{i} apparaît autant de fois dans l'ensemble des alternatives u_1, \dots, u_k que dans l'ensemble des alternatives v_1, \dots, v_k – se vérifie, on parvient en appliquant (2.1) à la contradiction $0 > 0$.

Dans la littérature, l'ensemble $\{u_1 \succ v_1, \dots, u_{k-1} \succ v_{k-1}\}$ des déclarations de préférences d'un décideur constitue ce que [Kadzinski et al., 2014] appellent un *preferential reduct* de (v_k, u_k) qui est déduit par calcul de la relation nécessaire². Dans leur article, ces auteurs suggèrent l'utilisation des *preferential reducts* dans un contexte interactif où, le décideur, aidé par un analyste, participe à la construction d'une décision qu'il doit prendre. En effet, [Kadzinski et al., 2014] estiment que, informé du lien entre la préférence (nécessaire) de l'alternative v_k sur l'alternative u_k et un nombre minimal (au sens de l'inclusion) de ses précédentes déclarations de préférences, le décideur acquerrait une meilleure compréhension de la méthode employée par l'analyste et pourrait, en cas de désaccord avec la conclusion, réviser ces dernières. À ce titre, les *preferential reducts* sont présentés comme une *explication* de la déduction faite. En nous autorisant un parallèle avec la situation que pose l'**Axiome 3**, nous pourrions également affirmer que

2. La relation nécessaire est un concept introduit dans la littérature par [Greco et al., 2008]. Le Chapitre 4 de ce travail traite essentiellement de l'explication des éléments de la relation nécessaire.

$\{u_1 \succ v_1, \dots, u_{k-1} \succ v_{k-1}\}$ est une explication de la déduction $v_k \succ u_k$. Cette façon d'expliquer nous paraîtrait alors insatisfaisante à plusieurs égards :

- la preuve complète de la déduction $v_k \succ u_k$ s'appuie (comme on l'a indiqué précédemment dans la preuve de l'Équation (2.3)) sur la structure de représentation de la relation \succ et
- surtout, l'explication correspondante n'explique pas les *mécanismes* en jeu dans la déduction : elle ne dit rien de la *façon* dont s'articulent et se combinent les éléments explicatifs $u_1 \succ v_1, \dots, u_{k-1} \succ v_{k-1}$ pour parvenir à la conclusion $v_k \succ u_k$.

L'ambition de ce chapitre est donc de remédier à ces inconvénients et faiblesses en proposant une explication de l'appartenance de la paire (v_k, u_k) à la relation \succ à partir de l'ensemble $\mathbb{E}_\succ = \{u_1 \succ v_1, \dots, u_{k-1} \succ v_{k-1}\}$, qui mettent en lumière les mécanismes à l'œuvre dans cette déduction. Une telle explication sera conforme à ce que l'on appellera un *schéma déductif*. Cette conception de l'explication se trouve être proche du concept de *Justification* de [Boixel and Endriss, 2020] qui, dans le domaine du Choix Social Computationnel, indiquent que la justification de la sélection d'un ou de plusieurs candidats se doit d'être « *une conséquence logique d'un certain nombre de principes normatifs de base qu'avec un peu de chance tout le monde accepterait* »³. Dans leur article, ces auteurs développent un modèle de justification expliquée (« *a step-by-step explanation* ») d'une telle sélection et qui est fondé sur différents axiomes de règles de vote ([Zwicker and Moulin, 2016]).

De la définition axiomatique des ROLA présentée en début de sous-section, il découle les 4 lemmes suivants :

Lemme 2.1 (Transitivité) \succ est transitive :

$$x \succ y \wedge y \succ z \implies x \succ z \quad \forall x, y, z \in 2^{[m]}$$

Démonstration 2.1 \succ étant une relation d'ordre linéaire (**Axiome 1**) alors \succ est transitive.

Lemme 2.2 (Congruence)

$$x \succ y \implies \tilde{x} \cup z \succ \tilde{y} \cup z \quad \forall x, y, z \in 2^{[m]}$$

où $\tilde{x} = x \setminus y$, $\tilde{y} = y \setminus x$ et $z \cap (\tilde{x} \cup \tilde{y}) = \emptyset$.

Démonstration 2.2 En prenant $k = 2$ (**Axiome 3**) et en posant : $(u_1, v_1) = (x, y) = (x \cup (x \cap y), y \cup (x \cap y))$ d'une part et $(u_2, v_2) = (\tilde{y} \cup z, \tilde{x} \cup z) = ((y \setminus (x \cap y)) \cup z, (x \setminus (x \cap y)) \cup z)$, on a, étant donné que $z \cap (\tilde{x} \cup \tilde{y}) = \emptyset$, la condition (2.2) qui est vérifiée et donc l'implication du Lemme 2.2.

3. Traduction personnelle.

Lemme 2.3 (Couverture)

$$x \succ y \wedge x' \succ y' \implies x \cup x' \succ y \cup y' \quad \forall x, y, x', y' \in 2^{[m]}$$

où $x \cap x' = \emptyset$ et $y \cap y' = \emptyset$.

Démonstration 2.3 En prenant $k = 3$ (**Axiome 3**) et en posant : $(u_1, v_1) = (x, y)$, $(u_2, v_2) = (x', y')$ et $(u_3, v_3) = (y \cup y', x \cup x')$, on a, étant donné que $x \cap x' = \emptyset$ et $y \cap y' = \emptyset$, la condition (2.2) qui est vérifiée et donc l'implication du Lemme 2.3.

Lemme 2.4 (Dominance)

$$x \succ y \quad \forall x, y \in 2^{[m]}$$

où $y \neq \emptyset$ et $y \subset x$.

Démonstration 2.4 On prouve le Lemme 2.4 en combinant l'**Axiome 2** au Lemme 2.2 : on a $x \setminus y \succ \emptyset$ et comme $(x \setminus y) \cup y = x$ et $\emptyset \cup y = y$, on déduit que : $x \succ y$.

Le Lemme 2.4 (**Dominance**) nous permet d'identifier un ensemble de comparaisons par paire communes à toutes les ROLA \succ définies sur $2^{[m]}$. Il s'agit des comparaisons (u, v) exprimant une dominance de Pareto (telles que v est un sous-ensemble, au sens strict, de u). Nous noterons par \mathbb{D}_m l'ensemble de telles comparaisons par paire et puis, dans une logique d'explication, nous considérerons que les éléments de \mathbb{D}_m , du fait de leur évidence, sont dispensés d'explication :

Hypothèse 2.1 (Statut de l'ensemble \mathbb{D}_m vis-à-vis de l'explication) Les comparaisons par paire (u, v) exprimant une dominance de Pareto – $v \subset u$ – constituant l'ensemble \mathbb{D}_m et communes à toutes les relations d'ordre linéaire additives définies sur $2^{[m]}$ sont considérées comme étant **admisses**. Ceci implique qu'elles peuvent être utilisées dans l'explication de toute comparaison par paire $(x, y) \in \succ$ sans qu'il ne soit nécessaire de les expliquer.

Dans la prochaine sous-section, nous définirons trois mécanismes de déduction qui découlent directement des lemmes présentés ci-dessus. Ces mécanismes de déduction permettront ensuite de définir dans la Section 2.3 ce qu'est un schéma d'explication.

2.2.2 Mécanismes de déduction associés

Les Lemmes 2.1, 2.2 et 2.3 ont pour point commun d'édicter les conditions sous lesquelles on parvient à montrer que deux alternatives sont en relation par \succ à partir d'une ou plusieurs comparaisons par paire de \succ :

- dans le cas du Lemme 2.1, il faut avoir deux comparaisons par paire $u_1 \succ v_1$ et $u_2 \succ v_2$ telles que : $u_2 = v_1$.

- dans le cas du Lemme 2.2, il faut avoir une comparaison par paire $u \succ v$ et une alternative z disjointe de $u \setminus v$ et de $v \setminus u$ (disjointe de l'ensemble des items qui sont spécifiques à u et v).
- dans le cas du Lemme 2.3, il faut avoir deux comparaisons par paire $u_1 \succ v_1$ et $u_2 \succ v_2$ telles que les alternatives u_1 et u_2 d'une part et les alternatives v_1 et v_2 d'autre part sont disjointes.

Cette remarque permet de définir 3 (simples) mécanismes de déduction (1 par Lemme). Ces mécanismes prennent la forme d'un ensemble de tuples de comparaisons par paire soumises à des conditions précises et qui permettent d'aboutir à une conclusion unique. Chaque tuple de comparaisons par paire est désigné sous le terme d'*instance*. On parlera donc d'*instance* d'un mécanisme de déduction qui induit une conclusion unique.

Les mécanismes que nous considérerons dans ce chapitre sont les suivants :

- Le mécanisme de déduction 2- Tr^\succ (et plus généralement k - Tr^\succ , $k \geq 2$) étant donné la ROLA \succ ,
- Le mécanisme de déduction 2- Cov^\succ (et plus généralement k - Cov^\succ , $k \geq 2$) étant donné la ROLA \succ ,
- Le mécanisme de déduction z - Cg^\succ ($z \in 2^{[m]}$) étant donné la ROLA \succ .

Comme on peut naturellement s'y attendre, ces mécanismes appliquent chacun une des trois propriétés des relations d'ordre linéaire additives exprimées à travers les Lemmes 2.1, 2.2 et 2.3.

La première des trois propriétés est la transitivité. On peut la paraphraser de la façon suivante : si l'alternative x est préférée à l'alternative y et l'alternative y à l'alternative z , alors x doit être préférée à z . Il s'agit d'une propriété assez courante et vérifiée par une bonne majorité des modèles de préférences que l'on rencontre dans la littérature. [Payne et al., 1999], proposant un ensemble de mesures de la qualité des méthodes d'aide à la décision relevant de l'approche constructiviste, font d'ailleurs remarquer que la plupart des décideurs estiment que leurs préférences devraient être transitives.

La congruence est la propriété que sous-tend le Lemme 2.2. Plus connue sous le nom de propriété d'*indépendance* (*Strong Extended Independence* plus précisément dans [Barberà et al., 2004]), cette propriété s'exprime en réalité sous la forme d'une équivalence : pour tout $x, y, z \in 2^{[m]}$ où $z \cap (x \cup y) = \emptyset$,

$$x \succ y \iff x \cup z \succ y \cup z \quad (2.4)$$

Elle peut être résumée en la formule suivante : les préférences doivent être indépendantes de l'ajout ou du retrait d'items-tiers, non fondamentaux dans la comparaison. Nous avons opté, dans la présentation du Lemme 2.2, pour une forme "beaucoup plus opérationnelle" de la propriété de congruence qui lui est équivalente : si x est préférée à y , en remplaçant dans x et y les items qui leur sont communs par un ensemble d'items

z disjoints des items restants, le sens de la préférence ne devrait pas changer. Ce choix de présentation nous a semblé plus en phase avec le concept de *mécanisme de déduction* introduit.

La couverture est la troisième et dernière des propriétés à partir desquelles nous définirons nos mécanismes de déduction. Le terme *couverture* est emprunté au célèbre problème de couverture par les ensembles (*Set Cover problem*), l'un des 21 problèmes NP-complets de [Karp, 1972] : la comparaison par paire $(x \cup x') \succ (y \cup y')$ étant “couverte” par $x \succ x'$ et $y \succ y'$. Dans la littérature consacrée à l'explication de préférences additives, on retrouve par exemple la contribution de [Belahcène et al., 2017] qui développe une méthode d'explication des déductions de la relation nécessaire à l'aide de *preference swaps* qui peuvent être vus comme des comparaisons par paire élémentaires couvrant la comparaison par paire dont on souhaite expliquer la déduction.

Dans la suite, nous proposerons une définition formelle aux trois mécanismes de déduction d'éléments de la ROLA \succ que nous considérerons dans ce chapitre. Nous indiquerons ensuite dans quelles mesures leur mise en œuvre permet d'entrevoir la question de l'explication au travers de schémas déductifs.

Définitions formelles des mécanismes de déduction

Nous illustrerons les définitions de ce paragraphe à partir de la ROLA \succ_ω représentée par la fonction de score ω de l'Exemple suivant :

Exemple 2.1 (Relation d'ordre linéaire additive illustrative)

$$\omega = (a : 3, b : 6, c : 8, d : 10) \quad (2.5)$$

On remarquera que ω est bien une *Subset-Sum-Distinct Sequence (SSDS)* comme l'atteste la Table 2.1 qui, par ordre décroissant, fournit les scores associés à chacune des alternatives de l'ensemble $2^{[4]} \setminus \{\emptyset\}$:

TABLE 2.1 – Classement des alternatives de $2^{[4]} \setminus \{\emptyset\}$ suivant ω (Équation 2.5)

| x | abcd | bcd | acd | abd | cd | abc | bd | bc | ad | ac | d | ab | c | b | a |
|-------------|------|-----|-----|-----|----|-----|----|----|----|----|----|----|---|---|---|
| $\omega(x)$ | 27 | 24 | 21 | 19 | 18 | 17 | 16 | 14 | 13 | 11 | 10 | 9 | 8 | 6 | 3 |

Définition 2.1 (Mécanisme de déduction $k\text{-Tr}^\succ$) Étant donné un entier naturel k où $k \geq 2$, on définit comme suit l'ensemble $k\text{-Tr}^\succ$:

$$k\text{-Tr}^\succ = \{(u_1 \succ v_1, \dots, u_k \succ v_k) \in \succ^k : u_{j+1} = v_j \text{ pour tout } 1 \leq j < k\}$$

Chaque instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ de ce mécanisme de déduction a pour **unique** conclusion $u_1 \succ v_k$ par application du Lemme 2.1 étendu à k comparaisons par paire.

Exemple 2.2 (Suite de l'Exemple 2.1) $(bd \succ_{\omega} ad, ad \succ_{\omega} ac)$, $(bd \succ_{\omega} bc, bc \succ_{\omega} ac)$ et $(bd \succ_{\omega} bc, bc \succ_{\omega} ad, ad \succ_{\omega} ac)$ sont les 3 (seules) instances des mécanismes $k\text{-Tr}^{\succ_{\omega}}$ ($k \in \{2, 3\}$) dont la conclusion est : $bd \succ_{\omega} ac$.

Il n'existe pas d'instance des mécanismes $k\text{-Tr}^{\succ_{\omega}}$ ($k \geq 2$) dont la conclusion est $c \succ_{\omega} b$. En effet, les alternatives c et b sont adjacentes suivant l'ordre induit par \succ_{ω} sur l'ensemble $2^{[4]} \setminus \{\emptyset\}$.

Définition 2.2 (Mécanisme de déduction $z\text{-Cg}^{\succ}$) Étant donné une alternative $z \in 2^{[m]}$, on définit l'ensemble $z\text{-Cg}^{\succ}$ comme suit :

$$z\text{-Cg}^{\succ} = \{(u, v) \in \succ : z \cap (\tilde{u} \cup \tilde{v}) = \emptyset\}$$

où $\tilde{u} = u \setminus v$, $\tilde{v} = v \setminus u$.

Chaque instance (u, v) de cet ensemble a pour **unique** conclusion $\tilde{u} \cup z \succ \tilde{v} \cup z$ par application du Lemme 2.2.

Exemple 2.3 (Suite de l'Exemple 2.1) La seule instance des mécanismes $z\text{-Cg}^{\succ_{\omega}}$ ($z \in 2^{[m]}$) dont la conclusion est $bd \succ_{\omega} ac$ est $bd \succ_{\omega} ac$ (elle-même) qui appartient au mécanisme $\emptyset\text{-Cg}^{\succ_{\omega}}$.

$ac \succ_{\omega} ab$ est l'une des instances du mécanisme $\emptyset\text{-Cg}^{\succ_{\omega}}$ dont la conclusion est $c \succ_{\omega} b$, le critère a commun à ac et ab ayant été remplacé par \emptyset .

Définition 2.3 (Mécanisme de déduction $k\text{-Cov}^{\succ}$) Étant donné un entier naturel k où $k \geq 2$, on définit l'ensemble $k\text{-Cov}^{\succ}$ de la manière suivante :

$$k\text{-Cov}^{\succ} = \{(u_1 \succ v_1, \dots, u_k \succ v_k) \in \succ^k : u_j \cap u_{j'} = v_j \cap v_{j'} = \emptyset \text{ pour tous } j \neq j'\}$$

Chaque instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ de ce mécanisme de déduction a pour **unique** conclusion $\bigcup_{j \in [1; k]} u_j \succ \bigcup_{j \in [1; k]} v_j$ par application du Lemme 2.3 étendu à k comparaisons par paire.

Exemple 2.4 (Suite de l'Exemple 2.1) Les séquences $(d \succ_{\omega} c, b \succ_{\omega} a)$ et $(b \succ_{\omega} a, d \succ_{\omega} c)$ sont les seules instances des mécanismes $k\text{-Cov}^{\succ_{\omega}}$ ($k \geq 2$) dont la conclusion est $bd \succ_{\omega} ac$.

De même qu'il n'existe pas d'instance des mécanismes $k\text{-Tr}^{\succ_{\omega}}$ ($k \geq 2$) dont la conclusion est $c \succ_{\omega} b$ (Exemple 2.2), il n'existe pas d'instance des mécanismes $k\text{-Cov}^{\succ_{\omega}}$ ($k \geq 2$) dont la conclusion est $c \succ_{\omega} b$. Cette inexistence d'instance de mécanismes de couverture s'explique par la structure élémentaire de la comparaison $c \succ_{\omega} b$ dont les deux alternatives sont des singletons.

Vers une formalisation de la notion d'explication

Nous avons précédemment présenté des mécanismes de déduction de l'appartenance d'une comparaison par paire (x, y) à une relation d'ordre linéaire additive \succ à partir d'une ou plusieurs autres comparaisons par paire appartenant à la même relation. Les mécanismes de déduction sont composés d'*instances* qui jouent le rôle de prémisses dans l'opération de déduction de la conclusion (x, y) . Chaque opération de déduction s'appuie sur une propriété bien précise des relations d'ordre linéaire additives. **Les propriétés retenues dans ce travail sont la transitivité, l'indépendance ou congruence et la couverture.** Nous admettons qu'il s'agit de propriétés dont l'utilisation dans une optique d'**explication** peut être envisagée.

Si l'on peut affirmer le cas échéant, que la conclusion (x, y) se **déduit** d'une *instance* d'un mécanisme associé à l'une de ces trois propriétés, peut-on dire que la conclusion (x, y) s'**explique** à partir de la même *instance*?

Dans ce chapitre, nous répondrons par la négative à cette question. Autrement, nous considérerions, dans le cas de l'Exemple 2.3, que la conclusion $bd \succ_{\omega} ac$ s'explique à partir d'elle-même par application de la propriété de congruence ; ce qui ne nous semble pas satisfaisant. Pour parler d'**explication**, il nous faut préciser l'ensemble des comparaisons par paire que l'on s'autorise à utiliser dans le "contenu" de l'explication. Ces comparaisons par paire doivent être considérées comme **admises** à l'image des comparaisons par paire exprimant une dominance de Pareto (Hypothèse 2.1) qui sont dispensables d'explication. Dans les travaux de [Boixel and Endriss, 2020] et de [Kadzinski et al., 2014], il s'agit respectivement d'un profil de préférences d'agents sur différents candidats et de l'ensemble des déclarations de préférences du décideur. Dans ce chapitre, nous proposerons des explications par l'intermédiaire de *schémas déductifs* (Section 2.3) en considérant un sous-ensemble \mathbb{B}_{\succ} (appelé base de référence) de la relation d'ordre linéaire additive \succ dont toutes les comparaisons par paire seront considérées comme **admises** (Hypothèse 2.2). Ce faisant, l'**explication** peut être vue comme un instrument de déduction qui opère par l'intermédiaire de *schémas déductifs* qui mettent en œuvre des mécanismes de déduction précis et à partir d'un ensemble de comparaisons par paire considérées comme **admises**.

Exemple 2.5 (Suite de l'Exemple 2.1) *En considérant comme base de référence de la ROLA \succ_{ω} représentée par la fonction de score ω (Équation 2.5), l'ensemble $\mathbb{B}_{\succ_{\omega}}$ suivant :*

$$\mathbb{B}_{\succ_{\omega}} = \{ac \succ_{\omega} d, d \succ_{\omega} ab, d \succ_{\omega} c, b \succ_{\omega} a\}$$

on a que :

1. $bd \succ_{\omega} ac$ s'explique (par couverture) à partir de $\mathbb{B}_{\succ_{\omega}}$. En effet,

$$d \succ_{\omega} c \wedge b \succ_{\omega} a \implies bd \succ_{\omega} ac$$

2. $c \succ_{\omega} b$ s'explique (par congruence puis transitivité) à partir de $\mathbb{B}_{\succ_{\omega}}$. En effet,

$$ac \succ_{\omega} ab \implies c \succ_{\omega} b$$

avec :

$$ac \succ_{\omega} d \wedge d \succ_{\omega} ab \implies ac \succ_{\omega} ab$$

On remarque, à travers cet exemple, que l'explication d'une comparaison par paire peut mobiliser plusieurs mécanismes de déduction et donc plusieurs propriétés. C'est le cas notamment de la comparaison par paire $c \succ_{\omega} b$ qui se déduit (en non s'explique) de $ac \succ_{\omega} ab$ (car n'appartenant pas à $\mathbb{B}_{\succ_{\omega}}$) qui s'explique à partir des comparaisons par paire (considérées comme **admises**) $ac \succ_{\omega} d$ et $d \succ_{\omega} ab$. L'explication de $c \succ_{\omega} b$ est donc indirecte et s'obtient grâce à une composition de mécanismes de déduction : d'abord la congruence puis la transitivité. Cet "enchaînement" définit un schéma déductif auquel l'on peut associer un algorithme dédié. Ce sera entre autres, l'objet du Chapitre 3.

Dans l'Exemple 2.5 nous avons formulé l'explication de deux comparaisons par paire dans un "langage" totalement informel (mélangeant du langage naturel et des sortes de formules logiques). Dans la section 2.3, nous spécifierons trois différents formats de représentation de l'explication d'une comparaison par paire (x, y) . Comme annoncé en introduction, une telle explication se conçoit dans un *contexte* qui rassemble la ROLA \succ considérée, une base de référence \mathbb{B}_{\succ} ($\subset \succ$) et un schéma déductif.

2.3 Formalisation de l'explication pour le modèle additif

À la fin de section précédente (juste avant l'Exemple 2.5), nous avons montré, sans proposer une définition formelle à la notion d'explication, comment cette dernière suppose l'existence d'une *base de référence* composée de comparaisons par paire à considérer comme **admises**. Ces comparaisons par paire seront structurées suivant des *schémas déductifs* qui mettent un œuvre un ou plusieurs mécanismes de déduction parmi les trois proposés en sous-section 2.2.2. Ces *schémas déductifs* sont assimilables aux *schémas d'argument* de [Walton, 1996] qui en donne une description qui cadre bien avec la logique dans laquelle nous en faisons usage ici. En effet, écrit-il (page 10) que : (les schémas d'argument) « ...sont normativement contraignants ... Si l'auditeur accepte les prémisses de l'argument du locuteur, et que l'argument est un exemple de schéma d'argument authentique et approprié (pour le type de dialogue dans lequel ils sont engagés), alors l'auditeur doit ou devrait de manière contraignante accepter la conclusion. »⁴. Dans notre cas, les « prémisses » sont à extraire de la *base de référence* et/ou de l'ensemble \mathbb{D}_m (Hypothèse 2.1).

4. Traduction personnelle.

L'« argument », c'est la mise en œuvre des mécanismes de déduction qui, basés sur des propriétés-axiomes des relations d'ordre linéaire additives, conduisent à une « conclusion ». L'objet de l'explication, c'est cette conclusion ; les schémas déductifs ne faisant que décrire formellement l'« enchaînement » des mécanismes de déduction permettant d'aboutir à cette conclusion et ce, à partir de comparaisons par paire considérées comme admises. Au sujet des comparaisons par paire de la base de référence et à l'image de l'Hypothèse 2.1, nous considérons :

Hypothèse 2.2 (Statut de la base de référence \mathbb{B}_{\succ} vis-à-vis de l'explication) *Soit \succ une relation d'ordre linéaire additive dont un sous-ensemble est \mathbb{B}_{\succ} , une base de référence à partir de laquelle on cherche à produire des explications suivant un schéma déductif. Les comparaisons par paire (u, v) de \mathbb{B}_{\succ} sont considérées comme étant **admises**. Ceci implique qu'elles peuvent être utilisées dans l'explication de toute comparaison par paire $(x, y) \in \succ$ sans qu'il ne soit nécessaire de les expliquer.*

Dans la sous-section 2.3.1, nous proposons au lecteur de découvrir par l'exemple, les trois différentes représentations de l'explication que nous adopterons dans ce travail. Pour ce faire, nous lui demandons de se prêter au « jeu » du calcul de l'explication de différentes comparaisons par paire. Nous convenons qu'il ne s'agit pas d'une tâche facile. Elle a cependant l'avantage de permettre que soit perçue et imaginée la variété des différents schémas déductifs qu'il est possible de construire à partir des trois mécanismes de déduction présentés en sous-section 2.2.2. Dans la sous-section 2.3.2 consacrée à la définition formelle de schémas déductifs, nous introduirons le concept de *niveau d'explication* qui pourrait servir à regrouper différents schémas selon leur structure. Enfin, dans la sous-section 2.3.3, nous proposerons une résolution du « jeu » de calcul d'explication présenté en sous-section 2.3.1.

2.3.1 Exemple introductif

Dans ce travail, nous considérerons trois représentations (équivalentes) de l'explication à l'aide de schémas déductifs d'une comparaison par paire (x, y) :

- une représentation sous forme de preuve logique exploitant directement les implications (logiques) induites par les *instances* des différents mécanismes de déduction (sous-section 2.2.2),
- une représentation textuelle (en langage naturel) et
- une représentation sous forme de diagramme de déduction.

En guise d'illustration, nous reprenons la ROLA représentée par la fonction de score de l'Équation 2.5 (déjà utilisée dans l'Exemple 2.5) et considérons la base de référence $\mathbb{B}_{\succ_{\omega}}$ suivante⁵ :

$$\mathbb{B}_{\succ_{\omega}} = \{d \succ_{\omega} ab \bullet ab \succ_{\omega} c \bullet bc \succ_{\omega} ad \bullet b \succ_{\omega} a\}$$

5. Les \bullet servant juste de séparateurs entre deux comparaisons par paire de la base de référence.

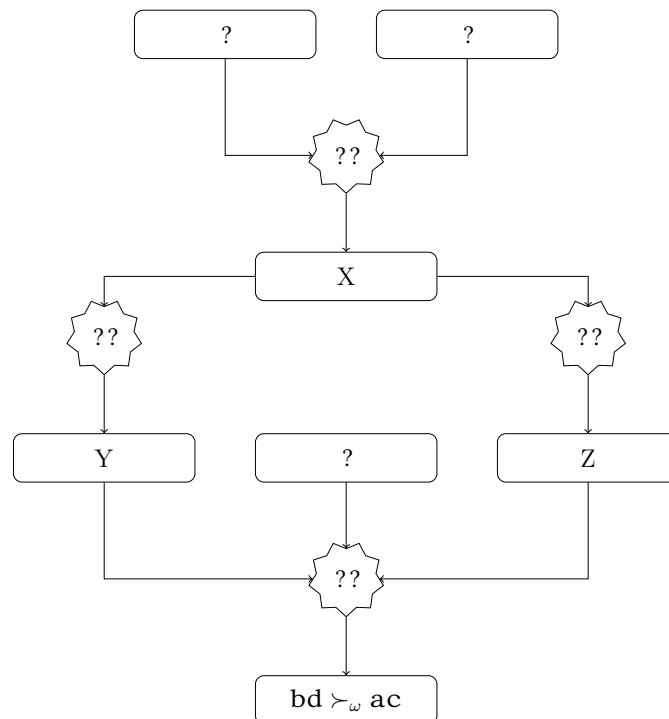
$\mathbb{B}_{\succsim_\omega}$ est composée de quatre comparaisons par paire qui permettent d'expliquer les trois comparaisons par paire suivantes :

$$\mathbb{E}_{\succsim_\omega} = \{d \succsim_\omega c \bullet c \succsim_\omega a \bullet bd \succsim_\omega ca\}$$

Le but du jeu consiste à remplir les trois structures à trous suivantes qui correspondent aux explications des comparaisons par paire de l'ensemble $\mathbb{E}_{\succsim_\omega}$. Pour ce faire, il faudra :

- ❑ remplacer les “?” par des comparaisons par paire considérées comme admises (Hypothèses 2.1 et 2.2)
- ❑ remplacer les “??” par les mécanismes de déduction ou les propriétés (*Tr*, *Cg*, ou *Cov*) qui conviennent.
- ❑ remplacer les variables X, Y et Z par les comparaisons par paire non-admises qui conviennent.

Explication de $bd \succsim_\omega ca$ (représentation sous forme de diagramme)



Explication de $d \succ_{\omega} c$ (représentation textuelle)

L'alternative d est préférée à l'alternative c car la comparaison par paire $d \succ_{\omega} c$ se déduit, par application du mécanisme $??$, des comparaisons par paire $?$ (admise) et $?$ (admise).

Explication de $c \succ_{\omega} a$ (représentation sous forme d'implications logiques)

$$\begin{array}{ccc}
 X & \implies & c \succ_{\omega} a & (??) \\
 & & \text{avec :} & \\
 ? \wedge ? \wedge ? & \implies & X & (??)
 \end{array}$$

2.3.2 Schémas déductifs

Cette sous-section est consacrée à la définition de schémas auxquels on pourrait recourir pour l'explication de la déduction d'une comparaison par paire (x, y) appartenant à une ROLA \succ à partir d'autres comparaisons par paire (considérées comme admises) de \succ . Nous considérerons deux grandes catégories de schémas déductifs :

- Les schémas déductifs de base ou de *niveau* 0 et
- Les schémas déductifs composés ou de *niveau* supérieur à 0.

Les schémas déductifs de base : tr , cov , cg

Il s'agit des trois schémas découlant directement des trois mécanismes de déduction présentés en sous-section 2.2.2. Étant donné une relation d'ordre linéaire additive \succ (définie sur $2^{[m]}$) dont un sous-ensemble noté \mathbb{B}_{\succ} sert de base de référence à l'explication d'autres comparaisons par paire de \succ , nous considérons les définitions suivantes :

Définition 2.4 (Schéma tr) Une conclusion (x, y) s'explique à l'aide du (ou suivant le) schéma tr – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_{\succ}, tr)$ – si et seulement si il existe une instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ du mécanisme $k\text{-Tr}^{\succ}$ ($k \geq 2$) dont la conclusion est (x, y) et telle que chaque comparaison par paire $u_j \succ v_j$ ($j \in \llbracket 1; k \rrbracket$) est considérée comme admise c'est-à-dire $(u_j, v_j) \in \mathbb{D}_m$ ou $(u_j, v_j) \in \mathbb{B}_{\succ}$.

De façon similaire, on définit le schéma déductif cov :

Définition 2.5 (Schéma *cov*) Une conclusion (x, y) s'explique suivant le schéma *cov* – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, cov)$ – si et seulement si il existe une instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ du mécanisme $k\text{-Cov}^\succ$ ($k \geq 2$) dont la conclusion est (x, y) et telle que chaque comparaison par paire $u_j \succ v_j$ ($j \in \llbracket 1; k \rrbracket$) est considérée comme admise c'est-à-dire $(u_j, v_j) \in \mathbb{D}_m$ ou $(u_j, v_j) \in \mathbb{B}_\succ$.

On définit comme suit le schéma déductif *cg* :

Définition 2.6 (Schéma *cg*) Une conclusion (x, y) s'explique à l'aide du schéma *cg* – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, cg)$ – si et seulement si il existe une instance $u \succ v$ du mécanisme $z\text{-Cg}^\succ$ ($z \in 2^{\llbracket m \rrbracket}$) dont la conclusion est (x, y) et telle que la comparaison par paire $u \succ v$ est considérée comme admise : $(u, v) \in \mathbb{D}_m$ ou $(u, v) \in \mathbb{B}_\succ$.

Au vu de ces définitions, il est évident que :

$$\mathcal{E}(\succ, \emptyset, tr) = \mathcal{E}(\succ, \emptyset, cov) = \mathcal{E}(\succ, \emptyset, cg) = \mathbb{D}_m \quad (2.6)$$

Autrement dit, seules les comparaisons par paire exprimant une dominance de Pareto sont explicables lorsque la base de référence est vide.

Les schémas déductifs composés : $sch_0/sch_1 \cdot \dots \cdot sch_n$ ($n \geq 1$)

Nous considérerons ici trois sous-catégories de schémas déductifs en fonction du type du schéma de base sch_0 (*tr*, *cg* ou *cov*).

Définition 2.7 (Schéma $tr/sch_1 \cdot \dots \cdot sch_n$, avec $n \geq 1$) On suppose que pour tout $i \in \llbracket 1; n \rrbracket$, sch_i est le schéma déductif de base *cg* ou *cov* ou un schéma composé ou $sch_i = cg/sch_{i,1}$ (Définition 2.9) ou $sch_i = cov/sch_{i,1} \cdot \dots \cdot sch_{i,n_i}$, avec $n_i \geq 1$ (Définition 2.8).

Une conclusion (x, y) s'explique à l'aide du schéma $tr/sch_1 \cdot \dots \cdot sch_n$ – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, tr/sch_1 \cdot \dots \cdot sch_n)$ – si et seulement si il existe une instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ du mécanisme $k\text{-Tr}^\succ$ ($k \geq 2$ et $k \geq n$) dont la conclusion est (x, y) et telle que pour chaque i ($1 \leq i \leq n$), il existe au moins un $j \in \{1, \dots, k\}$ tel que $(u_j, v_j) \in \mathcal{E}(\succ, \mathbb{B}_\succ, sch_i)$; les autres comparaisons par paire de l'instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ étant considérées comme admises (Hypothèses 2.1 et 2.2).

De façon similaire, lorsque $sch_0 = cov$, on a :

Définition 2.8 (Schéma $cov/sch_1 \cdot \dots \cdot sch_n$, avec $n \geq 1$) On suppose que pour tout $i \in \llbracket 1; n \rrbracket$, sch_i est le schéma déductif de base *cg* ou *tr* ou un schéma composé ou $sch_i = cg/sch_{i,1}$ (Définition 2.9) ou $sch_i = tr/sch_{i,1} \cdot \dots \cdot sch_{i,n_i}$, avec $n_i \geq 1$ (Définition 2.7).

Une conclusion (x, y) s'explique à l'aide du schéma $cov/sch_1 \cdot \dots \cdot sch_n$ – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, cov/sch_1 \cdot \dots \cdot sch_n)$ – si et seulement si il existe une instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ du mécanisme $k\text{-Cov}^\succ$ ($k \geq 2$ et $k \geq n$) dont la conclusion est (x, y) et

telle que pour chaque i ($1 \leq i \leq n$), il existe au moins un $j \in \{1, \dots, k\}$ tel que $(u_j, v_j) \in \mathcal{E}(\succ, \mathbb{B}_\succ, sch_i)$; les autres comparaisons par paire de l'instance $(u_1 \succ v_1, \dots, u_k \succ v_k)$ étant considérées comme admises (Hypothèses 2.1 et 2.2).

Les Définitions 2.8 et 2.7 décrivent une collection de schémas déductifs qui, pour expliquer la conclusion (x, y) , se servent d'une instance du mécanisme de déduction $k\text{-Tr}^\succ$ ou $k\text{-Cov}^\succ$ composée d'un ensemble de comparaisons par paire qui sont soit considérées comme admises ou qui sont explicables à l'aide des schémas sch_i , $1 \leq i \leq n$.

En suivant la même logique, on définit comme suit le schéma composé cg/sch :

Définition 2.9 (Schéma cg/sch) On suppose que sch est le schéma déductif tr ou cov ou un schéma composé $sch = tr/sch_1 \cdot \dots \cdot sch_n$ (Définition 2.7) ou $sch = cov/sch_1 \cdot \dots \cdot sch_n$, avec $n \geq 1$ (Définition 2.8).

Une conclusion (x, y) s'explique suivant le schéma cg/sch – on notera $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, cg/sch)$ – si et seulement si il existe une instance $u \succ v$ du mécanisme $z\text{-Cg}^\succ$ ($z \in 2^{[m]}$) dont la conclusion est (x, y) et telle que $(u, v) \in \mathcal{E}(\succ, \mathbb{B}_\succ, sch)$.

Au regard des définitions précédentes, on peut assimiler l'explication suivant un schéma déductif composé à une sorte de “processus” récursif qui procède de la mise en œuvre, à partir de comparaison(s) par paire explicable(s) ou considérée(s) comme admise(s), d'un mécanisme de déduction de la comparaison par paire expliquée. Avant de découvrir les différents schémas permettant d'expliquer les comparaisons par paire de l'ensemble $\mathbb{E}_{\succ_\omega}$ de la sous-section 2.3.1, nous introduisons ci-dessous, le concept de niveau d'explication qui peut servir à regrouper divers schémas déductifs (et les explications qui en résultent) en fonction des similitudes que présentent leurs représentations respectives.

Niveau d'explication et de schéma déductif

Dans l'introduction de la présente sous-section, nous avons indiqué que les schémas déductifs (dits de base) tr , cov et cg sont de niveau 0. Pour rappel (voir Définitions 2.4, 2.5 et 2.6), ces schémas s'appuient sur exactement un mécanisme de déduction (sous-section 2.2.2).

Quant aux schémas déductifs composés (Définitions 2.7, 2.8 et 2.9), on sait qu'ils mobilisent plus d'un mécanisme de déduction. Pour décrire une explication réalisée suivant un tel schéma composé et identifiée précédemment à une sorte de “processus” récursif, on pourrait s'intéresser au nombre total de propriétés différentes desquelles découlent l'ensemble des mécanismes de déduction mobilisés. Ce faisant, une telle explication pourrait recourir à deux ou au plus trois propriétés des relations d'ordre linéaire additives indépendamment de “la hauteur de l'arbre des appels récursifs” qui elle, a par contre un impact sur sa représentation et ce, quelle que soit sa nature (textuelle, formule(s) logique(s) ou diagramme).

Soit la définition suivante :

Définition 2.10 (Niveau d'une explication/schéma déductif) Le niveau \mathcal{L} d'un schéma déductif et de l'explication résultant de son utilisation se définit récursivement de la façon suivante :

- $\mathcal{L}(sch) = 0$ lorsque $sch \in \{tr, cov, cg\}$,
- $\mathcal{L}(sch_0/sch_1 \cdot \dots \cdot sch_n) = 1 + \max_{1 \leq i \leq n} \mathcal{L}(sch_i)$ lorsque $sch_0 \in \{tr, cov\}$,
- $\mathcal{L}(cg/sch) = 1 + \mathcal{L}(sch)$.

Exemple 2.6 Le tableau ci-après fournit les valeurs de niveau de différents schémas déductifs :

TABLE 2.2 – Niveaux de différents schémas déductifs

| sch | tr | cg/tr | $tr/cg/cov$ | $cg/cov/tr$ | $tr/cov \cdot cg/tr$ |
|--------------------|------|---------|-------------|-------------|----------------------|
| $\mathcal{L}(sch)$ | 0 | 1 | 2 | 2 | 2 |

2.3.3 Retour sur l'exemple introductif

Dans cette sous-section, nous proposons une résolution du “jeu” décrit en sous-section 2.3.1. Pour rappel, il s'agissait, à partir des comparaisons par paire de la base de référence $\mathbb{B}_{\gamma_\omega}$, d'expliquer à l'aide de schémas déductifs de base ou composés (sous-section 2.3.2) les comparaisons par paire de l'ensemble $\mathbb{E}_{\gamma_\omega}$.

$$\mathbb{B}_{\gamma_\omega} = \{ \mathbf{d} \gamma_\omega \mathbf{ab} \bullet \mathbf{ab} \gamma_\omega \mathbf{c} \bullet \mathbf{bc} \gamma_\omega \mathbf{ad} \bullet \mathbf{b} \gamma_\omega \mathbf{a} \}$$

$$\mathbb{E}_{\gamma_\omega} = \{ \mathbf{d} \gamma_\omega \mathbf{c} \bullet \mathbf{c} \gamma_\omega \mathbf{a} \bullet \mathbf{bd} \gamma_\omega \mathbf{ac} \}$$

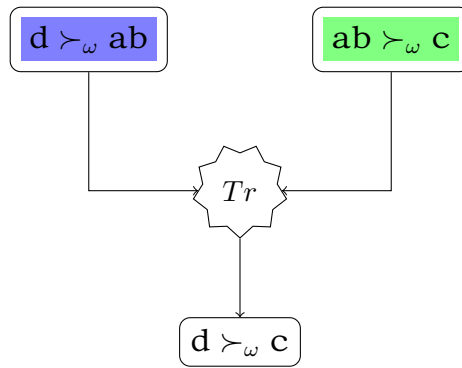
Pour bien montrer l'équivalence entre les (3) différentes représentations d'explication proposées dans ce travail, nous fournirons l'explication de chacune de ces comparaisons par paire en recourant à l'ensemble desdites représentations. Dans l'ordre, nous représenterons l'explication en langage naturel, puis à l'aide d'implication(s) logique(s) et enfin sous la forme de diagramme.

Dans chacune des représentations adoptées, les comparaisons considérées comme admises sont surlignées (le surlignage gris est réservé aux comparaisons par paire exprimant une dominance de Pareto). Dans la représentation en forme de diagramme, les comparaisons par paire admises sont les seuls “nœuds” sans “prédécesseurs” et le seul “nœud” sans successeur, c'est la comparaison par paire expliquée. Cette dernière est la seule comparaison par paire non soulignée située au bout d'une flèche d'implication \implies dans la représentation sous forme de formules logiques. Nous avons fait le choix de souligner (avec différents motifs) les comparaisons par paire non admises afin de faciliter la lecture de cette forme de représentation de l'explication.

Explication de $d \succ_{\omega} c$

L'alternative d est préférée à l'alternative c car la comparaison par paire $d \succ_{\omega} c$ se déduit, par application du mécanisme 2-Tr^{ω} , des comparaisons par paire $d \succ_{\omega} ab$ (admise) et $ab \succ_{\omega} c$ (admise).

$$d \succ_{\omega} ab \quad \wedge \quad ab \succ_{\omega} c \quad \implies \quad d \succ_{\omega} c \quad (2\text{-Tr}^{\omega})$$

FIGURE 2.1 – Explication de $d \succ_{\omega} c$ sous forme d'implications logiquesFIGURE 2.2 – Explication de $d \succ_{\omega} c$ sous forme de diagramme

On remarque à partir des différentes représentations ci-dessus que la structure d'explication de $d \succ_{\omega} c$ est conforme au schéma tr . Cette explication est donc de niveau 0.

Explication de $c \succ_{\omega} a$

L'alternative c est préférée à l'alternative a car la comparaison par paire $c \succ_{\omega} a$ se déduit, par application du mécanisme $\emptyset\text{-Cg}^{\omega}$, de la comparaison par paire $bc \succ_{\omega} ab$ (non admis) où :

- $bc \succ_{\omega} ab$ se déduit, par application du mécanisme 3-Tr^{ω} , de $bc \succ_{\omega} ad$ (admise), $ad \succ_{\omega} d$ (admise) et $d \succ_{\omega} ab$ (admise).

$$\underline{bc \succ_{\omega} ab} \implies c \succ_{\omega} a \quad (\emptyset\text{-Cg}^{\succ_{\omega}})$$

avec :

$$\underline{bc \succ_{\omega} ad} \wedge \underline{ad \succ_{\omega} d} \wedge \underline{d \succ_{\omega} ab} \implies \underline{bc \succ_{\omega} ab} \quad (3\text{-Tr}^{\succ_{\omega}})$$

FIGURE 2.3 – Explication de $c \succ_{\omega} a$ sous forme d'implications logiques

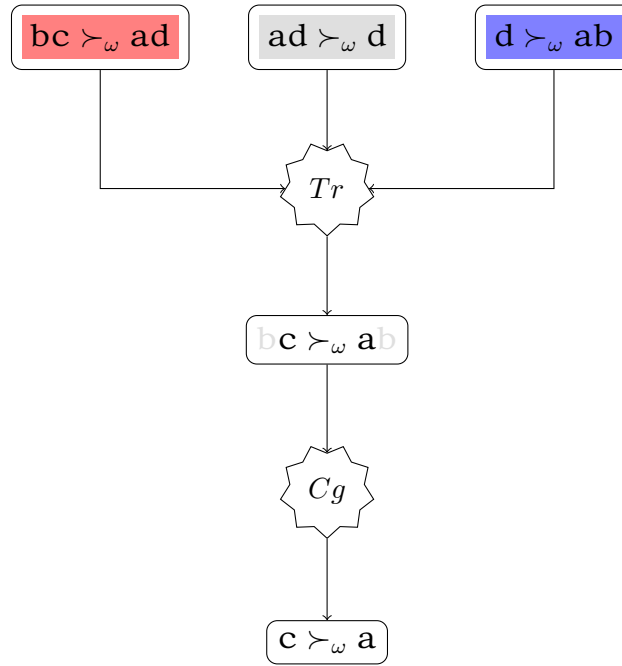


FIGURE 2.4 – Explication de $c \succ_{\omega} a$ sous forme de diagramme

L'explication de $c \succ_{\omega} a$ a été réalisée suivant le schéma cg/tr . Cette explication est donc de *niveau 1*.

Explication de $bd \succ_{\omega} ac$

L'alternative bd est préférée à l'alternative ac car la comparaison par paire $bd \succ_{\omega} ac$ se déduit, par application du mécanisme $3\text{-Tr}^{\succ_{\omega}}$, des comparaisons par paire $bd \succ_{\omega} bc$ (non admis), $\underline{bc \succ_{\omega} ad}$ (admise) et $ad \succ_{\omega} ac$ (non admis) où :

- $bd \succ_{\omega} bc$ se déduit, par application du mécanisme $b\text{-Cg}^{\succ_{\omega}}$, de la comparaison par paire $abd \succ_{\omega} abc$ (non admis) où :
 - $abd \succ_{\omega} abc$ se déduit, par application du mécanisme $2\text{-Cov}^{\succ_{\omega}}$, des comparaisons par paire $\underline{d \succ_{\omega} ab}$ (admise) et $\underline{ab \succ_{\omega} c}$ (admise).

- $ad \succ_{\omega} ac$ se déduit, par application du mécanisme $a-Cg^{\succ_{\omega}}$, de la comparaison par paire $abd \succ_{\omega} abc$ (non admis) où :
 - $abd \succ_{\omega} abc$ se déduit, par application du mécanisme $2-Cov^{\succ_{\omega}}$, des comparaisons par paire $d \succ_{\omega} ab$ (admise) et $ab \succ_{\omega} c$ (admise).

$$\underline{bd \succ_{\omega} bc} \wedge \underline{bc \succ_{\omega} ad} \wedge \underline{ad \succ_{\omega} ac} \implies bd \succ_{\omega} ac \quad (3-Tr^{\succ_{\omega}})$$

avec :

$$\underline{\underline{abd \succ_{\omega} abc}} \implies \underline{bd \succ_{\omega} bc} \quad (b-Cg^{\succ_{\omega}})$$

et

$$\underline{\underline{abd \succ_{\omega} abc}} \implies \underline{ad \succ_{\omega} ac} \quad (a-Cg^{\succ_{\omega}})$$

avec :

$$\underline{d \succ_{\omega} ab} \wedge \underline{ab \succ_{\omega} c} \implies \underline{\underline{abd \succ_{\omega} abc}} \quad (2-Cov^{\succ_{\omega}})$$

FIGURE 2.5 – Explication de $bd \succ_{\omega} ac$ sous forme d'implications logiques

La représentation sous forme de diagramme de l'explication de $bd \succ_{\omega} ca$ est fournie à la Figure 2.6.

L'explication de $bd \succ_{\omega} ca$ a été réalisée dans le contexte $(\omega, \mathbb{B}_{\succ_{\omega}}, tr/cg/cov)$. Cette explication est donc de niveau 2.

Il importe de remarquer ici qu'une comparaison par paire n'admet pas une explication unique. En effet, comme on peut s'en apercevoir sur la Figure 2.7, la comparaison par paire $bd \succ_{\omega} ac$ peut également s'expliquer suivant le schéma cov/tr de niveau 1.

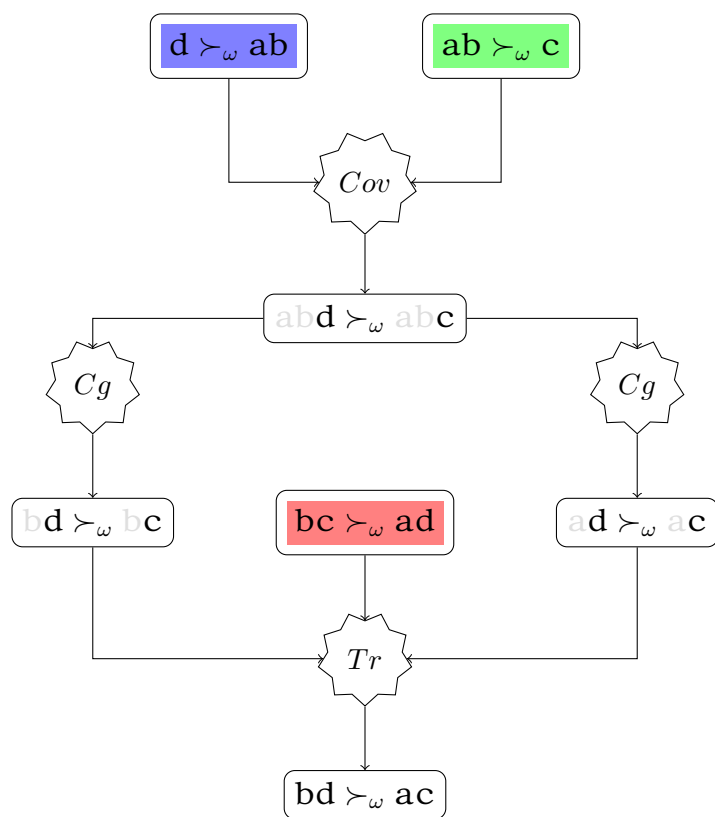


FIGURE 2.6 – Explication de $bd \gamma_\omega ac$ sous forme de diagramme

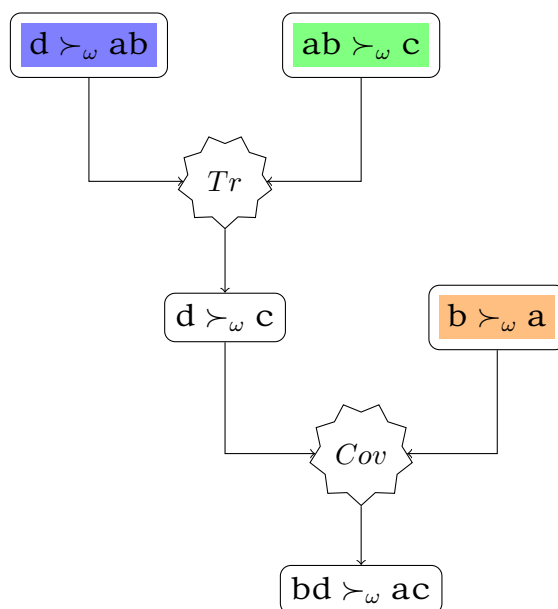


FIGURE 2.7 – Explication (alternative) de $bd \gamma_\omega ac$ sous forme de diagramme

2.4 Conclusion et perspectives

Ce chapitre développe une nouvelle approche de l'explication fondée sur les axiomes de la relation binaire représentant les préférences d'un décideur humain. La relation de préférences utilisée pour rendre compte de l'originalité de la proposition est la relation d'ordre linéaire additive. Les axiomes (propriétés) retenues sont la transitivité, l'indépendance et la couverture. À partir de ces dernières, nous avons défini des mécanismes de déduction qui ont été à la base de la construction de schémas déductifs : l'explication de la comparaison par paire (x, y) consistant dès lors à expliciter, à travers un schéma donné, les mécanismes en jeu dans sa déduction ; une telle déduction étant réalisée à partir de comparaisons par paire considérées comme admises. Enfin, nous avons proposé différentes représentations équivalentes de l'explication : l'une en langage naturel, les deux autres, sous la forme de formule(s) logique(s) et de diagramme.

Dans son esprit, cette proposition est proche des travaux de [Belahcène et al., 2017, Belahcène et al., 2019] en décision multicritère et de ceux de [Boixel and Endriss, 2020] en Choix Social Computationnel. Dans ces travaux, l'explication est conçue comme une preuve de la déduction d'une comparaison ou de l'élection d'un candidat ou plusieurs candidats à l'issue d'un processus de vote. Dans le prochain chapitre, nous proposerons quelques algorithmes dédiés au calcul de l'explication suivant un schéma déductif donné. Ces algorithmes vont ensuite nous servir à évaluer empiriquement le pouvoir explicatif des propriétés sur lesquelles se fondent les différents schémas déductifs décrits.

Chapitre 3

Pouvoir déductif de schémas d'explication

Résumé

Ce chapitre est essentiellement à visée expérimentale. L'objectif poursuivi est la mesure du pouvoir explicatif des schémas déductifs définis à partir des propriétés de transitivité, d'indépendance et de couverture des relations d'ordre linéaire additives. Pour ce faire, nous avons décrit trois grandes familles d'algorithmes. Chaque famille est spécifique au mécanisme utilisé dans la déduction de la conclusion à expliquer. Nous avons implémenté quinze algorithmes (cinq par famille) qui nous ont permis de mesurer le pouvoir explicatif des schémas déductifs auxquels ils se rapportent. Les expérimentations numériques que nous avons conduites ont porté sur l'ensemble de toutes les relations d'ordre linéaire additives décrites sur 4, 5 et 6 items. L'ensemble des comparaisons par paire explicatives retenu est essentiellement composé des paires critiques de chacune de ces relations binaires. La principale mesure réalisée est la distribution, sur l'ensemble des ordres linéaires additifs considérés, des proportions de comparaisons par paire disjointes non explicables à partir des algorithmes implémentés.

Par l'intermédiaire d'une procédure appelée "évolution dynamique du contexte", il nous a été possible de mesurer le pouvoir déductif des propriétés de transitivité, d'indépendance et de couverture des relations d'ordre linéaire additives. À ce sujet, l'un des principaux résultats du chapitre est que : la combinaison de ces trois propriétés suffit à expliquer toute comparaison par paire de toute relation d'ordre linéaire additive définie sur $m = 4, 5, 6$ items à partir des paires critiques de cette dernière.

Sommaire

| | | |
|------------|--|-----------|
| 3.1 | Introduction | 65 |
| 3.2 | Description générale d'algorithmes de calcul d'explication . . . | 66 |
| 3.2.1 | Le cas des algorithmes de la famille $\mathcal{T}r$ | 67 |
| 3.2.2 | Le cas des algorithmes de la famille $\mathcal{C}g$ | 68 |
| 3.2.3 | Le cas des algorithmes de la famille $\mathcal{C}or$ | 69 |
| 3.2.4 | Les algorithmes implémentés | 72 |
| 3.3 | Évaluation empirique du pouvoir déductif de schémas d'explication | 73 |
| 3.3.1 | Définitions et rappels | 74 |
| 3.3.2 | Protocole expérimental et résultats | 77 |
| 3.4 | Évaluation du pouvoir déductif des propriétés des ROLA | 83 |
| 3.4.1 | Évolution dynamique du contexte | 84 |
| 3.4.2 | Application | 87 |
| 3.5 | Conclusion et perspectives | 88 |

3.1 Introduction

Dans le précédent chapitre, nous avons développé une nouvelle approche de l'explication basée sur les axiomes d'une relation binaire représentant des préférences sur des ensembles d'items. Selon cette approche, l'explication est assimilée à une opération de déduction de la comparaison par paire expliquée à partir d'autres comparaisons par paire considérées comme admises. Elle est délivrée dans un *contexte* qui précise le *schéma* suivant lequel s'effectue la déduction. Nous avons distingué deux catégories de schémas déductifs : les schémas déductifs de base (ou de niveau 0) qui découlent de la mise en œuvre d'un unique mécanisme de déduction et les schémas déductifs composés (de niveau supérieur ou égal à 1) qui mettent en œuvre plusieurs mécanismes de déduction. Nous avons retenu trois mécanismes de déduction qui dérivent de l'application des axiomes de transitivité, de congruence et de couverture des relations d'ordre linéaire additives [Fishburn et al., 2002].

L'objectif de ce chapitre est de mesurer empiriquement le pouvoir explicatif des schémas déductifs que nous avons définis. Pour ce faire, nous avons implémenté quelques algorithmes qui calculent l'explication d'une comparaison par paire suivant un schéma spécifique (Table 3.1). Nous avons organisé ces algorithmes en familles en fonction du mécanisme utilisé dans la déduction de la conclusion à expliquer. La Section 3.2 propose une description générale de ces familles d'algorithmes tandis que la Section 3.3 décrit le protocole expérimental mis en place pour réaliser la mesure du pouvoir explicatif des schémas auxquels ils se rapportent.

Nous avons souhaité que ces mesures expérimentales soient exhaustives de sorte que les résultats numériques qui en découlent puissent refléter le véritable potentiel explicatif des schémas déductifs définis. C'est cette quête d'exhaustivité qui nous a amené, à nous servir des relations d'ordre linéaire additives pour illustrer notre nouvelle approche de l'explication. En effet, pour de petites valeurs de m (le nombre total d'items considérés), il est possible de générer l'intégralité de toutes les relations d'ordre linéaires additives. D'autre part, comme on sait (grâce à [Fishburn et al., 2002]) que l'ensemble des *comparaisons par paire critiques* de chaque relation d'ordre linéaire additive caractérise cette dernière, le pouvoir explicatif des schémas que nous avons définis peut être approché – en considérant comme comparaisons par paire admises (hors dominance de Pareto) lesdites paires critiques – par la proportion des éléments de cette relation qui sont explicables à partir des algorithmes que nous avons implémentés. Ces algorithmes offrent par ailleurs, l'opportunité de mesurer le pouvoir déductif des (seuls) trois axiomes des relations d'ordre linéaire additives que nous avons considérés à savoir la transitivité, la congruence et la couverture. À ce sujet, l'un des principaux résultats de ce chapitre peut être résumé comme suit : la combinaison de ces trois axiomes suffit à expliquer toute comparaison par paire de toute relation d'ordre linéaire additive définie sur $m = 4, 5, 6$ items à partir des paires critiques de cette dernière. Nous sommes parvenus à ce résul-

tat grâce à la mise en œuvre algorithmique d'une *évolution dynamique du contexte*, un concept introduit, illustré et décrit dans la Section 3.4.

3.2 Description générale d'algorithmes de calcul d'explication

L'objectif de cette section est de proposer une description générale d'algorithmes pouvant permettre de calculer l'explication de l'appartenance de la comparaison par paire (x, y) à la relation d'ordre linéaire additive \succ à partir de comparaison(s) par paire considérées comme admises (celles de la base de référence et celles exprimant une dominance de Pareto). Pour rappel, une telle explication consiste à déduire, par application d'une des **3 propriétés (transitivité, congruence et couverture)** de la relation \succ définie sur $2^{[m]}$, (x, y) à partir desdites comparaisons par paire admises et/ou une ou plusieurs autres non-admises mais explicables.

Nous allons, pour ce faire, considérer **3 familles d'algorithmes** notées \mathcal{T}_r , \mathcal{C}_g et \mathcal{C}_{or} suivant que la déduction de (x, y) dérive de l'application de la propriété de transitivité, de congruence ou de couverture. Comme les algorithmes d'une même famille s'implémentent de façon analogue, nous nous contenterons ici de décrire leurs points communs. Ces points communs sont :

- pour la famille \mathcal{T}_r , la recherche d'un chemin dans un graphe à définir,
- pour la famille \mathcal{C}_g , la vérification de l'explicabilité d'un ensemble de comparaisons par paire "équivalentes" à celle à expliquer et
- pour la famille \mathcal{C}_{or} , la résolution d'un programme linéaire en nombres entiers (PLNE).

La description que nous proposerons ici, portera donc sur la structure commune à tous les algorithmes d'une même famille. Cependant, quelle que soit sa famille, chaque algorithme prendra en entrées :

- (i) la comparaison par paire (x, y) à expliquer,
- (ii) l'ensemble des comparaisons par paire de la relation d'ordre linéaire additive \succ considérée ou tout simplement la représentation de cette dernière sous la forme de la fonction de score ω_\succ et
- (iii) le sous-ensemble de la relation \succ servant de base de référence.

En sortie de chaque algorithme, nous aurons un booléen indiquant si une explication a pu être calculée et les détails de celle-ci le cas échéant.

L'organisation de la section se présente comme suit : dans les sous-sections 3.2.1, 3.2.2 et 3.2.3, l'ossature des algorithmes des différentes familles sera détaillée. Dans la dernière

sous-section, nous indiquerons les caractéristiques des algorithmes que nous avons implémentés dans le cadre de ce travail et qui ont permis de réaliser les expérimentations numériques dont la présentation des résultats fait l'objet de la Section 3.3.

3.2.1 Le cas des algorithmes de la famille \mathcal{T}_r

Les algorithmes de la famille \mathcal{T}_r vont permettre de calculer des explications conformes au schéma de base tr (Définition 2.4) et aux schémas composés de la forme tr/sch (Définition 2.7). Ces algorithmes consisteront à rechercher un chemin dans un graphe orienté appelé *Graphe d'explication* et défini comme suit :

Définition 3.1 (Graphe d'explication $\mathcal{G}_{\mathbb{B}, \mathcal{S}}^{(x, y)}$) Soit $\mathbb{B} = \{(u_j, v_j)\}_{j \in \llbracket 1; k \rrbracket}$, un sous-ensemble d'une relation d'ordre linéaire additive \succ à partir de laquelle on déduit l'appartenance de $(x, y) \dot{\succ}$. Soit \mathbf{Algo}_{sch} , un algorithme permettant de calculer l'explication d'une comparaison par paire de \succ conforme au schéma sch .

On appelle **graphe d'explication** de la comparaison par paire (x, y) , le graphe noté $\mathcal{G}_{\mathbb{B}, \{sch\}}^{(x, y)}$ dont :

- l'ensemble V des sommets est composé de toutes les alternatives composant les comparaisons par paire de la base de référence \mathbb{B} en plus de x et y :

$$V = \bigcup_{j \in \llbracket 1; n \rrbracket} \{u_j, v_j\} \cup \{x, y\}$$

- l'ensemble $E = E_1 \cup E_2 \cup E_3 \cup E_4$ des arcs est tel que :
 - $E_1 = \mathbb{B}$.
 - $E_2 = \{(x, u_j) \in \succ : j \in \llbracket 1; k \rrbracket\}$ l'ensemble des arcs vérifiant $(x, u_j) \in \mathbb{D}_m$ ou tel que l'appel $\mathbf{Algo}_{sch}((x, u_j), \succ, \mathbb{B})$ retourne *True*.
 - $E_3 = \{(v_j, y) \in \succ : j \in \llbracket 1; k \rrbracket\}$ l'ensemble des arcs vérifiant $(v_j, y) \in \mathbb{D}_m$ ou tel que l'appel $\mathbf{Algo}_{sch}((v_j, y), \succ, \mathbb{B})$ retourne *True*.
 - $E_4 = \{(v_{j_1}, u_{j_2}) \in \succ : j_1, j_2 \in \llbracket 1; k \rrbracket \text{ avec } j_1 \neq j_2\}$ l'ensemble des arcs vérifiant $(v_{j_1}, u_{j_2}) \in \mathbb{D}_m$ ou tel que l'appel $\mathbf{Algo}_{sch}((v_{j_1}, u_{j_2}), \succ, \mathbb{B})$ retourne *True*.

Les algorithmes de la famille \mathcal{T}_r appliqueront à ce graphe orienté (sans circuit), un algorithme classique de recherche de chemin entre les sommets-alternatives x et y : si il existe un tel chemin, alors (x, y) est explicable dans le contexte (\succ, \mathbb{B}, tr) ou dans le contexte $(\succ, \mathbb{B}, tr/sch)$. Dans ce cas, l'algorithme appelé retournera *True* et les détails de l'explication (séquence de comparaisons par paire sur le chemin trouvé complétée le cas échéant des explications de comparaisons par paire non admises retournées par l'algorithme – des familles \mathcal{C}_g ou $Cor - \mathbf{Algo}_{sch}$). Nous rappelons que la recherche d'un

chemin entre deux sommets d'un graphe peut s'effectuer en réalisant un parcours de celui-ci.

En guise d'illustration, considérons l'algorithme Algo_{tr} consistant à calculer l'explication d'une comparaison par paire à l'aide du schéma tr . Dans le Paragraphe 2.3.3, on a pu montrer que la comparaison par paire (d, c) est explicable dans le contexte $(\succ_{\omega}, \mathbb{B}_{\succ_{\omega}}, tr)$ où \succ_{ω} est la relation d'ordre linéaire additive représentée par la fonction de score de l'Équation 2.5 et

$$\mathbb{B}_{\succ_{\omega}} = \{ \text{d} \succ_{\omega} \text{ab} \bullet \text{ab} \succ_{\omega} \text{c} \bullet \text{bc} \succ_{\omega} \text{ad} \bullet \text{b} \succ_{\omega} \text{a} \}$$

Algo_{tr} recherchera donc dans le graphe d'explication $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ ci-dessous un chemin entre les sommets-alternatives d et c .

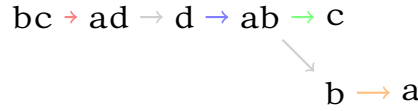


FIGURE 3.1 – Graphe d'explication $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ de $d \succ_{\omega} c$ (Paragraphe 2.3.3)

Les arcs de $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ sont colorés de couleurs permettant de mettre en évidence les comparaisons par paire correspondantes dans $\mathbb{B}_{\succ_{\omega}}$. Les arcs de couleur grise sont des comparaisons par paire (admisses) exprimant une dominance de Pareto.

3.2.2 Le cas des algorithmes de la famille \mathcal{C}_g

Les algorithmes de la famille \mathcal{C}_g vont permettre de calculer des explications conformes au schéma de base cg (Définition 2.6) et aux schémas composés de la forme cg/sch (Définition 2.9).

L'algorithme que l'on notera ici Algo_{cg} destiné à calculer l'explication d'une comparaison par paire à l'aide du schéma cg consistera à vérifier si cette dernière est considérée comme admise ou si elle est "équivalente" à une autre comparaison par paire qui, elle, est considérée comme admise. Si la comparaison par paire à expliquer est (x, y) , il s'agira donc de vérifier si $y \subset x$ ou si (x, y) est "équivalente" à une comparaison par paire de la base de référence. L'"équivalence" entre deux comparaisons par paire à laquelle nous faisons allusion ici est bien entendu, celle résultant de l'application de la propriété d'*indépendance* du modèle additif. Dans la suite, nous utiliserons le qualificatif *congruentes* pour désigner deux telles comparaisons par paire. De façon plus formelle, nous considérons la définition suivante :

Définition 3.2 (Comparaisons par paire congruentes) Deux comparaisons par paire (x, y) et (\tilde{x}, \tilde{y}) sont dites congruentes si $x \setminus y = \tilde{x} \setminus \tilde{y}$ et $y \setminus x = \tilde{y} \setminus \tilde{x}$ avec $x \neq \tilde{x}$ et $y \neq \tilde{y}$.

Il existe $2^{|neu(x,y)|} - 1$ comparaisons par paire congruentes à (x, y) où $neu(x, y) = [m] \setminus (x \setminus y \cup y \setminus x)$ est l'ensemble des items communs à x et y ou n'appartenant ni à x ni à y .

L'Algorithme générique 1 décrit la structure des autres algorithmes de la famille \mathcal{C}_g . Ceux-ci permettent de calculer l'explication de la comparaison par paire (x, y) à l'aide du schéma composé cg/sch . Ils s'appuient sur un algorithme \mathbf{Algo}_{sch} des familles \mathcal{T}_r ou \mathcal{C}_or calculant des explications conformes au schéma sch .

Algorithme 1 Algorithme générique calculant des explications conformes au schéma composé cg/sch

```

1: pour toute paire  $(\tilde{x}, \tilde{y})$  congruente à  $(x, y)$  faire
2:    $explainable, details \leftarrow \mathbf{Algo}_{sch}((\tilde{x}, \tilde{y}), \succ, \mathbb{B})$ 
3:   si  $explainable$  alors
4:     retourner  $\mathbf{True}, ((\tilde{x}, \tilde{y}), details)$ 
5:   sinon
6:     retourner  $\mathbf{False}, \mathbf{None}$ 
7:   fin du si
8: fin du pour

```

En guise d'illustration, considérons l'algorithme noté $\mathbf{Algo}_{cg/tr}$ s'appuyant sur l'algorithme \mathbf{Algo}_{tr} décrit dans la sous-section 3.2.1. Cet algorithme ($\mathbf{Algo}_{cg/tr}$) calcule l'explication de comparaison par paire explicable à l'aide du schéma cg/tr . Appliqué à la comparaison par paire (c, a) du Paragraphe 2.3.3, cet algorithme énumère les $2^2 - 1$ comparaisons par paire congruentes à (c, a) . Il s'agit de : (bc, ba) (cd, ad) (bcd, bad) .

La ligne 2 de l'algorithme générique 1 indique que plusieurs appels à l'algorithme \mathbf{Algo}_{tr} pourront être réalisés dont, en particulier, l'appel $\mathbf{Algo}_{tr}((bc, ab), \succ_\omega, \mathbb{B})$ qui retournera \mathbf{True} car comme on peut le voir sur le *graphe d'explication* $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ de la Figure 3.1 (identique au graphe $\mathcal{G}_{\mathbb{B}, \emptyset}^{(bc, ab)}$), il existe un chemin entre bc et ab .

3.2.3 Le cas des algorithmes de la famille \mathcal{C}_or

Les algorithmes de la famille \mathcal{C}_or vont principalement¹ permettre de calculer des explications conformes au schéma de base cov (Définition 2.5) et aux schémas composés de la forme cov/sch (Définition 2.8). La description de cette famille d'algorithmes nécessite que l'on introduise la définition suivante :

1. Ces algorithmes vont également permettre de capturer certains cas particuliers d'explication conforme aux schémas tr/cov et $tr/cov/sch$.

Définition 3.3 (Complémentaire de (x, y) par rapport à \mathbb{F}) Soit (x, y) une comparaison par paire appartenant à une relation d'ordre linéaire additive \succ .

Soit $\mathbb{F} = \{(u_j, v_j)\}_{j \in \llbracket 1; k \rrbracket}$ un ensemble de k comparaison(s) par paire ($k \geq 1$) de \succ telle que :

$$\bigcap_{j \in \llbracket 1; k \rrbracket} u_j = \bigcap_{j \in \llbracket 1; k \rrbracket} v_j = \emptyset \quad (3.1)$$

$$\bigcup_{j \in \llbracket 1; k \rrbracket} u_j \subseteq x \quad (3.2)$$

On appelle **complémentaire** de (x, y) par rapport à \mathbb{F} la comparaison par paire (x', y') appartenant à \succ et telle que :

$$(x', y') = (x \setminus \bigcup_{j \in \llbracket 1; k \rrbracket} u_j, y \setminus \bigcup_{j \in \llbracket 1; k \rrbracket} v_j)$$

La **cardinalité** du complémentaire (x', y') de (x, y) est : $|x'| + |y'|$.

Comme indiqué en introduction de section, ces algorithmes vont consister à résoudre un programme linéaire en nombres entiers (PLNE). Ce programme va permettre de déterminer un sous-ensemble \mathbb{F} de comparaison(s) par paire considérées comme admises et vérifiant les Équations (3.1) et (3.2) de sorte à **minimiser la cardinalité de la comparaison par paire (x', y') complémentaire de la comparaison (x, y) à expliquer**. Si cette cardinalité optimale est nulle, alors par définition (Définition 2.5), $(\bigcup_{j \in \llbracket 1; k \rrbracket} u_j, \bigcup_{j \in \llbracket 1; k \rrbracket} v_j) = (x, \bigcup_{j \in \llbracket 1; k \rrbracket} v_j)$ est explicable dans le contexte $(\succ, \mathbb{B}, cov)^2$ (où \mathbb{B} fait office de base de référence). Dans ce qui suit, on considérera que la base de référence s'écrit : $\mathbb{B} = \{(u_j, v_j)\}_{j \in \llbracket 1; k \rrbracket}$ et que la relation d'ordre linéaire additive est représentée par sa fonction de score ω .

Soit \mathcal{P} , le PLNE générique décrit comme suit :

Variables de décision. Deux catégories de variables sont à distinguer :

- Des variables booléennes b_j ($j \in \llbracket 1; k \rrbracket$) associées aux comparaisons par paire de la base de référence \mathbb{B} . Une variable b_j vaut 1 si et seulement la comparaison par paire correspondante (u_j, v_j) contribue à l'explication de (x, y) .
- Des variables booléennes c_i^x (respectivement c_i^y) associées à chaque item i composant x (respectivement y). La variable c_i^x (respectivement c_i^y) vaut 1 si et seulement si l'item i appartient à une alternative u_j (respectivement v_j) où (u_j, v_j) est une comparaison par paire qui contribue à l'explication de (x, y) .

2. Lorsque $y \subset \bigcup_{j \in \llbracket 1; k \rrbracket} v_j$, l'explication calculée est conforme au schéma tr/cov . Autrement (égalité de ces deux ensembles), (x, y) est explicable dans le contexte (\succ, \mathbb{B}, cov) .

Contraintes. On distinguera 3 catégories de groupes de contraintes :

- Les contraintes garantissant les conditions (3.1) et (3.2) de la Définition 3.3 et reliant les différents types de variables. Ce sont :

$$\sum_{j \in \llbracket 1; k \rrbracket : i \in u_j} b_j = 0 \quad \text{pour tout } i \notin x \quad (3.3)$$

$$\sum_{j \in \llbracket 1; k \rrbracket : i \in u_j} b_j = c_i^x \quad \text{pour tout } i \in x \quad (3.4)$$

$$\sum_{j \in \llbracket 1; k \rrbracket : i \in v_j} b_j \leq 1 \quad \text{pour tout } i \notin y \quad (3.5)$$

$$\sum_{j \in \llbracket 1; k \rrbracket : i \in v_j} b_j = c_i^y \quad \text{pour tout } i \in y \quad (3.6)$$

- La contrainte garantissant l'appartenance à \succ du complémentaire à (x, y) par rapport au sous-ensemble de comparaisons par paire à sélectionner :

$$\sum_{i_1 \in x} (1 - c_{i_1}^x) \times \omega_{i_1} \geq \sum_{i_2 \in y} (1 - c_{i_2}^y) \times \omega_{i_2} \quad (3.7)$$

- La contrainte de non-trivialité :

$$\sum_{j \in \llbracket 1; k \rrbracket} b_j \geq 1 \quad (3.8)$$

Fonction objectif.

$$\sum_{i_1 \in x} c_{i_1}^x + \sum_{i_2 \in x} c_{i_2}^y \quad (\text{à maximiser}) \quad (3.9)$$

Récupération de la solution. Lorsque le programme linéaire en nombres entiers décrit ci-dessus admet une solution, les valeurs b_j ($j \in \llbracket 1; k \rrbracket$) permettent d'identifier les comparaisons par paire \mathbb{F} participant à l'explication de (x, y) tandis que les valeurs des variables c_i^x et c_i^y nulles permettent de reconstruire le complémentaire de (x, y) par rapport à \mathbb{F} le cas échéant.

Lorsque toutes les valeurs c_i^y à l'optimum valent 1 alors (x, y) est explicable dans le contexte (\succ, \mathbb{B}, cov) ³ ou $(\succ, \mathbb{B}, tr/cov)$ ⁴ Cependant, lorsque la comparaison par paire (x', y') complémentaire à (x, y) n'est pas admise c'est-à-dire n'appartient pas à \mathbb{B} , les

3. Le cas où toutes les contraintes (3.5) sont non serrées (somme nulle).

4. Lorsqu'au moins une contrainte (3.5) est saturée (somme égale à 1).

autres algorithmes de la famille Cor (permettant principalement de calculer l'explication de la comparaison par paire (x, y) suivant le schéma composé cov/sch) consisteront à réaliser l'appel $\text{Algo}_{sch}((x', y'), \succ_\omega, \mathbb{B})$ où Algo_{sch} est un algorithme des familles \mathcal{T} ou \mathcal{C}_g calculant des explications conformes aux schémas sch .

En guise d'illustration, considérons l'algorithme noté $\text{Algo}_{cov/tr}$ qui utilise l'algorithme Algo_{tr} décrit dans la sous-section 3.2.1. Dans le Paragraphe 2.3.3, on a pu montrer que la comparaison par paire (bd, ac) est explicable dans le contexte $(\succ_\omega, \mathbb{B}_{\succ_\omega}, cov/tr)$ où \succ_ω est la relation d'ordre linéaire additive représentée par la fonction de score de l'Équation 2.5 et

$$\mathbb{B}_{\succ_\omega} = \{ \mathbf{d} \succ_\omega \mathbf{ab} \bullet \mathbf{ab} \succ_\omega \mathbf{c} \bullet \mathbf{bc} \succ_\omega \mathbf{ad} \bullet \mathbf{b} \succ_\omega \mathbf{a} \}$$

L'exécution du programme \mathcal{P} décrit précédemment sélectionnera uniquement la comparaison par paire $\mathbf{b} \succ_\omega \mathbf{a}$. En effet, les items a et c n'appartiennent pas à bd ; ce qui disqualifie $\mathbf{ab} \succ_\omega \mathbf{c}$ et $\mathbf{bc} \succ_\omega \mathbf{ad}$. Aussi, le complémentaire de (bd, ac) par rapport à l'ensemble singleton $\{ \mathbf{d} \succ_\omega \mathbf{ab} \}$ (si il existait) serait (b, c) ; or (b, c) n'appartient pas à $\succ_\omega : \omega(\mathbf{b}) = 6 < 8 = \omega(\mathbf{c})$.

L'étape suivante de l'exécution de l'appel $\text{Algo}_{cov/tr}((bd, ac), \succ_\omega, \mathbb{B})$ consiste à réaliser l'appel $\text{Algo}_{tr}((d, c), \succ_\omega, \mathbb{B})$ qui retournera `True` vu qu'il existe un chemin entre d et c dans le graphe d'explication $\mathcal{G}_{\mathbb{B}, \emptyset}^{(d, c)}$ (voir Figure 3.1).

3.2.4 Les algorithmes implémentés

La Table 3.1 liste les différents algorithmes implémentés dans le cadre de ce travail. Ces implémentations réalisées en Python obéissent aux descriptions données en Section 3.2. En tout, quinze (15) algorithmes ont été codés. Le code correspondant est accessible sur le dépôt que voici⁵. Comme le montre la Table 3.1, il s'agit d'algorithmes de calcul d'explications conformes aux 3 schémas de base et aux 12 schémas composés de la forme sch_1/sch_2 (en tout 6) et $sch_1/sch_2/sch_3$ (en tout 6 aussi) où les schémas de base sch_i , $i \in \{1, 2, 3\}$ sont tous différents.

Ces algorithmes vont nous permettre de mesurer le pouvoir explicatif des schémas de la colonne « **Type de schéma concerné** ».

Pour ce faire, il nous faudra aussi retenir l'ensemble des comparaisons par paire devant servir de base de référence. Dans la sous-section 3.3.1 de la Section 3.3, nous nommerons et décrirons les bases de référence que nous considérerons à travers les expérimentations numériques réalisées dans ce chapitre. Quant à la sous-section 3.3.2, elle détaillera le protocole expérimental mis en place et commentera les résultats obtenus.

5. <https://github.com/tetras92/SchemasDeductifsEnAMCD.git>

TABLE 3.1 – Ensemble des algorithmes implémentés

| n° | Nom | Type de schéma concerné | Niveau d'explication correspondant | Famille |
|-----------|---------------------------|-------------------------|------------------------------------|-----------------|
| 1 | Algo_{tr} | tr | 0 | \mathcal{T}_r |
| 2 | Algo_{cg} | cg | 0 | \mathcal{C}_g |
| 3 | Algo_{cov} | cov | 0 | $\mathcal{C}or$ |
| 4 | $\text{Algo}_{tr/cg}$ | tr/cg | 1 | \mathcal{T}_r |
| 5 | $\text{Algo}_{tr/cov}$ | tr/cov | 1 | \mathcal{T}_r |
| 6 | $\text{Algo}_{cg/tr}$ | cg/tr | 1 | \mathcal{C}_g |
| 7 | $\text{Algo}_{cg/cov}$ | cg/cov | 1 | \mathcal{C}_g |
| 8 | $\text{Algo}_{cov/tr}$ | cov/tr | 1 | $\mathcal{C}or$ |
| 9 | $\text{Algo}_{cov/cg}$ | cov/cg | 1 | $\mathcal{C}or$ |
| 10 | $\text{Algo}_{tr/cg/cov}$ | $tr/cg/cov$ | 2 | \mathcal{T}_r |
| 11 | $\text{Algo}_{tr/cov/cg}$ | $tr/cov/cg$ | 2 | \mathcal{T}_r |
| 12 | $\text{Algo}_{cg/tr/cov}$ | $cg/tr/cov$ | 2 | \mathcal{C}_g |
| 13 | $\text{Algo}_{cg/cov/cg}$ | $cg/cov/cg$ | 2 | \mathcal{C}_g |
| 14 | $\text{Algo}_{cov/tr/cg}$ | $cov/tr/cg$ | 2 | $\mathcal{C}or$ |
| 15 | $\text{Algo}_{cov/cg/tr}$ | $cov/cg/tr$ | 2 | $\mathcal{C}or$ |

3.3 Évaluation empirique du pouvoir déductif de schémas d'explication

Dans le précédent chapitre (sous-section 2.3.2), nous avons décrit différents schémas déductifs permettant de justifier l'appartenance d'une comparaison par paire (x, y) à une relation d'ordre linéaire additive \succ à partir d'autres comparaisons par paire appartenant à la même relation. Ainsi, étant donné un sous-ensemble \mathbb{B}_\succ de la relation \succ et un schéma déductif sch , nous avons indiqué (lorsque l'explication correspondante existe) que la comparaison par paire (x, y) s'explique à partir de \mathbb{B}_\succ et suivant le schéma sch et noté : $(x, y) \in \mathcal{E}(\succ, \mathbb{B}_\succ, sch)$. Le triplet $(\succ, \mathbb{B}_\succ, sch)$ a été désigné sous le vocable de *contexte*. Le *contexte* rassemble donc tous les éléments permettant d'explicitier la déduction de la comparaison par paire (x, y) à partir des comparaisons par paire de \mathbb{B}_\succ et suivant le schéma sch .

Dans la précédente section, nous avons proposé une description générale de la structure de différents algorithmes regroupés en familles (\mathcal{T}_r , \mathcal{C}_g et $\mathcal{C}or$) qui permettent de calculer des explications conformes à certains schémas déductifs définis en sous-section 2.3.2. Quelques algorithmes de ces familles ont été implémentés et listés dans la Table 3.1. Ces implémentations nous offrent l'occasion d'évaluer empiriquement la capacité déductive des ensembles de schémas auxquels elles se rapportent. Cette évaluation peut être quantitative, s'exprimant sous la forme du ratio :

$$\frac{|\mathcal{E}(\succ, \mathbb{B}_\succ, \mathcal{S})|}{|\mathcal{D}(\succ, \mathbb{B}_\succ)|} \quad (3.10)$$

où $|\mathbb{E}|$ désigne le cardinal de l'ensemble de comparaisons par paire \mathbb{E} , $\mathcal{D}(\succ, \mathbb{B}_\succ)$ est l'ensemble des comparaisons par paire de \succ pouvant être déduites à partir de \mathbb{B}_\succ par

application de l'**Axiome 3** de la définition (axiomatique) des relations d'ordre linéaire additives (voir sous-section 2.2.1) ou de façon équivalente à travers la résolution d'un programme linéaire en nombres entiers décrits dans [Fishburn, 1996] et $\mathcal{E}(\succ, \mathbb{B}_\succ, \mathcal{S})$, un abus de notation de l'ensemble des comparaisons par paire explicables dans l'ensemble des contextes $(\succ, \mathbb{B}_\succ, sch)$, $sch \in \mathcal{S}$.

L'objet de la présente section est de réaliser de façon exhaustive une telle évaluation quantitative; cette exhaustivité revêtant plusieurs aspects. D'une part, nous souhaitons mesurer le pouvoir déductif des différents schémas répertoriés dans la Table 3.1 sur l'ensemble des relations d'ordre linéaire additives définies sur $2^{[m]}$ et d'autre part, le faire à partir d'une base de référence à partir de laquelle se déduit (par application de l'**Axiome 3**) l'intégralité de la relation \succ considérée : il s'agit donc de prendre \mathbb{B}_\succ telle que $\mathcal{D}(\succ, \mathbb{B}_\succ)$ soit égal à \succ .

Alors que le besoin d'exhaustivité relative à l'ensemble des ROLA à considérer nous obligera à nous restreindre à de "petites" valeurs⁶ de nombre d'items m , celui de l'identification d'un sous-ensemble \mathbb{B}_\succ "général" intégralement \succ reste surmontable. En effet, [Fishburn et al., 2002] définit et décrit l'ensemble des *paires critiques*⁷ de la relation \succ comme déterminant intégralement cette dernière. Dans la sous-section 3.3.1, nous rappellerons ce qu'est une (*comparaison par*) *paire critique* et fournirons quelques données chiffrées sur l'ensemble des comparaisons par paire disjointes auquel elle appartient. Nous y rappellerons également les cardinalités (connues) des ensembles des relations d'ordre linéaire additives définies sur $2^{[m]}$ avec $m = 4, 5, 6, 7$ items. Comme l'évaluation quantitative à réaliser est empirique et repose sur les algorithmes listés dans la Table 3.1, nous concluons cette sous-section par la définition des ensembles de comparaisons par paire que ces algorithmes permettent d'expliquer. Tous ces rappels et définitions faciliteront ensuite la compréhension du protocole expérimental que nous avons mis en place ainsi que celle des résultats qui en découlent (sous-section 3.3.2).

3.3.1 Définitions et rappels

En désignant par \mathcal{R}_m l'ensemble des relations d'ordre linéaire additives \succ définies sur $2^{[m]}$, on rappelle (Table 3.2) les valeurs des quantités $|\mathcal{R}_m|/m!$ (extraites de [Fishburn et al., 2002, Maclagan, 1998]). La Table 3.2 nous apprend qu'il existe en tout 89 414 640 (= $124\,187 \times 6!$) relations d'ordre linéaire additives différentes définies sur l'ensemble

6. Il convient de noter qu'à notre connaissance, le nombre exact de relations d'ordre linéaire additives définies sur $2^{[m]}$, n'est connu que pour des valeurs de m inférieures ou égales à 7. C'est [Maclagan, 1998] qui à l'aide d'un programme récursif d'énumération, détermine ces valeurs exactes. Nous, en nous appuyant sur la définition de représentation des ROLA et en nous servant de la Programmation Linéaire en Nombres Entiers (PLNE), nous sommes arrivés à retrouver ces nombres et stocker sous la forme de séquences d'entiers naturels, l'intégralité des ROLA définies sur des ensembles d'items de tailles m inférieures ou égales à 6.

7. De l'anglais : *critical pairs*.

TABLE 3.2 – Nombre total de relations d'ordre linéaire additives en fonction de m

| m | $ \mathcal{R}_m / m!$ |
|-----|------------------------|
| 1 | 1 |
| 2 | 1 |
| 3 | 2 |
| 4 | 14 |
| 5 | 516 |
| 6 | 124 187 |
| 7 | 214 580 603 |

des parties d'un ensemble composé de $m = 6$ items. La division de \mathcal{R}_m par $m!$ dans la Table 3.2 se justifie du fait de la symétrie induite par le choix d'un ordre arbitraire sur les alternatives singletons. On remarque que le nombre total de relations d'ordre linéaire additives croît très (très) vite avec m ; ce qu'on a pu expérimenter lors de la génération exhaustive de tous les ROLA de \mathcal{R}_m à partir d'un algorithme basé sur de la Programmation Linéaire en Nombres Entiers (PLNE) (voir Annexe) : tandis que l'exécution du programme a duré moins d'une minute pour $m = 5$, il a fallu près de six heures pour qu'il termine en prenant $m = 6$.

Comme on sait (**Axiome 2** de la définition axiomatique, cf. Chapitre 2 Section 2.2) d'une part, que l'ensemble des comparaisons par paire exprimant une dominance de Pareto \mathbb{D}_m est incluse dans toute ROLA \succ définie sur $2^{[m]}$ et d'autre part, que $x \succ y \iff x \setminus y \succ y \setminus x$, chaque ROLA se distingue d'une autre suivant la façon dont elle ordonne les paires $\{u, v\}$ de sous-ensembles disjoints ($u \cap v = \emptyset$) d'items. On considérera les définitions suivantes :

Définition 3.4 (Ensembles \mathbb{T}_m , \mathbb{T}_\succ et $\mathbb{C}\mathbb{R}_\succ$) Soit m un entier naturel positif. Nous désignons par \mathbb{T}_m , l'ensemble des paires d'éléments de $2^{[m]} \setminus \emptyset$ disjointes :

$$\mathbb{T}_m = \{\{u, v\} \in 2^{[m]} \setminus \emptyset \times 2^{[m]} \setminus \emptyset : u \cap v = \emptyset\} \quad (3.11)$$

Soit \succ une relation d'ordre linéaire additive définie sur $2^{[m]}$, nous désignons par \mathbb{T}_\succ , l'ensemble des comparaisons par paire obtenues en ordonnant suivant \succ les paires de \mathbb{T}_m :

$$\mathbb{T}_\succ = \{(u, v) : u \succ v \text{ et } u \cap v = \emptyset\} \quad (3.12)$$

Enfin, l'ensemble des comparaisons par paire critiques $\mathbb{C}\mathbb{R}_\succ$ de \succ est composé des comparaisons par paire de \mathbb{T}_\succ dont les deux alternatives sont adjacentes dans l'ordre induit par \succ :

$$\mathbb{C}\mathbb{R}_\succ = \{(u, v) \in \mathbb{T}_\succ : \nexists z \in 2^{[m]}, u \succ z \text{ et } z \succ v\} \quad (3.13)$$

Exemple 3.1 Soit la ROLA $\succ_{\omega'}$ définie sur $2^{[3]}$ et représentée par la fonction de score suivante :

$$\omega' = (\mathbf{a} : 4, \mathbf{b} : 2, \mathbf{c} : 1)$$

L'ordre induit par $\succ_{\omega'}$ sur $2^{[3]}$ étant :

$$\mathbf{abc} \succ_{\omega'} \mathbf{ab} \succ_{\omega'} \mathbf{ac} \succ_{\omega'} \mathbf{a} \succ_{\omega'} \mathbf{bc} \succ_{\omega'} \mathbf{b} \succ_{\omega'} \mathbf{c} \succ_{\omega'} \emptyset$$

on a :

$$\mathbb{T}_{\succ_{\omega'}} = \{(\mathbf{a}, \mathbf{b}), (\mathbf{b}, \mathbf{c}), (\mathbf{a}, \mathbf{c}), (\mathbf{ac}, \mathbf{b}), (\mathbf{a}, \mathbf{bc}), (\mathbf{ab}, \mathbf{c})\}$$

et

$$\mathbb{CR}_{\succ_{\omega'}} = \{(\mathbf{b}, \mathbf{c}), (\mathbf{a}, \mathbf{bc})\}$$

Au sujet de la taille des ensembles \mathbb{T}_m , [Fishburn et al., 2002] montre que :

$$|\mathbb{T}_m| = \frac{3^m + 1}{2} - 2^m \quad \forall m \geq 1 \quad (3.14)$$

La Table 3.3 recense les valeurs $|\mathbb{T}_m|$ pour $m \in \llbracket 1; 7 \rrbracket$

TABLE 3.3 – Tailles des ensembles \mathbb{T}_m en fonction de m

| m | $ \mathbb{T}_m $ |
|-----|------------------|
| 1 | 0 |
| 2 | 1 |
| 3 | 6 |
| 4 | 25 |
| 5 | 90 |
| 6 | 301 |
| 7 | 966 |

Concernant les comparaisons par paire critiques, [Fishburn et al., 2002] démontre (Corollaire 3.8) que :

$$\mathbb{CR}_{\succ} \neq \mathbb{CR}_{\succ'} \iff \succ \neq \succ'$$

où \succ et \succ' sont deux ROLAs définies sur $2^{[m]}$.

En procédant par dénombrement, nous avons pu remarquer que :

- $|\mathbb{CR}_{\succ}| \in \llbracket 3; 6 \rrbracket$, pour toute ROLA \succ définie sur $2^{[4]}$; ce qui correspond à une proportion minimale (respectivement maximale) de 12% (respectivement 24%) de l'ensemble \mathbb{T}_4 .

3.3. ÉVALUATION EMPIRIQUE DU POUVOIR DÉDUCTIF DE SCHÉMAS D'EXPLICATION 77

- $|\mathbb{C}\mathbb{R}_{\succ}| \in \llbracket 4; 12 \rrbracket$, pour toute ROLA \succ définie sur $2^{[5]}$; ce qui correspond à une proportion minimale (respectivement maximale) de 4.4% (respectivement 13.3%) de l'ensemble \mathbb{T}_5 .
- $|\mathbb{C}\mathbb{R}_{\succ}| \in \llbracket 5; 21 \rrbracket$, pour toute ROLA \succ définie sur $2^{[6]}$; ce qui correspond à une proportion minimale (respectivement maximale) de 1.6% (respectivement 6.9%) de l'ensemble \mathbb{T}_6 .

Comme indiqué en début de section, les algorithmes implémentés dans le cadre de notre travail nous offrent l'occasion de mesurer le pouvoir déductif des différents schémas auxquels ils se rapportent (Table 3.1). Ce sont ces algorithmes que nous allons utiliser pour tenter d'**expliquer la déduction de chaque comparaison par paire de l'ensemble \mathbb{T}_{\succ} à partir des paires critiques $\mathbb{C}\mathbb{R}_{\succ}$** .

Soit \mathfrak{A} un ensemble d'algorithmes parmi ceux repertoriés dans la Table 3.1. Désignons par $\mathcal{S}^{\mathfrak{A}}$, l'ensemble des schémas déductifs auxquels sont conformes les structures d'explication retournées par les algorithmes de \mathfrak{A} et par $\mathcal{X}_{\text{pl}}^{\succ, \mathbb{C}\mathbb{R}_{\succ}, \mathfrak{A}}$, l'ensemble des comparaisons par paire de \mathbb{T}_{\succ} dont l'explication de la déduction a été rendue possible par au moins un algorithme de \mathfrak{A} . On a que :

$$\mathcal{X}_{\text{pl}}^{\succ, \mathbb{C}\mathbb{R}_{\succ}, \mathfrak{A}} \subseteq \bigcup_{sch \in \mathcal{S}^{\mathfrak{A}}} \mathcal{E}(\succ, \mathbb{C}\mathbb{R}_{\succ}, sch) \cap \mathbb{T}_{\succ} \subseteq \mathbb{T}_{\succ}$$

Ainsi, en comparant $|\mathcal{X}_{\text{pl}}^{\succ, \mathbb{C}\mathbb{R}_{\succ}, \mathfrak{A}}|$ à $|\mathbb{T}_{\succ}|$, on obtient une mesure (minorée) de la capacité des schémas de l'ensemble $\mathcal{S}^{\mathfrak{A}}$ à rendre compte de la déduction de l'appartenance des comparaisons par paire de \mathbb{T}_{\succ} à la relation \succ à partir de ses paires critiques $\mathbb{C}\mathbb{R}_{\succ}$.

3.3.2 Protocole expérimental et résultats

Les expérimentations numériques réalisées ici se veulent être exhaustives. En effet, à travers celles-ci, nous souhaitons décrire l'ensemble des relations d'ordre linéaire additives définies sur $2^{[m]}$ avec $m = 4, 5, 6$. Cette description nous permettra de découvrir et de comprendre comment sont construites, à partir de leurs paires critiques, les préférences basées sur ces relations et aussi d'observer leur diversité. Ces expérimentations ont été réalisées à partir de l'ensemble des relations d'ordre linéaire additives de \mathcal{R}_m ($m = 4, 5, 6$) vérifiant l'ordre (de préférence) arbitraire sur les alternatives singletons suivant :

$$\underbrace{f \succ e \succ \overbrace{d \succ c \succ b \succ a}^{m=5}}_{m=6} \quad (3.15)$$

Leur objectif, c'est la production d'une mesure du pouvoir déductif des schémas auxquels sont conformes les explications retournées par les algorithmes implémentés (Table

3.1). Dans la suite de cette sous-section, nous désignerons par \mathfrak{A}_l ($0 \leq l \leq 2$) l'ensemble des algorithmes implémentés calculant des explications de niveau au plus égal à l . Par exemple, lorsque $l = 1$, il s'agit des algorithmes (n°) 1 à 9 de la Table 3.1.

En guise de mesure, nous avons fait le choix de calculer **la fonction de répartition du nombre de comparaisons par paire dont l'explication de la déduction n'a pas été rendue possible par chaque ensemble \mathfrak{A}_l ($0 \leq l \leq 2$) d'algorithmes**. Concrètement, pour chaque valeur $m \in \{4, 5, 6\}$ et chaque valeur $l \in \{0, 1, 2\}$, nous avons calculé :

$$\mathbf{P}(\#\text{Unxp}_l \leq \mathbf{x}_m) \quad \forall \mathbf{x}_m \in \llbracket 0; |\mathbb{T}_m| \rrbracket$$

où $\#\text{Unxp}_l$ est la variable aléatoire définie comme suit :

$$\#\text{Unxp}_l = |\mathbb{T}_m| - |\mathbb{X}_{\text{rol}}^{\succ, \text{CR}_{\succ}, \mathfrak{A}_l}| \quad \text{pour tout } \succ \in \mathcal{R}_m$$

Pour des valeurs m et l données, $\mathbf{P}(\#\text{Unxp}_l \leq 0)$ correspond donc à la proportion de ROLA de \mathcal{R}_m dont toutes les comparaisons par paire disjointes (\mathbb{T}_m) ont pu être expliquées grâce aux algorithmes de l'ensemble \mathfrak{A}_l (Table 3.1). Quant au produit $|\mathcal{R}_m| \times \mathbf{P}(\#\text{Unxp}_l \leq \mathbf{x}_m)$, il nous fournit le nombre exact de relations d'ordre linéaire additives \succ dont le nombre de comparaisons par paire de \mathbb{T}_{\succ} qui sont "inexplicables" à l'aide des algorithmes de l'ensemble \mathfrak{A}_l est inférieur ou égal à \mathbf{x}_m ($\mathbf{x}_m \in \llbracket 0; |\mathbb{T}_m| \rrbracket$).

Dans les commentaires des résultats expérimentaux obtenus, nous analyserons tout particulièrement :

- Les valeurs des proportions $\mathbf{P}(\#\text{Unxp}_l \leq 0)$ qui, multipliées par $|\mathcal{R}_m|$ renseignent sur le nombre minimal de ROLA dont l'intégralité des comparaisons par paire s'explique à partir des schémas déductifs de niveau l recensés dans la Table 3.1.
- L'évolution suivant les valeurs de niveau l des indicateurs de dispersion (minimum, quartiles (Q_1 , Q_2 et Q_3) et maximum) de la variable $\#\text{Unxp}_l$.
- Les temps d'exécution qu'ont requis les différentes expérimentations. Ceux-ci, en l'absence d'une analyse rigoureuse de complexité des algorithmes proposés (Table 3.1), nous éclaireront sur les possibilités d'utilisation de ces algorithmes en situation réelle.

Le cas $m = 4$

La Table 3.4 compile les résultats obtenus sur l'ensemble des $14 \times 4! = 336$ relations d'ordre linéaire additives définies sur $2^{[4]}$. On y lit qu'environ 37,7% de ces relations sont composés de (25) comparaisons par paire disjointes qui sont toutes explicables suivant les schémas de niveau 0 répertoriés dans la Table 3.1 et à partir de leurs paires critiques. Dès que l'on considère l'ensemble des schémas déductifs de niveau 1 de la Table 3.1, on remarque ($\mathbf{P}(\#\text{Unxp}_1 \leq 0) = 1$) que toutes les comparaisons par paire de toutes les ROLA sont explicables à partir de leurs paires critiques .

TABLE 3.4 – Fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_4 inexplicables

| x_4 | 0 | 1 | 2 | 3 |
|--|----------|----------|----------|----------|
| $\mathbf{P}(\#\text{Unxp}_0 \leq x_4)$ | 37,7% | 71,4% | 85,7% | 100% |
| $\mathbf{P}(\#\text{Unxp}_1 \leq x_4)$ | 100% | - | - | - |
| $\mathbf{P}(\#\text{Unxp}_2 \leq x_4)$ | 100% | - | - | - |

Temps d'exécution :⁸ La durée totale d'exécution des algorithmes sur l'ensemble des comparaisons par paire de toutes les 14 ROLA définies sur $2^{[4]}$ est inférieure à la seconde.

Le cas $m = 5$

La Figure 3.2 présente graphiquement les trois fonctions de répartition du nombre total $x_5 \in \llbracket 0; 90 \rrbracket$ de comparaisons par paire disjointes que les algorithmes que nous avons implémentés ne sont pas arrivés à expliquer. Seules environ 1,7% des ROLA définies sur $2^{[5]}$ sont intégralement explicables à l'aide de leurs paires critiques et suivant les schémas de niveau 0 à savoir *tr*, *cg* et *cov*. En considérant également les schémas de niveau 1, cette proportion monte à 22,2%. En recourant à l'ensemble des 15 algorithmes implémentés (3.1), on a remarqué que plus de la moitié (56,2%) des ROLA définies sur $2^{[5]}$ sont intégralement explicables à partir de leurs paires critiques.

On note, en ce qui concerne les valeurs x_5^{max} minimales telles que $\mathbf{P}(\#\text{Unxp}_l \leq x_5^{max}) = 1$, elles sont de 31 lorsque $l = 0$, 16 lorsque $l \leq 1$ et de 14 lorsque $l \leq 2$. Quant aux quartiles $x_5^{Q_3}$, ils passent de 15 à 4 puis à 1 au fur et à mesure de l'introduction de schémas de niveau supérieur. C'est ce que l'on peut distinctement observer sur les diagrammes en boîte de la Figure 3.3. Ces diagrammes sont une synthèse des fonctions de répartition représentées à la Figure 3.2. Ils se concentrent sur les principaux indicateurs de dispersion des variables $\#\text{Unxp}_l$. En ce qui concerne l'évolution de la médiane $x_5^{Q_2}$, valeur des $\#\text{Unxp}_l$ divisant l'ensemble des 516 ROLA définies sur $2^{[5]}$ et vérifiant (3.15) en (à peu près) deux sous-ensembles de tailles égales, on remarquera qu'elle baisse fortement (de 12 à 2) dès l'introduction des schémas de niveau 1 puis s'annule lorsque l'on considère également les schémas de niveau 2. Enfin, pour toute ROLA \succ définie sur $2^{[5]}$, au plus 14 (sur 90) comparaisons par paire de \mathbb{T}_\succ seraient⁹ inexplicables après recours à l'ensemble des quinze algorithmes implémentés.

8. Tous les calculs ont été réalisés sur une machine Linux Intel Core i7 à partir d'un programme écrit en Python (Version 3.7). Les programmes linéaires ont été résolus à l'aide du solveur Gurobi (licence académique) sous la forme d'une librairie externe de Python.

9. L'usage du conditionnel ici s'impose car les algorithmes des ensembles \mathfrak{A}_1 et \mathfrak{A}_2 ne retournent pas toutes les explications de niveaux 1 et 2 conformes aux schémas déductifs auxquels ils se rapportent.

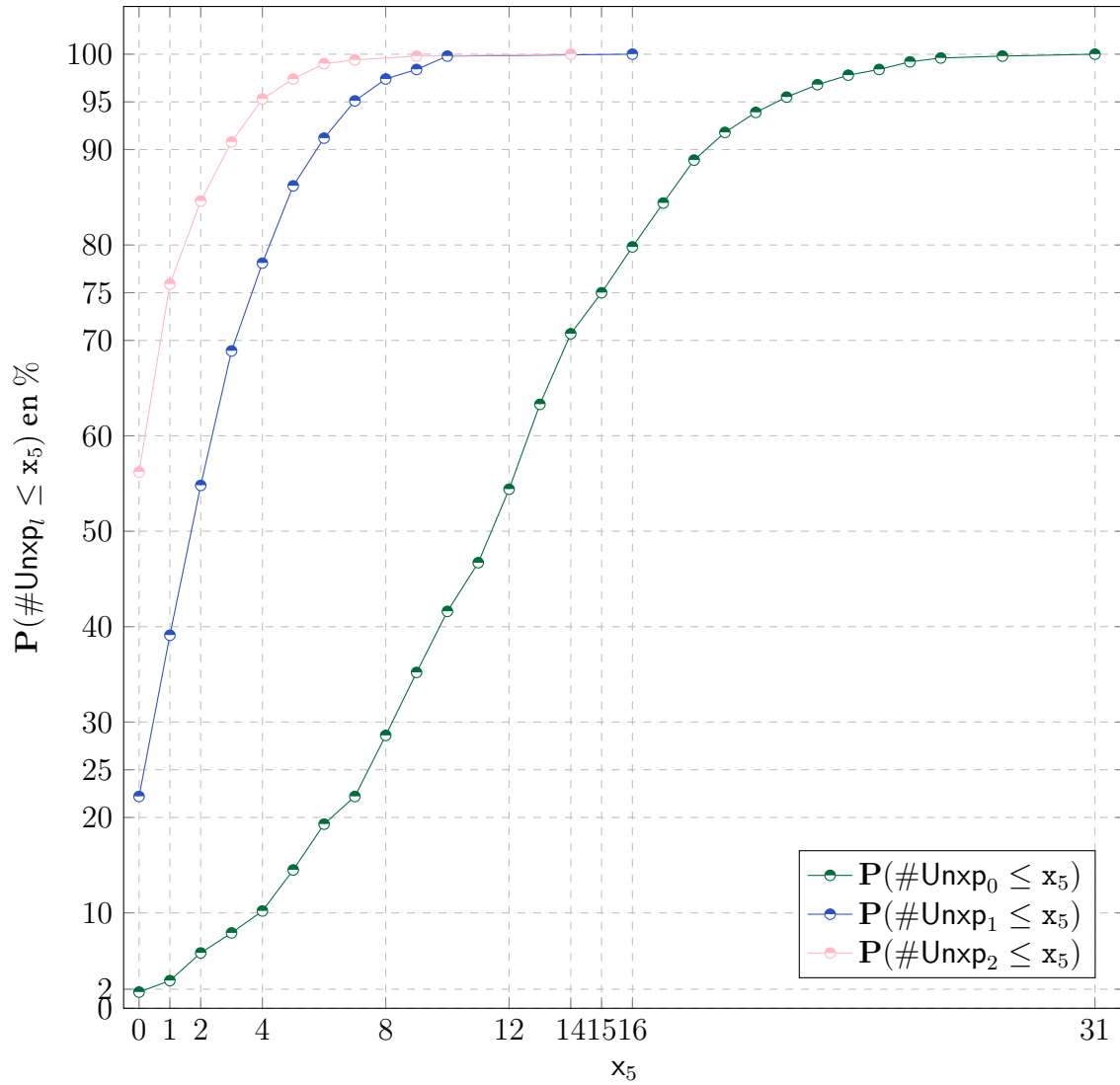


FIGURE 3.2 – Fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_5 inexplicables

3.3. ÉVALUATION EMPIRIQUE DU POUVOIR DÉDUCTIF DE SCHÉMAS D'EXPLICATION 81

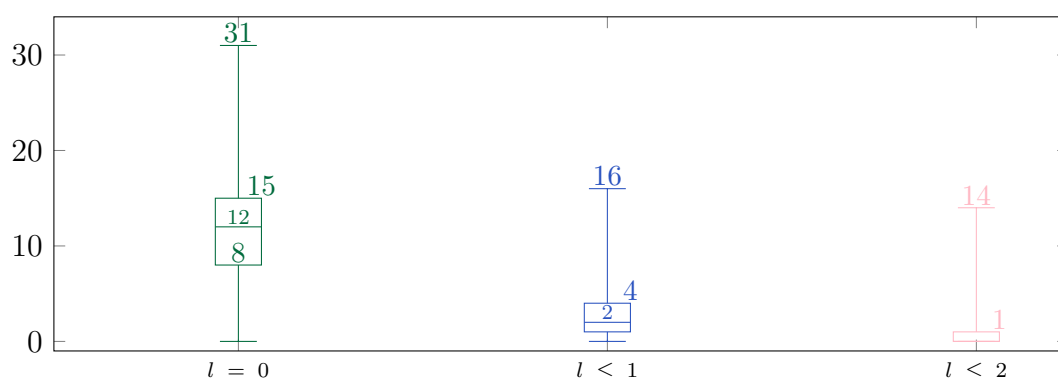


FIGURE 3.3 – Diagrammes en boîte synthétisant les fonctions de répartition de la Figure 3.2

Temps d'exécution : La durée totale du calcul d'explication réalisé à partir des algorithmes de la Table 3.1 sur l'ensemble des 90 comparaisons par paire de chacune des 516 ROLA définies sur $2^{[5]}$ a été inférieure à 2 minutes soit moins d'un tiers de seconde pour chaque ROLA.

Le cas $m = 6$

La Figure 3.4 ($m = 6$), à l'image de la Figure 3.3 ($m = 5$), présente une synthèse des fonctions de répartition (observée) des variables $\#\text{Unxp}_l$ ($l \leq 2$). On observe que la valeur x_6^{\min} minimale telle que $\mathbf{P}(\#\text{Unxp}_0 \leq x_6) > 0$ est nulle. Ceci signifie qu'il existe aussi dans le cas $m = 6$, une proportion non nulle de ROLA dont toutes les comparaisons par paire disjointes sont explicables suivant des schémas de niveau 0. Pour rappel, lorsque $m = 4$ (respectivement $m = 5$), ces proportions étaient d'environ 37,7% (respectivement 1,7%). Ici, cette proportion est d'environ 0,012%¹⁰. On observe (Figure 3.4) par ailleurs, qu'au moins 75% des relations d'ordre linéaire additives laissent inexplicables au plus $x_6^{Q_3} = 6$ (sur 301) comparaisons par paire disjointes lorsqu'on se sert des algorithmes de calcul d'explications conformes à tous les schémas répertoriés dans la Table 3.1 ; les valeurs correspondant au 1^{er} quartile et à la médiane étant respectivement $x_6^{Q_1} = 1$ et $x_6^{Q_2} = 2$.

10. "Existe-t-il un nombre d'items m_{lim} supérieur ou égal à 7 à partir duquel cette proportion devient nulle ? Si oui, laquelle ? Si non, est-on en mesure de caractériser les relations d'ordre linéaire additives dont toutes les comparaisons par paire disjointes s'expliquent, à partir de leurs paires critiques, suivant les schémas de base tr , cg et cov ?". Voici quelques questions de recherche qui se dégagent des observations expérimentales faites ici. Leur apporter des réponses enrichira à coup sûr la connaissance que l'on a (ou croit avoir) des préférences représentables par un modèle "aussi simple" que le modèle additif.

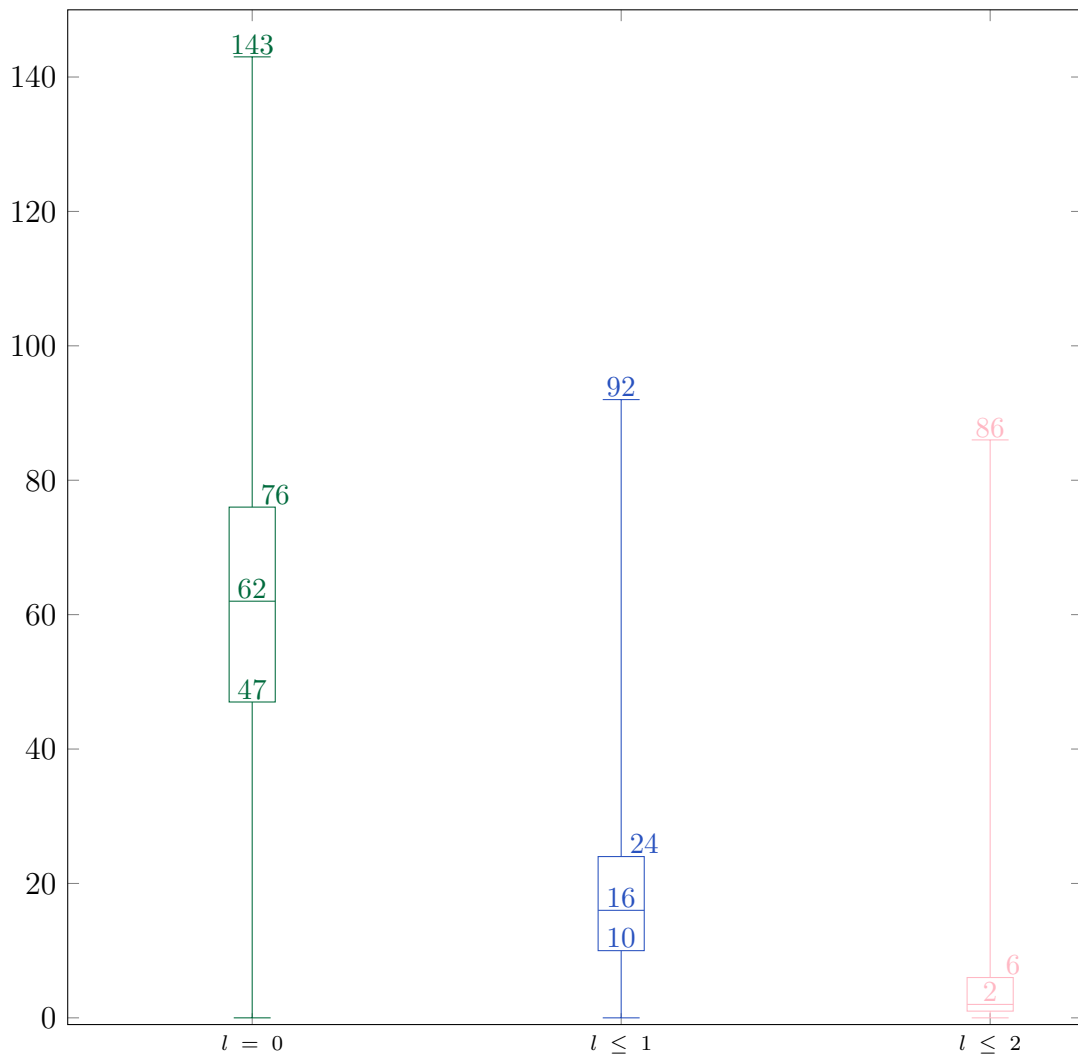


FIGURE 3.4 – Synthèse des fonctions de répartition du nombre de comparaisons par paire de \mathbb{T}_6 inexplicables

Temps d'exécution : La durée totale du calcul des explications à partir des algorithmes de la Table 3.1 sur l'ensemble des 301 comparaisons par paire de chacune des 124 187 ROLA définies sur $2^{[6]}$ et vérifiant (3.15) a été d'environ 54 heures soit moins de 2 secondes pour chaque ROLA.

3.4 Évaluation du pouvoir déductif des propriétés des ROLA

Dans la précédente section, nous avons mis en place et exécuté un protocole expérimental destiné à évaluer le pouvoir déductif des schémas répertoriés dans la Table 3.1. Cette évaluation a été rendue possible grâce aux algorithmes que nous avons implémentés. Pour rappel, tout schéma déductif (de base ou composé) met en œuvre un ou plusieurs mécanismes de déduction reposant sur trois propriétés fondamentales des relations d'ordre linéaire additives. Il s'agit des propriétés de transitivité, d'indépendance et de couverture.

Dans cette section, nous nous proposons d'évaluer le pouvoir déductif de ces trois propriétés à partir de l'ensemble des comparaisons par paire critiques de toute relation d'ordre linéaire additive \succ définie sur $2^{[m]}$, $m = 4, 5, 6$. On sait déjà, au vu des résultats expérimentaux précédents, que l'application de ces trois propriétés permet de déduire à partir de leurs paires critiques, toutes les ROLA définies sur $2^{[4]}$. Mais qu'en est-il des ROLA définies sur $2^{[5]}$ et $2^{[6]}$?

Pour répondre à cette question, deux options s'offrent à nous :

1. Identifier un ensemble fini \mathcal{S}^* de schémas déductifs reposant sur l'application des propriétés de transitivité, d'indépendance et de couverture et pour lequel l'on disposerait d'un ensemble $\mathfrak{A}_{\mathcal{S}^*}$ d'algorithmes calculant des explications conformes à ces schémas et qui, appliqués à toute comparaison par paire de toute ROLA définie sur $2^{[5]}$ et $2^{[6]}$ retourne une explication.
2. Tenter par d'autres moyens d'approcher la performance de l'ensemble $\mathfrak{A}_{\mathcal{S}^*}$.

Nous opterons pour la seconde solution en procédant à ce que nous appellerons *une évolution dynamique du contexte*.

Pour rappel, on appelle *contexte*, le triplet $(\succ, \mathbb{B}_{\succ}, sch)$ composé d'une relation d'ordre linéaire additive \succ , d'une base de référence \mathbb{B}_{\succ} , sous-ensemble de \succ et d'un schéma déductif sch suivant lequel on tentera d'expliquer la déduction de comparaisons par paire de la relation \succ à partir des comparaisons par paire admises (de \mathbb{B}_{\succ} et \mathbb{D}_m). L'idée derrière le concept d'*évolution dynamique du contexte* est simple. Il s'agit, lorsque l'on souhaite expliquer non pas une seule comparaison par paire mais un ensemble de plusieurs comparaisons par paire, de s'autoriser à compléter itérativement la base de référence \mathbb{B}_{\succ} des comparaisons par paire qu'on a réussi à expliquer à l'aide d'algorithmes de calcul d'explication. Ce faire, va donc potentiellement permettre d'accroître la proportion de comparaisons par paire explicables. De plus, on calculerait ainsi indirectement des explications conformes à des schémas autres que ceux considérés dans la Table 3.1.

Dans la sous-section 3.4.1, nous illustrerons et détaillerons la procédure d'évolution dynamique du contexte que nous utiliserons (en sous-section 3.4.2) dans le but de mesurer expérimentalement le pouvoir déductif des propriétés de transitivité, d'indépendance

et de couverture.

3.4.1 Évolution dynamique du contexte

Considérons l'exemple suivant.

Exemple 3.2 Soit la relation d'ordre linéaire additive $\succ_{\omega''}$ définie sur $2^{[5]}$ et représentable par la fonction de score ω'' suivante :

$$\omega'' = (a : 12, b : 15, c : 16, d : 20, e : 22)$$

Soit la base de référence $\mathbb{B}_{\succ_{\omega''}}$ suivante :

$$\mathbb{B}_{\succ_{\omega''}} = \{ \text{abc } \succ_{\omega''} \text{ de} \bullet \text{ be } \succ_{\omega''} \text{ cd} \bullet \text{ c } \succ_{\omega''} \text{ b} \bullet \text{ bd } \succ_{\omega''} \text{ ae} \bullet \text{ ad } \succ_{\omega''} \text{ bc} \}$$

à partir de laquelle on déduit l'ensemble de comparaisons par paire $\mathbb{E}_{\succ_{\omega''}}$ suivant :

$$\mathbb{E}_{\succ_{\omega''}} = \{ \text{ae } \succ_{\omega''} \text{ bc} \bullet \text{ d } \succ_{\omega''} \text{ c} \bullet \text{ ab } \succ_{\omega''} \text{ e} \}$$

L'ensemble des algorithmes \mathfrak{A} (Table 3.1) appliqués à $\mathbb{B}_{\succ_{\omega''}}$ échouent à expliquer les comparaisons par paire $\text{d } \succ_{\omega''} \text{ c}$ et $\text{ab } \succ_{\omega''} \text{ e}$; l'explication de $\text{ae } \succ_{\omega''} \text{ bc}$ se présentant par contre comme suit :

L'alternative ae est préférée à l'alternative bc car la comparaison par paire $\text{ae } \succ_{\omega''} \text{ bc}$ se déduit, par application du mécanisme $2\text{-Tr}^{\succ_{\omega''}}$, des comparaisons par paire $\text{ae } \succ_{\omega''} \text{ ad}$ (non admis) et $\text{ad } \succ_{\omega''} \text{ bc}$ (admise) où :

- $\text{ae } \succ_{\omega''} \text{ ad}$ se déduit, par application du mécanisme $\text{a-Cg}^{\succ_{\omega''}}$, de la comparaison par paire $\text{bce } \succ_{\omega''} \text{ bcd}$ (non admis) où :
 - $\text{bce } \succ_{\omega''} \text{ bcd}$ se déduit, par application du mécanisme $2\text{-Cov}^{\succ_{\omega''}}$, des comparaisons par paire $\text{be } \succ_{\omega''} \text{ cd}$ (admise) et $\text{c } \succ_{\omega''} \text{ b}$ (admise).

Cette explication est conforme au schéma tr/cg/cov de niveau 2.

En désignant par $\mathbb{B}_{\succ_{\omega''}}^1$, l'ensemble obtenu en rajoutant $\text{ae } \succ_{\omega''} \text{ bc}$ à $\mathbb{B}_{\succ_{\omega''}}$:

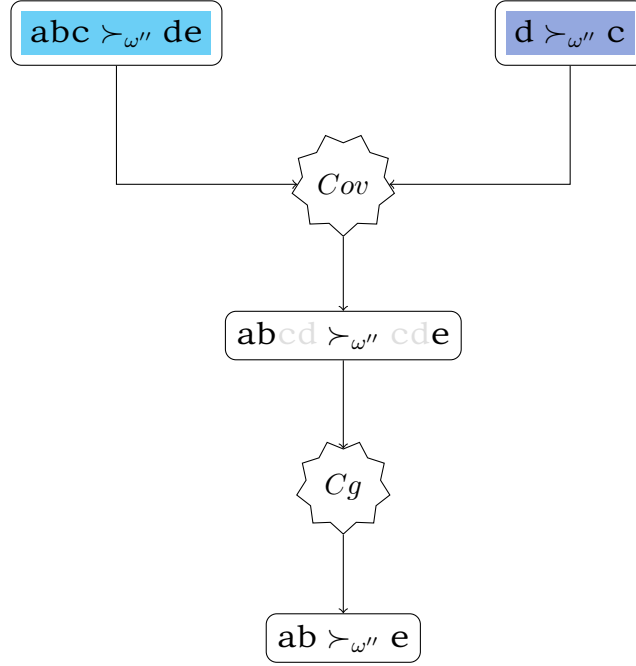
$$\mathbb{B}_{\succ_{\omega''}}^1 = \mathbb{B}_{\succ_{\omega''}} \cup \{ \text{ae } \succ_{\omega''} \text{ bc} \}$$

l'exécution de l'appel $\text{Algo}_{\text{cg/tr}}((\text{d}, \text{c}), \succ_{\omega''}, \mathbb{B}_{\succ_{\omega''}})$ produit l'explication suivante :

$$\underline{\text{bd } \succ_{\omega''} \text{ bc}} \implies \text{d } \succ_{\omega''} \text{ c} \quad (\emptyset\text{-Cg}^{\succ_{\omega''}})$$

avec :

$$\text{bd } \succ_{\omega''} \text{ ae} \wedge \text{ae } \succ_{\omega''} \text{ bc} \implies \underline{\text{bd } \succ_{\omega''} \text{ bc}} \quad (2\text{-Tr}^{\succ_{\omega''}})$$

FIGURE 3.5 – Explication de $ab \succ_{\omega''} e$ sous forme de diagramme.

Cette explication est conforme au schéma cg/tr de niveau 1.

Enfin, en désignant par $\mathbb{B}_{\succ_{\omega''}}^2$, l'ensemble obtenu en rajoutant $d \succ_{\omega''} c$ à $\mathbb{B}_{\succ_{\omega''}}^1$:

$$\mathbb{B}_{\succ_{\omega''}}^2 = \mathbb{B}_{\succ_{\omega''}}^1 \cup \{d \succ_{\omega''} c\}$$

l'exécution de l'appel $\text{Algo}_{cg/cov}((ab, e), \succ_{\omega''}, \mathbb{B}_{\succ_{\omega''}})$ produit l'explication représentée sous forme de diagramme à la Figure 3.5.

Cette explication est conforme au schéma cg/cov de niveau 1.

Cet exemple illustre bien le concept d'évolution dynamique du contexte en montrant comment l'explication de $ae \succ_{\omega''} bc$ a rendu possible celle de $d \succ_{\omega''} c$ qui a ensuite rendu possible celle de $ab \succ_{\omega''} e$. On remarquera que : $ab \succ_{\omega''} e$ est explicable à partir des comparaisons par paire $\mathbb{B}_{\succ_{\omega''}}$, suivant le schéma $cg/cov/cg/tr/cg/cov$ de niveau 5. La représentation en diagramme de la Figure 3.4.1 détaille ladite explication.

De façon générale, si l'on souhaite expliquer un ensemble \mathbb{E} de comparaisons par paire suivant un ensemble \mathcal{A} d'algorithmes de calcul d'explication, on peut procéder itérativement comme l'indique l'Algorithme 2.

Dans cet algorithme, \mathbb{B}_{\succ}^0 et \mathbb{E}^0 sont respectivement la base de référence \mathbb{B}_{\succ} et l'ensemble \mathbb{E} des comparaisons par paire à expliquer.

La procédure $\text{Explain}(x, y, \mathbb{B}_{\succ}^t, \mathcal{A})$ consiste à exécuter séquentiellement les algorithmes d'explication de l'ensemble \mathcal{A} avec comme arguments la comparaison par paire

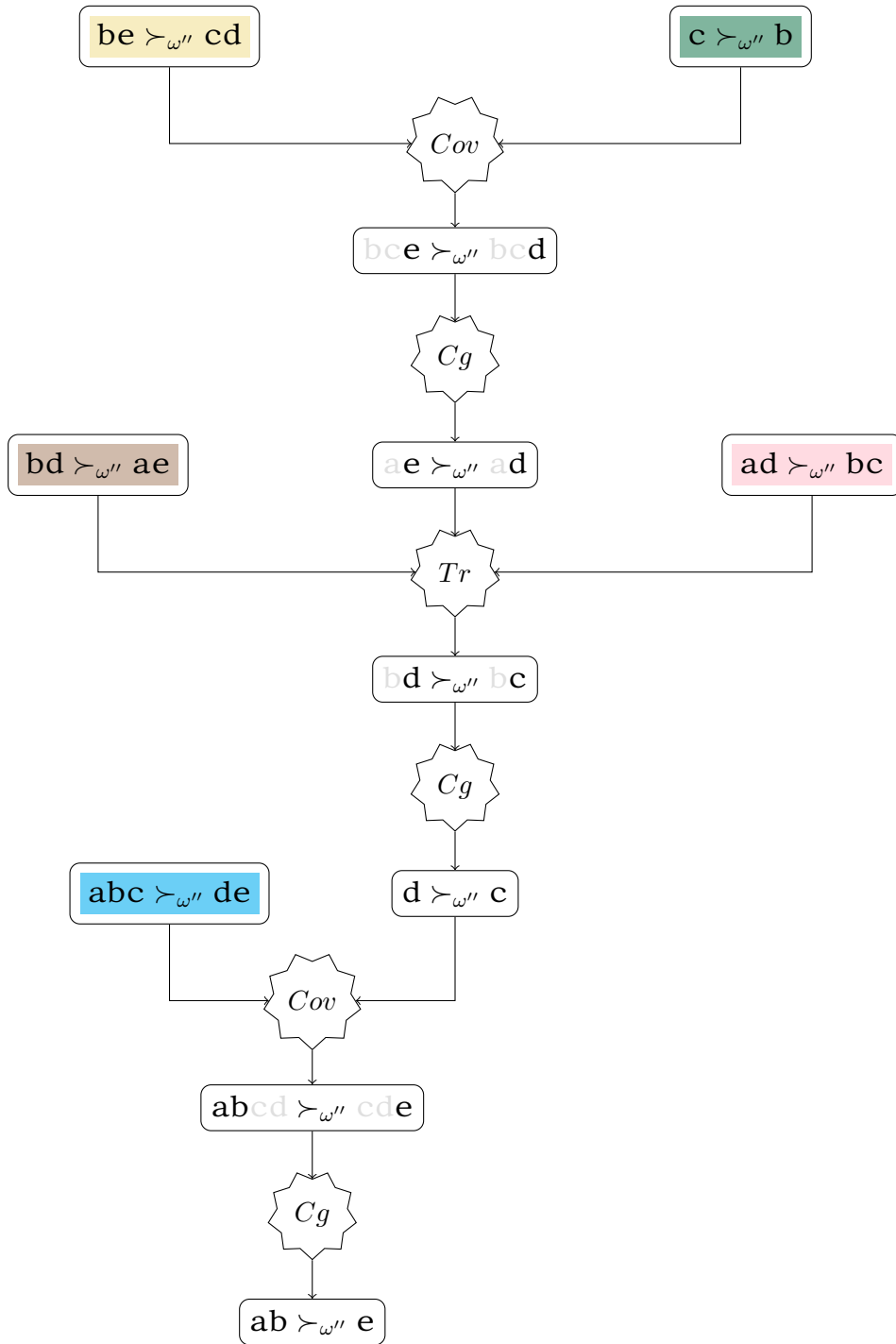


FIGURE 3.6 – Évolution dynamique du contexte : explication de $ab \succ_{\omega''} e$ sous forme de diagramme à partir de $\mathbb{B}_{\succ_{\omega''}}$.

Algorithme 2 Explication d'un ensemble \mathbb{E} de comparaisons par paire en appliquant une évolution dynamique du contexte.

Entrée(s) $\mathbb{E}, \mathbb{B}_{\succ}, \mathfrak{A}$ un ensemble d'algorithmes de calcul d'explication

Sortie(s) Les détails d'explication des comparaisons par paire de \mathbb{E} à travers la table TX.

```

1: Initialiser TX
2: convergence  $\leftarrow$  False
3:  $t \leftarrow 0$ 
4: tant que  $|\mathbb{E}^t| \neq 0$  and not convergence faire
5:    $\mathbb{B}_{\succ}^{t+1} \leftarrow \mathbb{B}_{\succ}^t$ 
6:    $\mathbb{E}^{t+1} \leftarrow \emptyset$ 
7:   convergence  $\leftarrow$  True
8:   pour chaque comparaison par paire  $(x, y) \in \mathbb{E}^t$  faire
9:     Exécuter la procédure  $\text{Explain}(x, y, \mathbb{B}_{\succ}^t, \mathfrak{A})$ 
10:    si  $\text{Explain}(x, y, \mathbb{B}_{\succ}^t, \mathfrak{A})$  échoue alors
11:      Rajouter  $(x, y)$  à  $\mathbb{E}^{t+1}$ 
12:    sinon
13:      Rajouter à TX l'explication correspondante de  $(x, y)$ 
14:      Rajouter  $(x, y)$  à  $\mathbb{B}_{\succ}^{t+1}$ 
15:      convergence  $\leftarrow$  False
16:    fin du si
17:  fin du pour
18:   $t \leftarrow t + 1$ 
19: fin du tant que
20: retourner TX

```

(x, y) et la base de référence \mathbb{B}_{\succ}^t . Elle échoue lorsqu'aucun de ces algorithmes n'arrivent à produire une explication de (x, y) .

3.4.2 Application

En exécutant l'Algorithme 2 avec comme entrées $\mathbb{T}_{\succ}, \mathbb{C}\mathbb{R}_{\succ}$ et \mathfrak{A} (Table 3.1) où $\succ \in \mathcal{R}_m$ ($m = 5, 6$), on parvient à expliquer toutes les comparaisons par paire de \mathbb{T}_{\succ} :

Observation 3.1 *Toute comparaison par paire de toute relation d'ordre linéaire additive \succ définie sur $2^{[m]}$ ($m = 4, 5, 6$) est explicable dans au moins un contexte $(\succ, \mathbb{C}\mathbb{R}_{\succ}, sch)$ où sch est un schéma déductif basé sur aucune autre propriété que la transitivité, l'indépendance et la couverture.*

Autrement dit, **lorsqu'elles sont appliquées aux comparaisons par paire critiques des ROLA de \mathcal{R}_m ($m = 4, 5, 6$), les propriétés de transitivité, d'indépen-**

dance et de couverture ont un pouvoir déductif que l'on pourrait qualifier de total.

Temps d'exécution : La durée totale d'exécution observée pour l'ensemble des 516 ROLA de \mathcal{R}_5 vérifiant (3.15) est de moins de 3 minutes et l'Algorithme 2 converge au maximum au bout de 2 itérations.

En ce qui concerne les 124 187 ROLA de \mathcal{R}_6 vérifiant (3.15), on a observé une convergence de l'Algorithme 2 au maximum au bout de 3 itérations. Le temps total qu'il a fallu pour ce calcul est d'environ 70 heures; ce qui correspond en moyenne à environ 2 secondes pour calculer l'explication des 301 comparaisons par paire disjointes de chaque ROLA.

3.5 Conclusion et perspectives

L'essentiel des contributions de ce chapitre est expérimental. Il a consisté à décrire des familles d'algorithmes destinés à calculer l'explication d'une comparaison par paire dans un *contexte* donné. Le *contexte* décrit la relation d'ordre linéaire additive à laquelle appartient la comparaison par paire à expliquer, son sous-ensemble de comparaisons par paire considérées comme admises et le schéma explicitant les mécanismes en jeu dans la déduction de celle-ci. Quinze algorithmes organisés en trois familles ont été implémentés et ont d'une part, permis de mesurer le pouvoir explicatif des schémas déductifs (de niveau inférieur ou égal à 2) auxquels ils se rapportent et d'autre part, de prouver que la combinaison des axiomes de transitivité, congruence et couverture, suffit à expliquer toute comparaison par paire de toute relation d'ordre linéaire additive définie sur $m = 4, 5, 6$ items lorsque les paires critiques de cette dernière sont incluses dans l'ensemble des comparaisons par paire admises.

Nous estimons que la nouvelle approche de l'explication que nous proposons, bien qu'illustrée ici à partir des relations d'ordre linéaire additives, peut être étendue à toute relation de préférences pour laquelle on dispose d'un système d'axiomes dont la transformation en mécanismes de déduction à des fins d'explication peut être envisagée. C'est ce que nous essayerons de montrer à travers les prochains chapitres où les préférences du décideur, représentables par un modèle additif, sont exprimées sur un ensemble fini d'alternatives.

Deuxième partie

Explication de préférences robustes et de recommandations

Chapitre 4

Fonctions de l'explication de préférences robustes

Résumé

On considère un modèle additif de décision multicritère dans lequel les critères sont évalués sur une échelle binaire. Le décideur fournit une information préférentielle sur la base de laquelle une relation de préférences nécessaire est établie. Le type de recommandation considéré étant un choix, les conclusions robustes pouvant découler de cette relation sont soit des choix nécessaires (certitude qu'une certaine alternative appartient au sous-ensemble des meilleures alternatives à recommander) soit des choix impossibles (certitude qu'une certaine alternative ne pourrait être recommandée). Dans ce chapitre, nous nous intéressons à l'explication de telles conclusions robustes tant dans une optique de justification que d'élicitation. L'approche d'explication adoptée est la même que celle décrite dans les chapitres précédents et qui se base sur la définition d'un *contexte*. Dans la première partie du chapitre, nous montrons comment des travaux antérieurs d'explication de déductions de la relation nécessaire s'inscrivent dans ce cadre. Nous indiquons ensuite comment ces travaux peuvent être étendus en complétant l'ensemble des comparaisons par paire admises des déclarations de préférences du décideur et en diversifiant les schémas déductifs utilisés. Des expérimentations numériques sont conduites dans le but de quantifier l'intérêt de ces changements lorsqu'il s'agit de justifier les choix nécessaires et/ou impossibles déduits de l'information préférentielle fournie par le décideur. Dans la deuxième partie du chapitre, nous montrons que l'explication de conclusions robustes peut également remplir une fonction d'élicitation des préférences du décideur. À cette fin, nous proposons une ébauche de protocole interactif à travers lequel l'explication sert d'occasion de collecte d'information préférentielle additionnelle susceptible de faire aboutir le processus d'aide à la détermination précise de la recommandation.

Sommaire

| | | |
|------------|--|------------|
| 4.1 | Introduction | 93 |
| 4.2 | Brève revue de littérature et définitions | 95 |
| 4.2.1 | Relation nécessaire | 95 |
| 4.2.2 | Conclusions robustes dans le cadre du choix multiple | 97 |
| 4.2.3 | Explications <i>questionnable</i> et à l'aide de <i>preference swaps</i> | 99 |
| 4.3 | Fonction de justification de l'explication | 102 |
| 4.3.1 | Exemple introductif | 102 |
| 4.3.2 | Justification des conclusions robustes de l'exemple introductif | 104 |
| 4.3.3 | Expérimentations numériques | 110 |
| 4.4 | Fonction d'élicitation de l'explication | 117 |
| 4.4.1 | Exemple introductif | 118 |
| 4.4.2 | Illustration de la fonction d'élicitation sur l'exemple introductif | 120 |
| 4.4.3 | Ébauche d'un protocole interactif | 125 |
| 4.4.4 | Formalisation et modélisation | 127 |
| 4.5 | Conclusion et perspectives | 134 |

4.1 Introduction

Le concept de “robustesse” en aide à la décision a fait l’objet, ces deux dernières décennies, de multiples travaux scientifiques ; la quête (ou préoccupation) de robustesse se justifiant par la présence « d’à-peu-près, de zones d’ignorances susceptibles d’affecter le travail de modélisation de l’analyste » [Roy, 2009]. En décision multicritère plus spécifiquement, cette préoccupation de robustesse a conduit à la proposition d’un nombre considérable de méthodes destinées à la formulation de recommandations (choix, classement et tri) à partir d’affirmations de préférences du décideur représentables par différents types de modèles (additif, intégrale de Choquet, ...) ¹. Dans le cas du modèle additif, on peut citer les méthodes UTA^{GMS} [Greco et al., 2008] et $UTADIS^{GMS}$ [Greco et al., 2010a] qui relèvent toutes deux du paradigme de désagrégation des préférences (“*disaggregation–aggregation approach*” en anglais) selon lequel l’éllicitation s’effectue de façon indirecte : plutôt que de fournir directement les valeurs des paramètres du modèle de préférences à utiliser, le décideur est invité à exprimer ses préférences sous la forme d’exemples de jugements holistiques ; de ces exemples, sont ensuite déduites des valeurs compatibles desdits paramètres qui permettront la formulation de « conclusions robustes » [Roy, 2009]. Une conclusion est une information déduite du modèle et fournie au décideur durant le processus d’aide à la décision ; il peut s’agir de la solution du problème (c’est-à-dire la recommandation) mais aussi d’une propriété, un fait qui peut lui être utile [Bouyssou et al., 2006b]. Une conclusion est robuste si elle est vérifiée par l’ensemble de toutes les valeurs de paramètres compatibles avec les préférences exprimées par le décideur. Lorsque le problème à traiter est un problème de type choix, une conclusion robuste peut prendre la forme de l’assertion de la présence “nécessaire” de l’alternative x parmi les k meilleures alternatives ou simplement de la préférence “nécessaire” de l’alternative x par rapport à une autre alternative y [Greco et al., 2008]. Nous nous intéresserons dans ce chapitre à l’explication de telles conclusions.

Le cadre de décision que nous considérerons met en présence un décideur et un analyste qui interagissent dans le but de parvenir au choix d’un sous-ensemble non vide A^* d’alternatives parmi un ensemble (plus grand) A ($A^* \subset A$) d’alternatives décrites sur m critères ayant exactement deux niveaux d’évaluation (“fort” et “faible”). Les préférences du décideur sont collectées sous la forme d’un ensemble noté \mathbb{P} (Preference Information) de comparaisons par paire d’alternatives (x, y) qui traduisent sa préférence de l’alternative x sur l’alternative y . Ces alternatives, sur lesquelles le décideur sait exprimer des préférences, forment l’ensemble A^R des alternatives dites de *référence* ; les ensembles A^R et A étant le plus souvent disjoints. Nous faisons l’hypothèse que les préférences du décideur sont représentables par un modèle additif. Sous cette hypothèse, l’ensemble des comparaisons par paire déduites de l’ensemble \mathbb{P} forment ce que [Greco et al., 2008] appellent la *relation nécessaire* et que nous noterons $\mathbb{N}_{\mathbb{P}}$. Les conclusions robustes qui

1. Le lecteur intéressé pourra consulter la référence suivante : [Greco et al., 2010b].

feront l'objet d'explication seront donc fondées sur les déductions de la relation nécessaire.

Concrètement, l'objectif du chapitre est d'appliquer notre approche de l'explication, axée autour de la définition d'un *contexte* (Chapitres 2 et 3), au cas des déductions de la relation nécessaire. Nous montrerons qu'il s'agit d'un cadre dans lequel peuvent s'inscrire un certain nombre de travaux antérieurs dont notamment ceux de [Belahcène et al., 2017] et [Amoussou et al., 2023a]. Ces derniers, traitant (tout au moins partiellement) du même cadre de décision que celui décrit au paragraphe précédent, ont contribué à illustrer les fonctions de **justification** et d'**élicitation** que peut remplir l'explication en aide multicritère à la décision. Ce chapitre a vocation à étendre ces travaux en :

- accroissant la capacité à produire des explications de conclusions robustes via l'élargissement de l'ensemble des comparaisons par paire considérées comme admises aux comparaisons par paire de l'ensemble $\mathbb{P}\mathbb{I}$ et en considérant d'autres schémas déductifs que le seul schéma de couverture, pour ce qui est de la fonction de **justification**,
- en proposant une ébauche de protocole interactif dans lequel l'explication d'une déduction de la relation nécessaire permet de capturer de l'information préférentielle supplémentaire susceptible de faire aboutir le processus d'aide à la détermination précise de la recommandation (fonction d'**élicitation**)²

Le chapitre sera organisé de la façon suivante : composé de quatre sections dont deux consacrés à l'illustration des fonctions de **Justification** et d'**Élicitation** de l'explication, il débutera par une brève revue de littérature (Section 4.2) et s'achèvera par une conclusion et l'évocation de perspectives futures (Section 4.5). La Section 4.2 rappellera la notion de relation nécessaire [Greco et al., 2008] puis, précisant le cadre de décision considéré, détaillera les conclusions robustes qui seront sujettes à explication. Elle s'achèvera par un positionnement des contributions de ce chapitre par rapport aux travaux de [Belahcène et al., 2017] et [Amoussou et al., 2023a]. Pour rappel, lesdites contributions sont l'illustration des fonctions de justification et d'élicitation de l'explication de préférences robustes en aide multicritère à la décision. Elles feront respectivement l'objet des Sections 4.3 et 4.4. Chacune de ces deux sections sera introduite par un exemple décrivant une situation d'aide à la décision au cours de laquelle sont produites des explications. Dans la Section 4.3, ces explications auront vocation à justifier des conclusions robustes auxquelles est parvenu l'analyste et dont il souhaite faire part au décideur. Elles intégreront dans leur formulation, les déclarations de préférences du décideur. Les algorithmes permettant leur calcul seront brièvement décrits et l'adresse du dépôt où sont accessibles leurs implémentations sera fournie. Cette section sera clôturée par des expérimentations numériques destinées à mesurer le potentiel explicatif des algorithmes

2. Cette proposition a fait l'objet d'un travail ([Amoussou et al., 2023b]) soumis et présenté à l'occasion des 17^{èmes} Journées d'Intelligence Artificielle Fondamentale (JIAF 2023) .

implémentés. Le cœur de la Section 4.4 est la proposition d'une ébauche de protocole interactif à travers lequel l'explication pourra pleinement remplir sa fonction d'élicitation. Ce protocole sera illustré à travers le déroulé d'un second exemple de situation d'aide à la décision qui, grâce à la production d'explications soumises au décideur, aboutira à la détermination exacte de la recommandation à lui faire. La dernière partie de cette section sera consacrée à des questions de formalisation et modélisation du problème de décision associé au calcul de telles explications.

Du point de vue de la lecture de ce (gros) chapitre, les Sections 4.3 et 4.4 peuvent être considérées comme des sous-chapitres indépendants pour lesquels la Section 4.2 fait office de (grande) introduction.

4.2 Brève revue de littérature et définitions

Dans tout le chapitre, nous faisons l'hypothèse que les préférences du décideur sont représentables par un modèle additif. Le décideur exprime ses préférences sur un ensemble \mathbf{A}^R d'alternatives dites de référence. Ces alternatives, tout comme les alternatives de l'ensemble \mathbf{A} (sur lequel doit porter la recommandation) sont évaluées suivant m critères binaires. Dans la suite, l'ensemble des critères sera noté $[m]$. Les deux niveaux d'évaluation des critères seront symbolisés de la façon suivante : le niveau "fort" représenté par 1 et le niveau "faible" par 0. On admettra que (pour tout décideur), le niveau "fort" est préféré au niveau "faible" : $1 \succ_i 0 \forall i \in [m]$. Une alternative pourra donc être représentée, soit, par un vecteur de m composantes – valuées à 1 ou 0 – soit sous une forme "ensembliste" : l'ensemble des critères sur lesquels elle vaut 1. On notera \mathbf{X} le produit cartésien des m échelles binaires ordonnées $X_i = \{0, 1\}$:

$$\mathbf{X} = \prod_{i \in [m]} X_i$$

On a donc : $|\mathbf{X}| = 2^m$. Aussi, généralement, a-t-on : $|\mathbf{A}| \ll |\mathbf{X}|$ et $|\mathbf{A}^R| \ll |\mathbf{X}|$.

4.2.1 Relation nécessaire

Comme précisé dans l'introduction, les préférences du décideur sont collectées sous la forme d'un ensemble noté \mathbb{P} (Preference Information) de comparaisons par paire d'alternatives (x, y) indiquant sa préférence de l'alternative x sur l'alternative y . La Définition 4.1 formalise l'hypothèse de représentabilité de telles préférences par un modèle additif :

Définition 4.1 (Modèle additif avec critères binaires) *Une relation de préférences \succsim sur \mathbf{X} est représentable par un modèle additif si et seulement si il existe une fonction de score $\omega : \langle \omega_i \rangle_{i \in [m]}$ avec $\omega_i : X_i \rightarrow \mathbb{R}^+$ telle que :*

$$x \succsim y \iff \omega(x) = \sum_{i \in [m]} \omega_i(x_i) \geq \sum_{j \in [m]} \omega_j(y_j) = \omega(y)$$

où x_i (resp. y_j) symbolise l'évaluation (0 ou 1) de l'alternative x (resp. y) sur le i -ème (resp. j -ième) critère.

Sans perte de généralité, nous considérerons dans la suite que $\omega_i(0) = 0$ avec comme effet, une simplification de l'écriture de la fonction de score ω et $\omega_i(1) > 0$ pour tout $i \in [m]$.

Nous faisons également l'hypothèse que l'ensemble \mathbb{P} des préférences du décideur est *consistant* c'est-à-dire qu'il se « fond » dans au moins une relation de préférences \succsim sur \mathbf{X} représentable par un modèle additif : $\mathbb{P} \subset \succsim$. L'ensemble des comparaisons composant \mathbb{P} peut donc être restitué par une ou plusieurs fonctions de score ω qui seront dites *compatibles* avec \mathbb{P} .

Définition 4.2 (Fonction de score compatible avec \mathbb{E}) Soit \mathbb{E} un ensemble de comparaisons par paire (x, y) . La fonction de score ω est dite *compatible* avec \mathbb{E} si :

$$\omega(x) \geq \omega(y) \quad \forall (x, y) \in \mathbb{E} \quad (4.1)$$

Lorsqu'il n'existe aucune fonction de score compatible avec \mathbb{E} alors \mathbb{E} est dit *inconsistant*. Dans le cas contraire, \mathbb{E} est dit *consistant*.

L'existence d'une fonction de score compatible avec \mathbb{E} peut être testée à l'aide d'un programme linéaire (sans fonction objectif) dont les contraintes sont définies par les inéquations (4.1).

On désignera, dans la suite, par $\Omega_{\mathbb{E}}$ l'ensemble des fonctions de score compatibles avec l'ensemble \mathbb{E} . Cet ensemble est vide ($|\Omega_{\mathbb{E}}| = 0$) si et seulement si \mathbb{E} est *inconsistant*.

Dans la suite, nous désignerons par \mathbb{A} (resp. \mathbb{X}) l'ensemble des comparaisons par paire de l'ensemble $\mathbf{A} \times \mathbf{A}$ (resp. $\mathbf{X} \times \mathbf{X}$).

L'ensemble \mathbb{P} des comparaisons par paire collectées par l'analyste auprès du décideur lui permet de faire des déductions sur l'ensemble \mathbf{A} des alternatives sur lesquels doit porter la recommandation. Seront, dans cette perspective, d'un intérêt particulier, les déductions de la « relation nécessaire » qui est un concept faisant référence à une certaine idée de robustesse et que l'on doit à [Greco et al., 2008] et qui se définit comme suit :

Définition 4.3 (Relation nécessaire étant donné \mathbb{P}) Étant donné un ensemble consistant de comparaisons par paire \mathbb{P} et deux alternatives $x, y \in \mathbf{X}$ décrites sur m critères binaires, on dit que x est *nécessairement préférée* à y si et seulement si on a : $\omega(x) - \omega(y) \geq 0$ pour toute fonction de score ω compatible avec \mathbb{P} ($\omega \in \Omega_{\mathbb{P}}$).

On notera $\mathbb{N}_{\mathbb{P}}$ l'ensemble des comparaisons par paire $(x, y) \in \mathbb{X}$ telles que x est *nécessairement préférée* à y .

$\mathbb{N}_{\mathbb{P}\parallel}$ est une relation binaire réflexive et transitive [Greco et al., 2008].

L'ensemble \mathbb{D}_m des comparaisons par paire (x, y) tel que : $y_i = 1 \Rightarrow x_i = 1$ traduisant la dominance de Pareto est (trivialement) inclus dans $\mathbb{N}_{\mathbb{P}\parallel}$.

De même, par définition, on a : $\mathbb{P}\parallel \subset \mathbb{N}_{\mathbb{P}\parallel}$.

La Proposition 4.1 suggère une méthode de calcul des éléments de la relation nécessaire basée sur la résolution d'un programme linéaire tandis que la Proposition 4.2 reprenant une partie du Théorème 5 – *Caractérisation de la relation nécessaire à l'aide des covecteurs* (Définition 4.4) – de [Belahcène et al., 2017] offre un moyen algébrique de vérification de la déduction par calcul de la relation nécessaire d'une comparaison par paire (x, y) .

Proposition 4.1 *Étant donné un ensemble consistant de comparaisons par paire $\mathbb{P}\parallel$, la comparaison par paire $(x, y) \in \mathcal{X}$ appartient à $\mathbb{N}_{\mathbb{P}\parallel}$ si l'ensemble $\mathbb{P}\parallel \cup \{(y, x)\}$ est inconsistent (voir Définition 4.2).*

Définition 4.4 (Covecteur de la comparaison par paire (x, y)) *Le covecteur – noté $(x, y)^*$ – de la comparaison par paire (x, y) est le vecteur à m composantes défini comme suit :*

$$(x, y)^* = (\lambda_i)_{i \in [m]}$$

où :

$$\lambda_i = \begin{cases} -1 & \text{si } y_i = 1 \text{ et } x_i = 0 \\ 0 & \text{si } y_i = x_i \\ 1 & \text{si } x_i = 1 \text{ et } y_i = 0 \end{cases}$$

Proposition 4.2 ([Belahcène et al., 2017]) *Étant donné un ensemble consistant de comparaisons par paire $\mathbb{P}\parallel$, la comparaison par paire $(x, y) \in \mathcal{X}$ appartient à $\mathbb{N}_{\mathbb{P}\parallel}$ si $(x, y)^*$ peut s'écrire sous la forme d'une combinaison linéaire à coefficients positifs de covecteurs de comparaison(s) par paire représentant les préférences du décideur – $(u, v) \in \mathbb{P}\parallel$ – et de celles exprimant une dominance de Pareto – $(u, v) \in \mathbb{D}_m$ –.*

Dans la prochaine sous-section, nous indiquerons et définirons les types de conclusions robustes auxquelles permettent d'aboutir les déductions du calcul de la relation nécessaire et qui vont nous intéresser dans ce travail.

4.2.2 Conclusions robustes dans le cadre du choix multiple

Comme \mathbf{A} est l'ensemble des alternatives sur lequel va porter la recommandation de l'analyste, les déductions de l'ensemble $\mathbb{N}_{\mathbb{P}\parallel} \cap \mathbf{A}$ que nous noterons par la suite $\mathbb{N}_{\mathbb{P}\parallel}^{\mathbf{A}}$ (déductions du calcul de la relation nécessaire sur l'ensemble d'alternatives \mathbf{A}) permettront de

construire des conclusions robustes ([Roy, 2009]) en lien avec le type de recommandation à faire. Pour rappel, le type de recommandation qui nous intéressera dans ce chapitre est la recommandation de type choix ; choix du sous-ensemble \mathbf{A}^* composé des k meilleures (suivant les préférences du décideur) alternatives de l'ensemble \mathbf{A} . Nous considérerons tout au long du chapitre que $|\mathbf{A}^*| = k$ et $|\mathbf{A}| = n$ avec $1 \leq k < n$.

En rapport avec la recommandation de type k -best à fournir au décideur, nous considérerons ici, exactement deux catégories de « conclusions robustes » : la conclusion robuste de type **choix nécessaire** et la conclusion robuste de type **choix impossible**. La conclusion robuste de type **choix nécessaire** est l'information qui, compte tenu de l'information préférentielle \mathbb{P} fournie par le décideur, lui indique qu'une alternative $x \in \mathbf{A}$ appartient nécessairement à l'ensemble \mathbf{A}^* des k meilleures alternatives à recommander. À l'inverse, la conclusion robuste de type **choix impossible** est l'information destinée au décideur qui lui garantit qu'une alternative $y \in \mathbf{A}$ ne pourra pas appartenir, compte tenu de l'information préférentielle \mathbb{P} qu'il a fournie, à l'ensemble \mathbf{A}^* des k meilleures alternatives à recommander. Au sujet des deux types de conclusions robustes, nous considérons les propositions suivantes :

Proposition 4.3 *Étant donné un ensemble consistant de comparaisons par paire \mathbb{P} , une alternative $x \in \mathbf{A}$ est un **choix nécessaire** si et seulement si on a :*

$$|\{y \in \mathbf{A} \setminus \{x\}, (x, y) \in \mathbb{N}_{\mathbb{P}}^{\mathbf{A}}\}| \geq n - k$$

c'est-à-dire si x est nécessairement préférée à au moins $n - k$ autres alternatives de l'ensemble \mathbf{A} .

Proposition 4.4 *Étant donné un ensemble consistant de comparaisons par paire \mathbb{P} , une alternative $y \in \mathbf{A}$ est un **choix impossible** si et seulement si on a :*

$$|\{x \in \mathbf{A} \setminus \{y\}, (x, y) \in \mathbb{N}_{\mathbb{P}}^{\mathbf{A}}\}| \geq k$$

c'est-à-dire si il existe au moins k autre(s) alternative(s) de l'ensemble \mathbf{A} nécessairement préférée(s) à y .

Ces (deux) types d'informations robustes sont à la fois pertinentes et utiles pour l'analyste. D'une part, elles permettent de réduire le nombre d'alternatives à (encore) considérer en vue de la formulation de la recommandation finale. D'autre part, l'examen minutieux des alternatives **choix nécessaires** et **choix impossibles** pourrait lui permettre d'identifier les critères et combinaisons de critères pertinents pour le décideur. L'analyste pourrait se servir de cette connaissance pour convenablement orienter ses discussions avec le décideur. Les conclusions robustes pourraient en outre servir de « feed-back » intéressants à partager avec celui-ci. Elles pourraient dès lors, être sujettes à questionnements de la part de ce dernier d'où l'importance, pour l'analyste, de pouvoir les « expliquer » de façon intelligible, à la fois dans une optique de justification (voir

Section 4.3) que dans une logique de facilitation de leur interaction qui sera potentiellement génératrice d'information préférentielle supplémentaire à capturer (voir Section 4.4). Dans la prochaine sous-section, nous détaillerons comment les travaux de [Belahcène et al., 2017] et [Amoussou et al., 2023a] tentent de faire remplit à l'explication de conclusions robustes des fonctions de **justication** et d'**élicitation** et surtout, montrons comment leurs propositions s'inscrivent dans le cadre de la nouvelle approche de l'explication que nous proposons dans ce travail.

4.2.3 Explications *questionnable* et à l'aide de *preference swaps*

La multiplicité des critères et leur conflictualité étant la principale source de difficulté en décision multicritère, [Belahcène et al., 2017] et [Amoussou et al., 2023a], en s'inspirant des travaux de [Hammond et al., 1998], traitent la question de l'explication sous l'angle de la décomposition de la comparaison par paire à expliquer en sous-comparaisons élémentaires impliquant un tout petit nombre de critères.

Soit (u, v) une comparaison par paire de \mathcal{X} . On considèrera les définitions suivantes :

Définition 4.5 (Ensembles **pro, **con** et **neutr**)** *Les ensembles $\mathbf{pro}(u, v)$, $\mathbf{con}(u, v)$ et $\mathbf{neutr}(u, v)$ de la comparaison par paire (u, v) sont définis comme suit :*

$$\mathbf{pro}(u, v) = \{i \in [m], u_i = 1 \text{ et } v_i = 0\}$$

$$\mathbf{con}(u, v) = \{j \in [m], u_j = 0 \text{ et } v_j = 1\}$$

$$\mathbf{neutr}(u, v) = \{k \in [m], u_k = v_k\}$$

*On remarquera que les trois ensembles $\mathbf{pro}(u, v)$, $\mathbf{con}(u, v)$ et $\mathbf{neutr}(u, v)$ partitionnent l'ensemble des critères $[m]$ et représentent respectivement l'ensemble des critères en faveur (**pro**), en défaveur (**con**) et neutre(s) (**neutr**) de la préférence traduite par la comparaison par paire (u, v) .*

*Lorsque $|\mathbf{pro}(u, v)| = |\mathbf{con}(u, v)| = 1$ et $|\mathbf{neutr}(u, v)| = 0$, (u, v) est un *preference swap*. Les explications de [Belahcène et al., 2017] et [Amoussou et al., 2023a] expliquent une comparaison par paire en la décomposant en sous-comparaisons élémentaires de type *preference swap*.*

Soit \mathbb{P} , l'ensemble des préférences du décideur sur l'ensemble des alternatives de référence \mathbf{A}^R .

Soit $\Omega_{\mathbb{P}}$, l'ensemble des fonctions de score compatibles (Définition 4.2) avec \mathbb{P} . On suppose que $|\Omega_{\mathbb{P}}| \neq 0$. Soit ω , un élément de $\Omega_{\mathbb{P}}$ et \succsim_{ω} la relation de préférences sur \mathcal{X} induite par la fonction de score ω . On notera $\mathcal{S}_{\succsim_{\omega}}$, l'ensemble des *preference swaps* de la relation \succsim_{ω} :

$$\mathbb{S}_{\succsim_\omega} = \{(u, v) \in \mathbb{X} : u \succsim_\omega v \text{ et } |\mathbf{pro}(u, v)| = |\mathbf{con}(u, v)| = 1 \text{ et } |\mathbf{neutr}(u, v)| = 0\} \quad (4.2)$$

L'intersection de tous les ensembles $\mathbb{S}_{\succsim_\omega}$, $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ forme l'ensemble des *preference swaps* (u, v) vérifiant $\omega(u) \geq \omega(v)$ pour toute fonction de score ω compatible avec la $\mathbb{P}\mathbb{I}$. Par définition (4.3), cet ensemble est inclus dans $\mathbb{N}_{\mathbb{P}\mathbb{I}}$, l'ensemble des déductions du calcul de la relation nécessaire sur \mathbb{X} . Nous le noterons $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ et qualifierons de “nécessaires” les *preference swaps* de $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$:

$$\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}} = \bigcap_{\omega \in \Omega_{\mathbb{P}\mathbb{I}}} \mathbb{S}_{\succsim_\omega} \quad (4.3)$$

Comme indiqué précédemment, les explications selon [Belahcène et al., 2017] et [Amoussou et al., 2023a] sont composés de *preference swaps*. Dans ces travaux, ces deux types d'explication s'appliquent aux éléments de la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}$. Ils se distinguent de par l'“origine” ($\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ ou $\mathbb{S}_{\succsim_\omega}$ avec $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$) des *preference swaps* utilisés comme le précisent les rappels de définition suivants :

Définition 4.6 (Explication à l'aide de *preference swaps* [Belahcène et al., 2017]) La comparaison par paire déduite $(x, y) \in \mathbb{N}_{\mathbb{P}\mathbb{I}}$ est explicable à l'aide de *preference swaps* nécessaires si et seulement si il existe une application injective ϕ de l'ensemble $\mathbf{con}(x, y)$ vers l'ensemble $\mathbf{pro}(x, y)$ telle que pour tout $j \in \mathbf{con}(x, y)$, la comparaison par paire $(\phi(j), j)$ appartient à $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ ([Belahcène et al., 2017, Théorème 6]).

Définition 4.7 (Explication questionnable [Amoussou et al., 2023a]) Il existe une explication questionnable de (x, y) si et seulement si il existe une fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ et une application injective ϕ de l'ensemble $\mathbf{con}(x, y)$ vers l'ensemble $\mathbf{pro}(x, y)$ telle que pour tout $j \in \mathbf{con}(x, y)$, la comparaison par paire $(\phi(j), j) \in \mathbb{S}_{\succsim_\omega}$ où \succsim_ω est la relation de préférences induite par la fonction de score ω ([Amoussou et al., 2023a, Théorème 3.1]). Une telle explication est une séquence transitive de comparaisons par paire $[(u_1, v_1), \dots, (u_l, v_l)]$ appartenant à \succsim_ω et telle que :

- $u_1 \equiv x, v_l \equiv y$ et $u_k \equiv v_{k-1}$ pour tout $k \in \llbracket 2; l \rrbracket$
- pour tout $j \in \mathbf{con}(x, y)$, il existe un unique $k \in \llbracket 1; l \rrbracket$ tel que : $\mathbf{pro}(\phi(j), j) = \mathbf{pro}(u_k, v_k)$ et $\mathbf{con}(\phi(j), j) = \mathbf{con}(u_k, v_k)$ c'est-à-dire $(\phi(j), j)$ et (u_k, v_k) sont congruentes (Définition 3.2, Chapitre 3); les $(l - |\mathbf{con}(x, y)|)$ autres comparaisons par paire $(u_{k'}, v_{k'})$ appartenant à \mathbb{D}_m .

De ces deux définitions, se dégagent les remarques et les propositions ci-après :

- Les explications à l'aide de *preference swaps nécessaires* ([Belahcène et al., 2017]) sont un cas particulier d'explications questionnables [Amoussou et al., 2023a].

- Tandis que l’explication selon [Belahcène et al., 2017] est une preuve de l’appartenance à la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}$ de la comparaison par paire expliquée (x, y) , l’explication questionnable (plus générale) [Amoussou et al., 2023a] est la preuve de l’appartenance de (x, y) à une relation \succsim_ω induite par une fonction de score compatible ω compatible avec $\mathbb{P}\mathbb{I}$; chacune de ces preuves étant construite essentiellement à partir de *preference swaps*.

Proposition 4.5 *Si $(x, y) \in \mathbb{N}_{\mathbb{P}\mathbb{I}}$ est explicable à l’aide de *preference swaps* nécessaires alors (x, y) est explicable dans le contexte $(\mathbb{N}_{\mathbb{P}\mathbb{I}}, \mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}, cov)$ lorsque $|\mathbf{pro}(x, y)| \geq 2$ et explicable dans le contexte $(\mathbb{N}_{\mathbb{P}\mathbb{I}}, \mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}, cg)$ lorsque $|\mathbf{pro}(x, y)| = 1 : (x, y) \in \mathcal{E}(\mathbb{N}_{\mathbb{P}\mathbb{I}}, \mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}, sch)$ avec $sch \in \{cg, cov\}$.*

Démonstration 4.1 *En effet, lorsque $|\mathbf{pro}(x, y)| \geq 2$, l’application injective ϕ de la Définition 4.6 traduit l’existence d’une instance du mécanisme de couverture (Définition 2.3, Chapitre 2) composée des comparaisons par paire $(\phi(j), j)$, $j \in \mathbf{con}(x, y)$ et éventuellement – lorsque $|\mathbf{pro}(x, y)| > \mathbf{con}(x, y)$ – de la comparaison par paire $(x', \emptyset) \in \mathbb{D}_m$ où x' est l’alternative dont toutes les évaluations x'_i valent 0 à l’exception de celles portant sur les critères $i \in \mathbf{pro}(x, y)$ sans antécédent par ϕ . Cette instance a donc pour unique conclusion, la comparaison par paire (\tilde{x}, \tilde{y}) où $\tilde{x} \equiv \mathbf{pro}(x, y)$ et $\tilde{y} \equiv \mathbf{con}(x, y)$. Comme la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}$ est réflexive, en complétant l’instance du mécanisme de couverture décrit par ϕ des éventuels *preference swaps* nécessaires (\mathbf{i}, \mathbf{i}) tels que $x_i = y_i = 1$, on obtient une nouvelle instance dont l’unique conclusion est (x, y) .*

*Lorsque $|\mathbf{pro}(x, y)| = 1$, (x, y) et l’unique *preference swap* $(\phi(j), j)$ sont congruentes par définition, d’où le résultat.*

Proposition 4.6 *Si (x, y) admet une explication questionnable alors il existe une fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ telle que (x, y) est explicable dans le contexte $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, cov)$ ou $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, tr/cg)$ lorsque $|\mathbf{pro}(x, y)| \geq 2$ et explicable dans le contexte $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, cg)$ lorsque $|\mathbf{pro}(x, y)| = 1$ où \succsim_ω est la relation de préférence induite par la fonction de score ω .*

Démonstration 4.2 *À l’image de la démonstration 4.1, la preuve de la Proposition 4.6 découle de la Définition 4.7 et de celle du schéma déductif composé *tr/cg* (Définition 2.7, Chapitre 2).*

Les Propositions 4.5 et 4.6 montrent comment les différents types d’explication questionnable ([Belahcène et al., 2017, Amoussou et al., 2023a]) s’intègrent au cadre d’explication que nous proposons dans ce travail. Pour rappel, notre approche de l’explication est axée autour de la définition d’un *contexte* qui comprend la relation à laquelle appartient la comparaison par paire à expliquer, un sous-ensemble de cette relation composée des comparaisons par paire admises et d’un schéma déductif qui explicite les mécanismes en jeu dans la déduction. Comme indiqué dans l’introduction, **l’objectif de ce chapitre est d’étendre les travaux précédemment mentionnés en proposant de nouveaux**

types d'explication qui considèrent, en plus des *preference swaps* nécessaires, les éléments de la $\mathbb{P}\mathbb{I}$ comme comparaisons par paire admises et qui sont construits suivant des schémas déductifs autres que les schémas *cg*, *cov* et *tr/cg*. Ce faisant, nous contribuerions à accroître les capacités de production d'explication de déduction de la relation nécessaire tant à des fins de **justification** que d'**élicitation**. En effet, dans une optique de justification de conclusions robustes à fournir au décideur, une explication de la déduction (x, y) dans un *contexte* $(\mathbb{N}_{\mathbb{P}\mathbb{I}}, \mathbb{P}\mathbb{I} \cup \mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}, sch)$ où *sch* est un schéma déductif donné, s'apparente à une réponse à la question – du décideur – « *Pourquoi devrais-je préférer l'alternative x à l'alternative y ?* ». Elle fournit une preuve de la préférence nécessaire de l'alternative x sur l'alternative y à partir des déclarations de préférences du décideur. La remise en cause d'une telle explication par ce dernier conduira à une situation d'inconsistance qu'il faudra résoudre ([Mousseau et al., 2003]) ou à l'abandon de l'hypothèse de représentabilité de ses préférences par un modèle additif. Quant à l'explication de la déduction (x, y) de la relation nécessaire dans un *contexte* $(\succsim_{\omega}, \mathbb{P}\mathbb{I} \cup \mathbb{S}_{\succsim_{\omega}}, sch)$ où \succsim_{ω} est la relation de préférences induite par la fonction de score ω compatible avec la $\mathbb{P}\mathbb{I}$, elle répond à la question « *Pourquoi pourrais-je préférer l'alternative x à l'alternative y ?* ». Elle pourrait donc permettre de capturer de l'information préférentielle supplémentaire sous la forme de *preference swap* validé ou contesté par le décideur attribuant de ce fait à l'explication une fonction d'élicitation.

Les deux prochaines sections sont consacrées à l'illustration de ces deux fonctions et à la proposition d'algorithmes de calcul d'explication. Nous commencerons par la fonction de **justification** (Section 4.3) et terminerons par celle d'**élicitation** (Section 4.4).

4.3 Fonction de justification de l'explication

Dans cette section consacrée à la fonction de justification de l'explication de préférences robustes, nous présenterons en sous-section 4.3.1 un exemple de situation d'aide à la décision dans laquelle sont justifiés des *choix nécessaires* et *choix impossibles* résultant de la prise en compte des préférences du décideur sur l'ensemble A^R des alternatives de référence.

Dans la sous-section 4.3.2, nous fournirons les détails des explications correspondantes obtenus à partir d'algorithmes que nous avons implémentés et achèverons la section par des expérimentations numériques.

4.3.1 Exemple introductif

Dans cet exemple, les ensembles A et A^R sont respectivement composés de 12 et 8 alternatives (voir Tables 4.1 et 4.2). Ces alternatives sont décrites suivant $m = 7$ critères binaires. L'ensemble des critères est $[m] = \{a, b, c, d, e, f, g\}$. Les Tables 4.1

et 4.2 fournissent à la fois la représentation ensembliste (dernière colonne) et vectorielle des alternatives. Chaque alternative de l'ensemble \mathbf{A} (respectivement \mathbf{A}^R) est symbolisée par x^{id} (respectivement r^{id}) où id est l'entier en base décimale correspondant à la conversion de son vecteur d'évaluation.

L'analyste a pour rôle d'aider le décideur à opérer le choix de $k = 3$ alternatives parmi les 12 que compte l'ensemble \mathbf{A} .

TABLE 4.1 – Description de l'ensemble \mathbf{A} .

| | a | b | c | d | e | f | g | |
|-----------|---|---|---|---|---|---|---|-------|
| x^{113} | 1 | 1 | 1 | 0 | 0 | 0 | 1 | abcg |
| x^{91} | 1 | 0 | 1 | 1 | 0 | 1 | 1 | acdfg |
| x^{69} | 1 | 0 | 0 | 0 | 1 | 0 | 1 | aeg |
| x^{37} | 0 | 1 | 0 | 0 | 1 | 0 | 1 | beg |
| x^{98} | 1 | 1 | 0 | 0 | 0 | 1 | 0 | abf |
| x^{116} | 1 | 1 | 1 | 0 | 1 | 0 | 0 | abce |
| x^{104} | 1 | 1 | 0 | 1 | 0 | 0 | 0 | abd |
| x^{78} | 1 | 0 | 0 | 1 | 1 | 1 | 0 | adef |
| x^{92} | 1 | 0 | 1 | 1 | 1 | 0 | 0 | acde |
| x^{46} | 0 | 1 | 0 | 1 | 1 | 1 | 0 | bdef |
| x^{56} | 0 | 1 | 1 | 1 | 0 | 0 | 0 | bcd |
| x^{30} | 0 | 0 | 1 | 1 | 1 | 1 | 0 | cdef |

Le décideur fournit, à cet effet, l'ensemble suivant de préférences :

$$\mathbb{P} = \{(r^{73}, r^{45}), (r^{45}, r^{15}), (r^{15}, r^{66}), (r^{66}, r^{84}), (r^{84}, r^{52}), (r^{52}, r^{22}), (r^{22}, r^{28})\} \quad (4.4)$$

Il s'agit d'un ordre total sur l'ensemble des alternatives de référence \mathbf{A}^R (que le décideur connaît bien). L'ensemble \mathbb{P} est consistant comme le prouve la Table 4.3 qui compile les scores des alternatives de référence calculés à partir de la fonction ω suivante :

$$\omega = (a : 61, b : 44, c : 15, d : 10, e : 6, f : 24, g : 161) \quad (4.5)$$

À partir de l'ensemble \mathbb{P} (Équation 4.4), on déduit sur \mathbf{A} un ensemble de comparaisons par paire appartenant à la relation nécessaire (Définition 4.3). Quelques-unes de ces comparaisons par paire sont représentées à la Figure 4.1. Sur cette figure, l'arc reliant une alternative x à une alternative y symbolise le fait que « x est nécessairement préférée à y ». On a fait le choix d'y omettre, pour des raisons de lisibilité, les arcs de réflexivité et de transitivité. L'ensemble des arcs représentés permet d'identifier les alternatives **choix nécessaires** et **choix impossibles** (Propositions 4.3 et 4.4) qui pourront faire l'objet de conclusions robustes à expliquer au décideur.

TABLE 4.2 – Description de l'ensemble \mathbf{A}^R .

| | a | b | c | d | e | f | g | |
|----------|---|---|---|---|---|---|---|------|
| r^{73} | 1 | 0 | 0 | 1 | 0 | 0 | 1 | adg |
| r^{45} | 0 | 1 | 0 | 1 | 1 | 0 | 1 | bdeg |
| r^{15} | 0 | 0 | 0 | 1 | 1 | 1 | 1 | defg |
| r^{66} | 1 | 0 | 0 | 0 | 0 | 1 | 0 | af |
| r^{84} | 1 | 0 | 1 | 0 | 1 | 0 | 0 | ace |
| r^{52} | 0 | 1 | 1 | 0 | 1 | 0 | 0 | bce |
| r^{22} | 0 | 0 | 1 | 0 | 1 | 1 | 0 | cef |
| r^{28} | 0 | 0 | 1 | 1 | 1 | 0 | 0 | cde |

TABLE 4.3 – Scores des alternatives de référence (Table 4.2) suivant ω (Équation 4.5)

| | r^{73} | r^{45} | r^{15} | r^{66} | r^{84} | r^{52} | r^{22} | r^{28} |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|
| $\omega(\cdot)$ | 232 | 221 | 201 | 85 | 82 | 65 | 45 | 31 |

Une seule alternative est un **choix nécessaire** étant donné \mathbb{P} . Il s'agit de l'alternative x^{91} mise en évidence par la couleur **verte** : cette alternative est l'une des $k = 3$ qui, étant donné \mathbb{P} (Équation 4.4), seront recommandées au décideur. En effet, on observe sur la Figure 4.1 que l'alternative x^{91} est nécessairement préférée à $n - k = 9$ autres alternatives (toutes sauf x^{113} et x^{78}) de l'ensemble \mathbf{A} . Sur cette même figure, les alternatives **choix impossibles** sont identifiables grâce à la couleur **rouge**. Pour rappel, il s'agit d'alternatives que l'information préférentielle \mathbb{P} fournie par le décideur suffit à disqualifier. Dans le présent exemple, il s'agit des alternatives x^{37} , x^{92} , x^{56} , x^{46} et x^{30} auxquelles sont nécessairement préférées au moins 3 autres alternatives de l'ensemble \mathbf{A} .

4.3.2 Justification des conclusions robustes de l'exemple introductif

Nous nous intéressons ici à la justification des conclusions robustes de types **choix nécessaire** et **choix impossible** de l'Exemple 4.3.1. Les Propositions 4.3 et 4.4 nous rappellent sous quelles conditions une alternative $x \in \mathbf{A}$ (respectivement $y \in \mathbf{A}$) peut être considérée comme un **choix nécessaire** (respectivement **choix impossible**). Conformément à ces propositions, nous rappelons que la justification d'un **choix nécessaire** x (respectivement **choix impossible** y) consiste à expliquer au moins $(n - k) = 9$ comparaisons par paire (x, y') , $y' \in \mathbf{A} \setminus \{x\}$ (respectivement $k = 3$ comparaisons par paire (x', y) , $x' \in \mathbf{A} \setminus \{y\}$). À cette fin, nous tenterons dans un premier temps d'expliquer l'ensemble des conclusions robustes de l'Exemple 4.3.1 à l'aide de *preference swaps nécessaires*

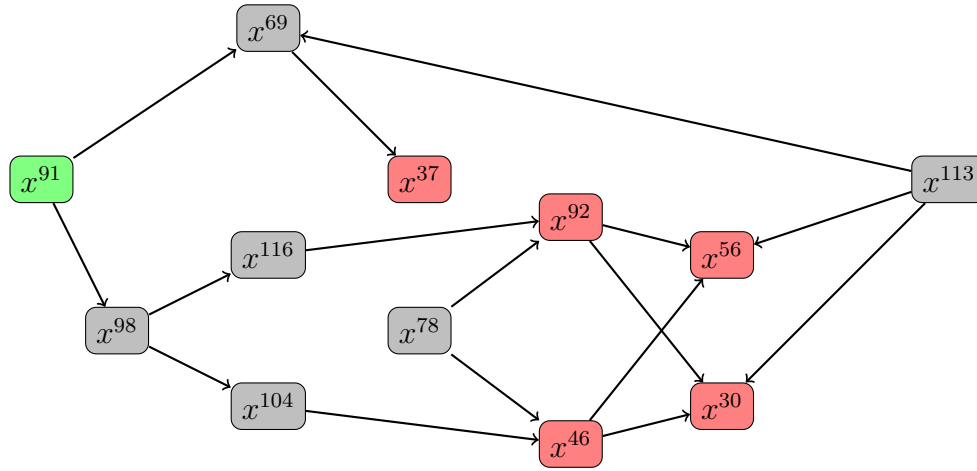


FIGURE 4.1 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}^A$ sans les arcs de réflexivité et de transitivité (Exemple 4.3.1).

([Belahcène et al., 2017]). Ceci nous permettra d'observer les limites de ce type d'explication. Ensuite, nous montrerons comment l'élargissement des comparaisons par paire considérées comme admises aux éléments de la $\mathbb{P}\mathbb{I}$ de même que le recours à d'autres schémas déductifs rendra possible la justification de toutes les conclusions robustes de l'Exemple 4.3.1.

Dans cette partie, pour des raisons de lisibilité, les explications de comparaisons par paire seront produites sous la forme de diagrammes.

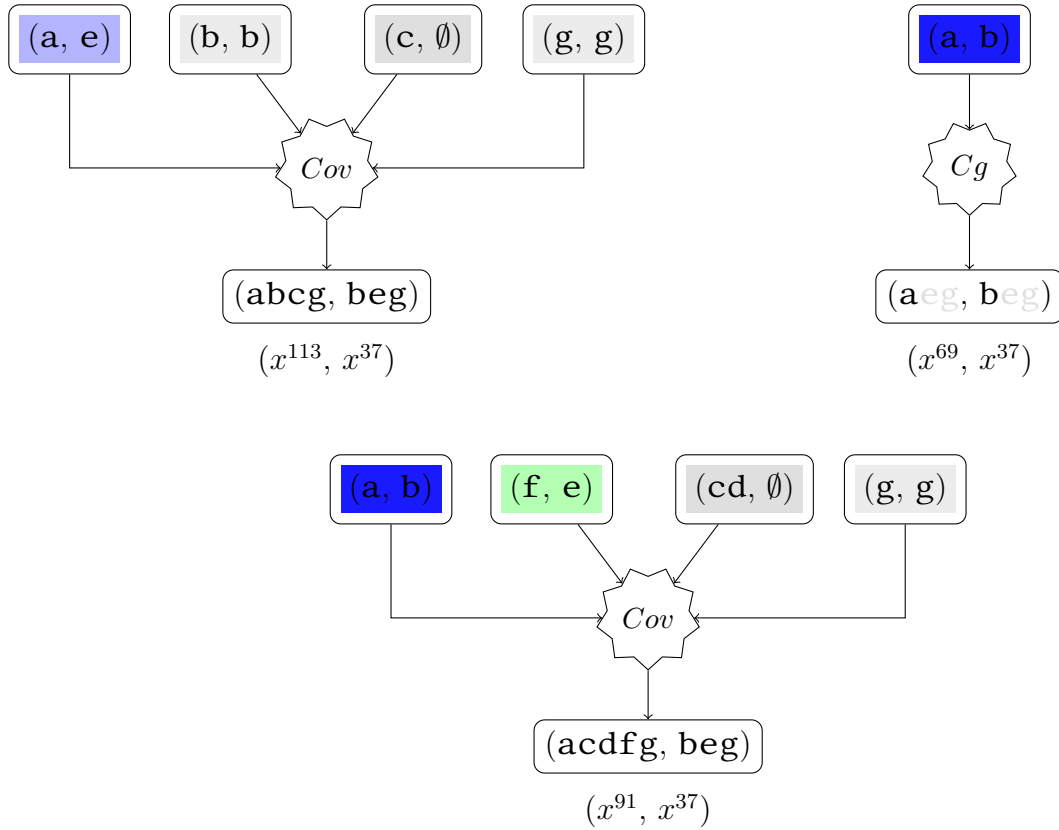
Justification des conclusions robustes x^{37} , x^{92} , x^{56} et x^{30} à l'aide de *preference swaps nécessaires*

L'ensemble $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ (Équation 4.3) des *preference swaps* déduits des déclarations de préférence $\mathbb{P}\mathbb{I}$ du décideur se présente comme suit³ :

$$\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}} = \{ (a, b), (a, c), (a, d), (a, e), (a, f), (b, c), (b, d), (b, e), (b, f), (f, c), (f, d), (f, e) \} \quad (4.6)$$

Les Figures 4.2, 4.3, 4.4 et 4.5 détaillent les justifications des *choix impossibles* x^{37} , x^{92} , x^{56} et x^{30} .

3. Cet ensemble est obtenu en résolvant tous les programmes linéaires décrits par la Proposition 4.1 ; le nombre total de programmes linéaires à résoudre à cette fin est de l'ordre $\mathcal{O}(m^2)$. On rappelle qu'un programme linéaire se résout en temps polynomial à l'aide de la méthode des points intérieurs.

FIGURE 4.2 – Justification du choix impossible x^{37}

Impossibilité de justification des conclusions robustes x^{46} et x^{91} à l'aide de *preference swaps nécessaires*

À la différence des conclusions robustes x^{37} , x^{92} , x^{56} et x^{30} traitées précédemment, le **choix impossible** $x^{46} \equiv bdef$ ne peut être justifié à l'aide de *preference swaps nécessaires*. En effet, sur la Figure 4.1, on remarque que quatre alternatives de l'ensemble A sont nécessairement préférées à x^{46} , il s'agit des alternatives : x^{91} , x^{98} , x^{104} et x^{78} . Pour que le **choix impossible** $x^{46} \equiv bdef$ puisse être justifié à l'aide de *preference swaps nécessaires*, il fallait (Proposition 4.4) que, pour au moins trois des quatre comparaisons par paire (x^{91}, x^{46}) , (x^{98}, x^{46}) , (x^{104}, x^{46}) et (x^{78}, x^{46}) , il existât ;) une application injective ϕ (Définition 4.6) de leur ensemble de critères "**con**" vers celui de critères "**pro**" telle que les *preference swaps* représentant les correspondances de ϕ appartiennent tous à S_{NPI} (Équation 4.6). À partir la Table 4.4 et de l'Équation 4.6, on conclut que seule la comparaison par paire (x^{78}, x^{46}) est explicable à l'aide de *preference swaps nécessaires* : les comparaisons par paire (x^{98}, x^{46}) et (x^{104}, x^{46}) ne le sont pas – pour des raisons "structurelles" : strictement moins de critères **pro** que de critères **con** alors que ϕ est

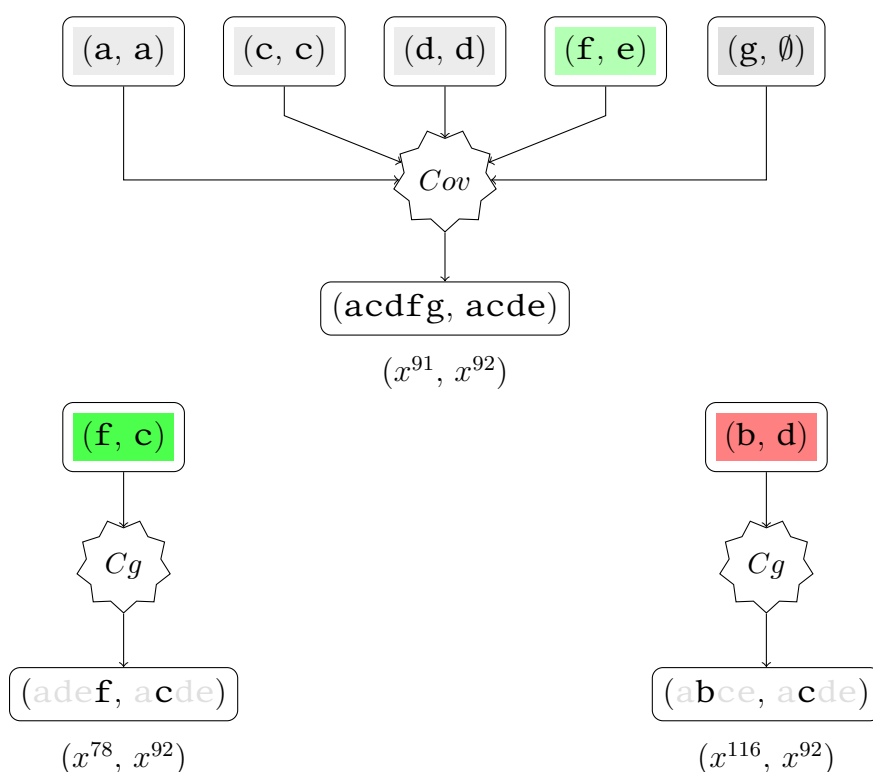


FIGURE 4.3 – Justification du choix impossible x^{92}

une application – et la comparaison (x^{91}, x^{46}) n'est pas explicable à l'aide de *preference swaps nécessaires* car ni (c, b) ni (c, e) n'appartiennent à $S_{N_{PI}}$ (Équation 4.6).

TABLE 4.4 – Justification du **choix impossible** x^{46} : Table d'ensembles **pro** et **con**.

| (x, y) | (x^{91}, x^{46}) | (x^{98}, x^{46}) | (x^{104}, x^{46}) | (x^{78}, x^{46}) |
|---------------------|--------------------|--------------------|---------------------|--------------------|
| pro (x, y) | {a, c} | {a} | {a} | {a} |
| con (x, y) | {b, e} | {d, e} | {b, e} | {b} |

Un raisonnement similaire (voir Table 4.5) permet de montrer que le (seul) **choix nécessaire** $x^{91} \equiv acdfg$ de l'Exemple 4.3.1 n'est pas explicable à l'aide de *preference swaps nécessaires*.

TABLE 4.5 – Justification du **choix nécessaire** x^{91} : Table d'ensembles **pro** et **con**.

| (x, y) | (x^{91}, x^{98}) | (x^{91}, x^{116}) | (x^{91}, x^{104}) | (x^{91}, x^{78}) | (x^{91}, x^{46}) |
|---------------------|--------------------|---------------------|---------------------|--------------------|--------------------|
| pro (x, y) | {c, d, g} | {d, f, g} | {c, f, g} | {c, g} | {a, c, g} |
| con (x, y) | {b} | {b, e} | {b} | {e} | {b, e} |

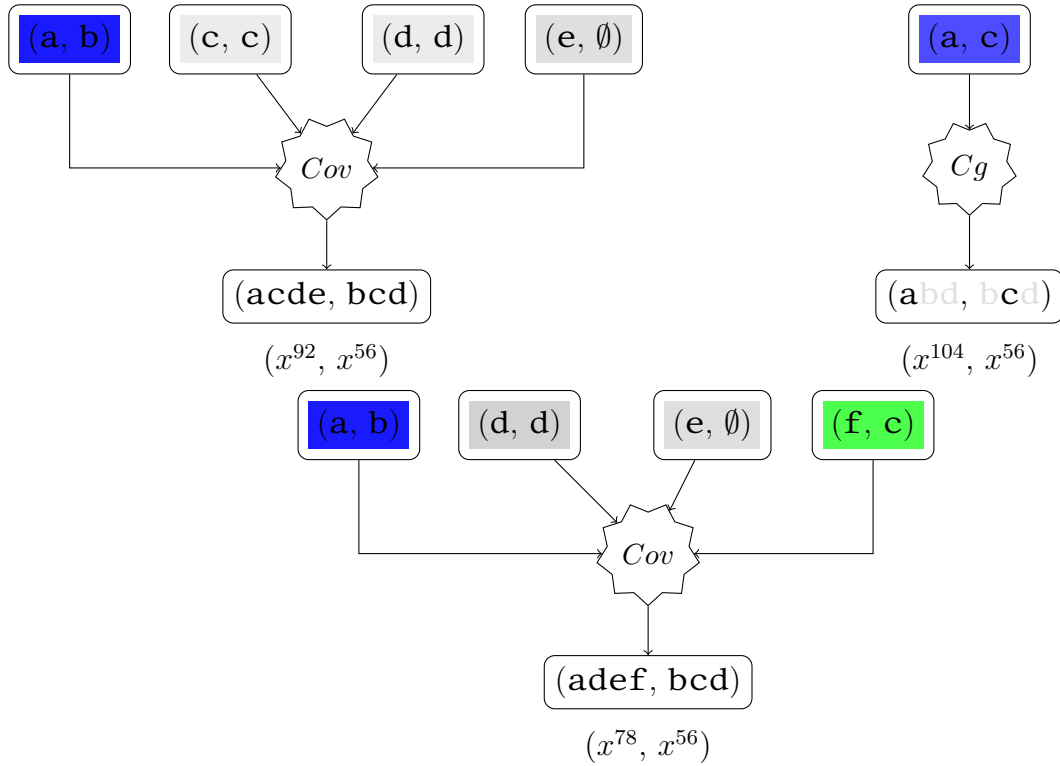


FIGURE 4.4 – Justification du choix impossible x^{56}

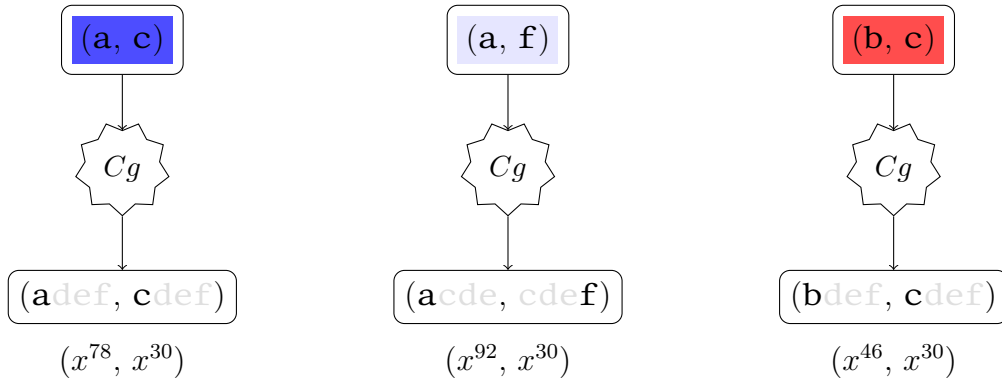


FIGURE 4.5 – Justification du choix impossible x^{30}

Dans cette Table, nous avons uniquement recensé les comparaisons par paire (x^{91}, y) , $y \in \mathbf{A} \setminus \{x^{91}\}$ non explicables à partir de *preference swaps nécessaires*⁴. Il y en a en tout cinq sur les dix alternatives auxquelles x^{91} est nécessairement préférée alors que, pour

4. Les comparaisons par paire (x^{91}, y) , $y \in \mathbf{A} \setminus \{x^{91}\}$ explicables à partir de *preference swaps* sont détaillées à la Figure 4.7.

justifier son **choix nécessaire**, il aurait fallu (Proposition 4.3) pouvoir expliquer au moins neuf comparaisons par paire (x^{91}, y) , $y \in \mathbf{A} \setminus \{x^{91}\}$.

Justification des conclusions robustes x^{46} et x^{91} à partir d'autres types d'explication

Comme indiqué en début de section, la principale contribution de ce chapitre consiste à étendre les travaux de [Belahcène et al., 2017] qui proposent une explication des déductions de la relation nécessaire à l'aide de *preference swaps nécessaires*. Pour ce faire, nous avons montré (sous-section 4.2.3) comment cette proposition s'inscrit dans le cadre plus général de l'approche de l'explication que nous adoptons dans ce travail et suggéré, conformément à celle-ci, que soient considérés d'autres *contextes* qui élargissent l'ensemble des comparaisons par paire admises aux déclarations de préférences du décideur et utilisent d'autres schémas déductifs que ceux correspondant à l'explication à l'aide de *preference swaps nécessaires*.

Ici, nous montrerons comment il est possible de justifier les **choix impossible** x^{46} et **choix nécessaire** x^{91} . Dans les explications produites à cette fin, ont été utilisés, en plus des *preference swaps* de $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ (Équation 4.6), les éléments de la $\mathbb{P}\mathbb{I}$ (Équation 4.4) suivants :

$$\begin{aligned} (r^{73}, r^{45}) &\equiv (\text{adg}, \text{bdeg}) \\ (r^{45}, r^{15}) &\equiv (\text{bdeg}, \text{defg}) \\ (r^{15}, r^{66}) &\equiv (\text{defg}, \text{af}) \\ (r^{66}, r^{84}) &\equiv (\text{af}, \text{ace}) \end{aligned}$$

La Figure 4.6 présente les explications de trois comparaisons par paire qui suffisent à justifier le **choix impossible** x^{46} tandis que les Figures 4.7, 4.8 et 4.9 concernent la justification du **choix nécessaire** de x^{91} . On y remarque l'utilisation de schéma déductif de *niveau* 4 tel que le schéma $cg/tr/cg/tr/cg$ (explication de (x^{91}, x^{104}) sur la Figure 4.8). Sur la Figure 4.9, on remarque la réutilisation de la comparaison par paire (x^{91}, x^{98}) préalablement expliquée dans celle de la comparaison par paire (x^{91}, x^{116}) : il s'agit de l'application de la procédure d'évolution dynamique du contexte détaillée au chapitre précédent (sous-section 3.4.1).

Toutes ces explications ont été calculées à partir d'algorithmes dont la structure est très similaire à celle des algorithmes des familles Tr , Cg et Cor décrites à la Section 3.2. Comme on peut le voir sur les Figures 4.6, 4.7, 4.8 et 4.9, dans nos algorithmes, les schémas de couverture Cov ont été limités aux seules comparaisons par paire de type *preference swaps nécessaires*. Ce faisant, les gains d'expressivité observés et mesurés à la sous-section 4.3.3 sont uniquement imputables à l'ajout des éléments de la $\mathbb{P}\mathbb{I}$ à l'ensemble des comparaisons par paire admises et à l'application à ces derniers des mécanismes de transitivité et de congruence.

Le code correspondant est accessible sur le dépôt que voici ⁵. Il convient de signaler pour finir que les explications retournées par nos algorithmes sont de *niveau* (Définition 2.10) au plus égal à 5.

4.3.3 Expérimentations numériques

Dans la précédente sous-section, nous avons perçu les limites des *preference swaps nécessaires* de [Belahcène et al., 2017] à justifier certaines conclusions robustes de l'Exemple 4.3.1. L'objectif de cette sous-section est de quantifier expérimentalement le "potentiel justificatif" des *preference swaps nécessaires* ⁶ et de le comparer à celui des types d'explication que retournent les algorithmes que nous avons implémentés.

Dans toute la sous-section, nous désignerons sous le terme d'*instance*, le quadruplet $(m, \mathbf{A}, \mathbf{A}^R, \omega)$ censé résumer un problème fictif ⁷ de situation d'aide à la décision multicritère :

- m est la dimension du problème (nombre total de critères),
- \mathbf{A} est l'ensemble des alternatives sur lequel doit porter la recommandation,
- \mathbf{A}^R est l'ensemble des alternatives de référence, que le décideur connaît bien et
- la fonction de score ω représente les préférences de ce dernier.

Description du protocole expérimental

Nous avons constitué 4 grands groupes d'instances aléatoires en fonction de la taille du problème qu'elles représentent : $m \in \llbracket 7; 10 \rrbracket$. Étant donné m , chaque instance est obtenue (indépendamment des autres instances du même groupe) en générant aléatoirement $n = |\mathbf{A}|$ alternatives sur lesquelles doit porter la recommandation, $r = |\mathbf{A}^R|$ alternatives de référence et une fonction de score ω qui, appliquée à ces dernières, simule les déclarations de préférences du décideur (PII). Chaque fonction de score est tirée (pour chaque valeur $m \in \llbracket 7; 10 \rrbracket$) suivant la méthode de [Butler et al., 1997]. Quant aux tirages aléatoires des ensembles \mathbf{A} et \mathbf{A}^R , ils ont été réalisés (de façon indépendante l'une de l'autre) de la manière suivante :

- Tirage uniforme sans remise de n ou r entiers naturels entre 0 et $2^m - 1$,
- Conversion en binaire des entiers tirés,

5. <https://github.com/tetras92/SchemasDeductifsEnAMCD.git>

6. Dans [Amoussou et al., 2023a], de similaires expérimentations numériques ont été conduites. À la différence de celles-ci, elles ont porté sur des comparaisons par paire spécifiques considérées individuellement plutôt que sur des ensembles de comparaisons par paire constitutives d'une conclusion robuste. De plus, ont été écartées les comparaisons par paire dont l'inexplicabilité est "structurelle", liée à la cardinalité des ensembles **pro** et **con** (voir comparaisons par paire (x^{98}, x^{46}) et (x^{104}, x^{46}) de la Table 4.4)

7. Fictif car la fonction de score ω représentant les préférences du décideur, dans les situations réelles d'aide à la décision, n'est pas connue et parce que le décideur est supposé s'y conformer parfaitement.

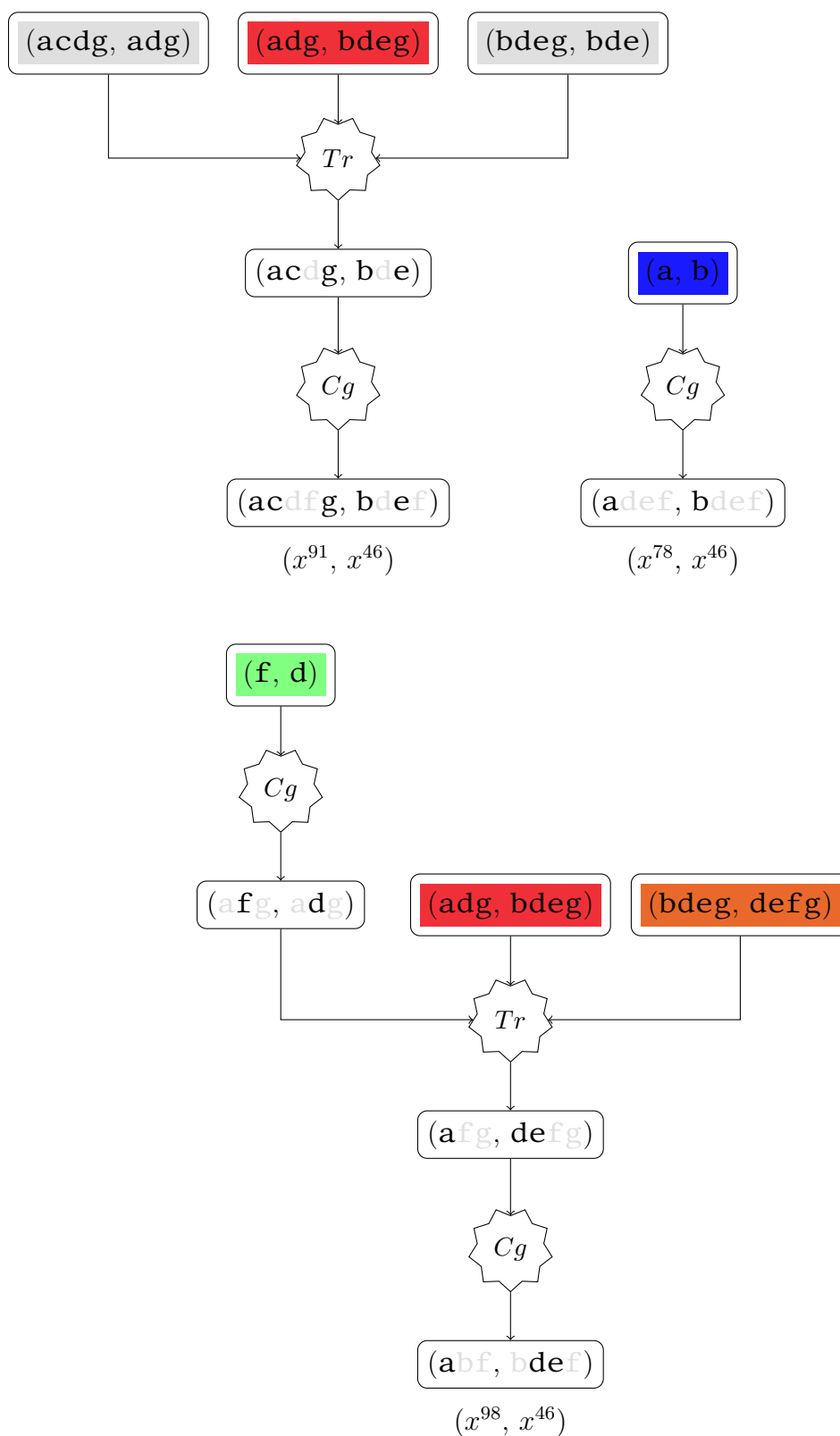


FIGURE 4.6 – Justification du choix impossible x^{46}

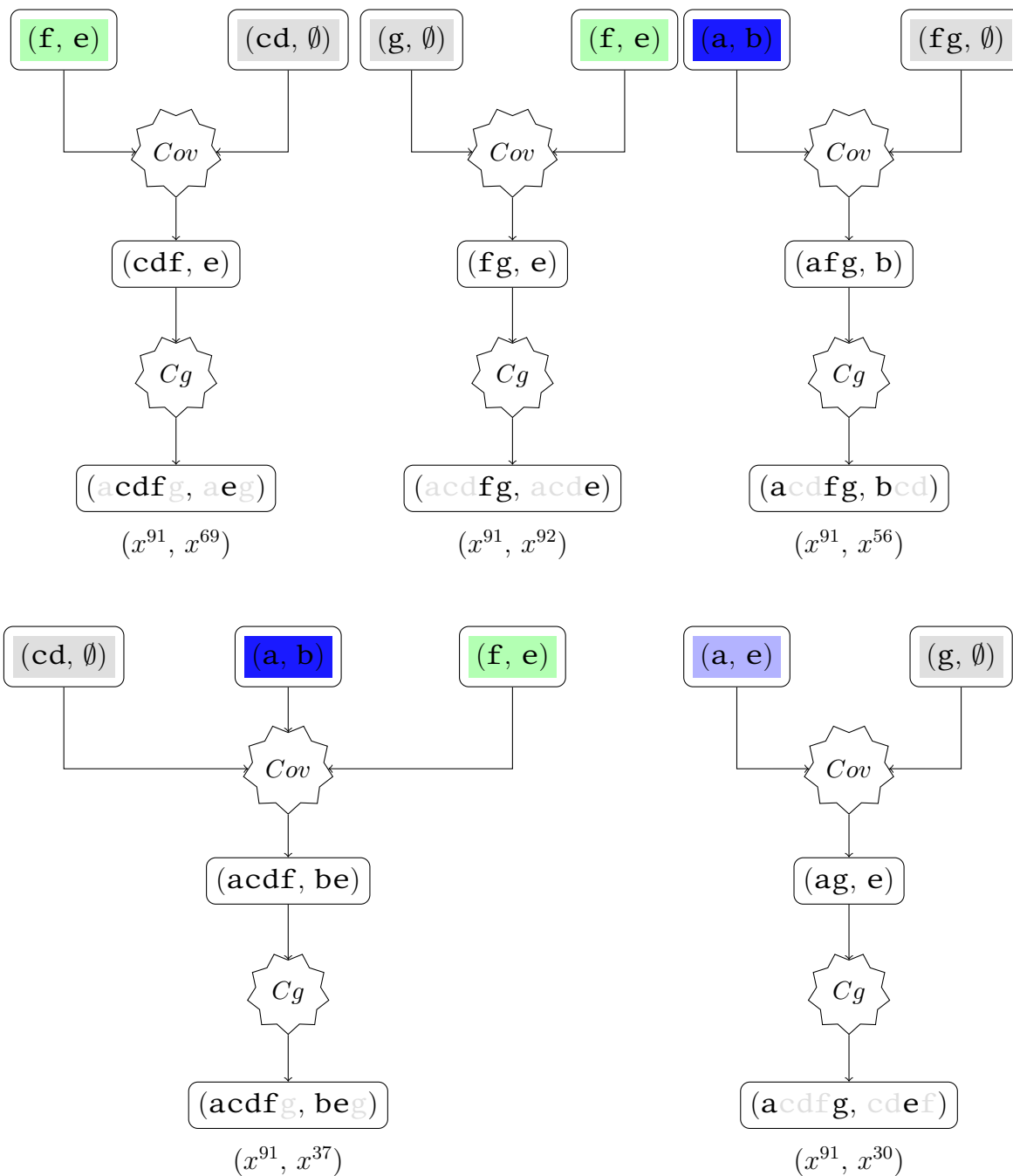


FIGURE 4.7 – Justification du choix nécessaire x^{91} (partie 1 / 3)

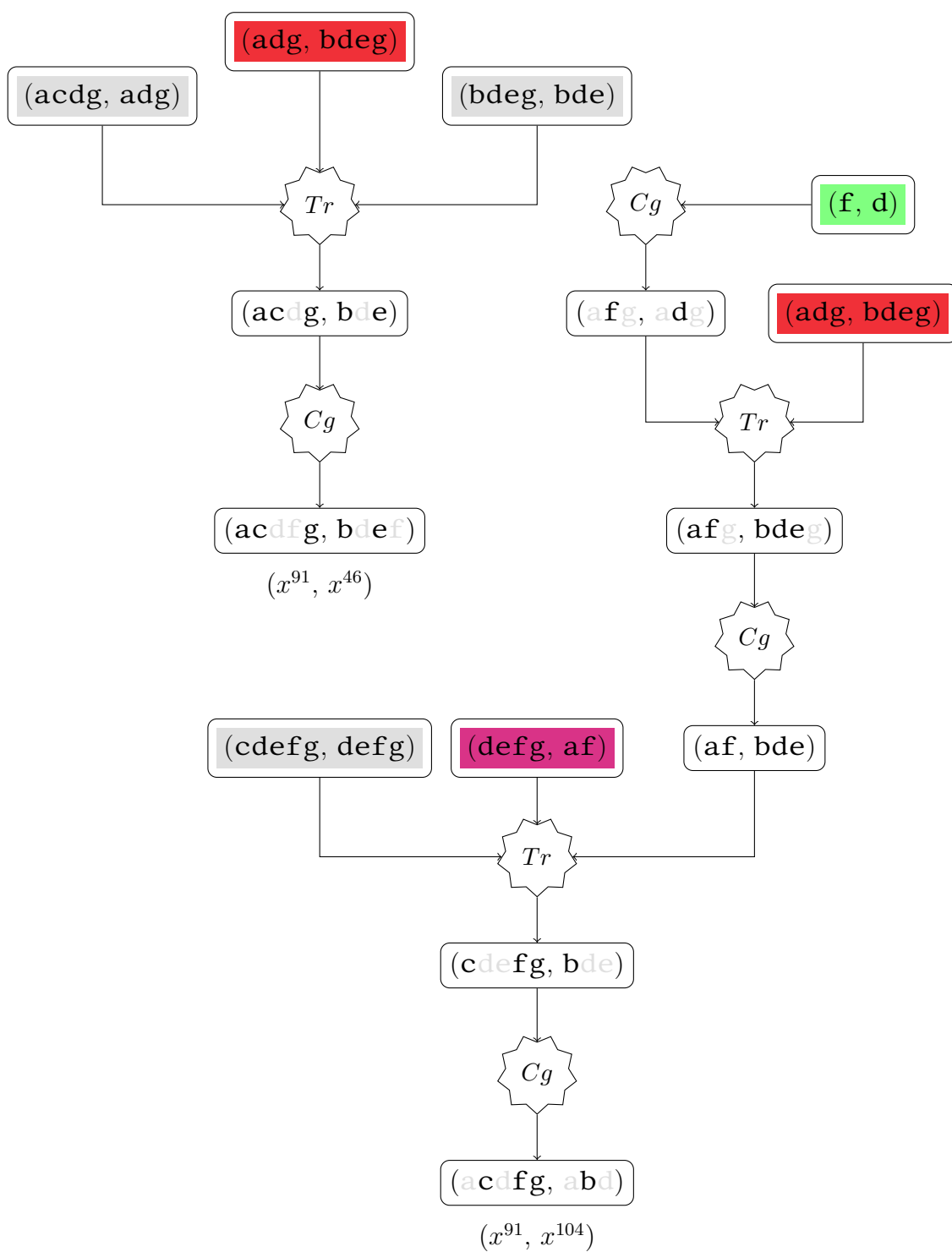
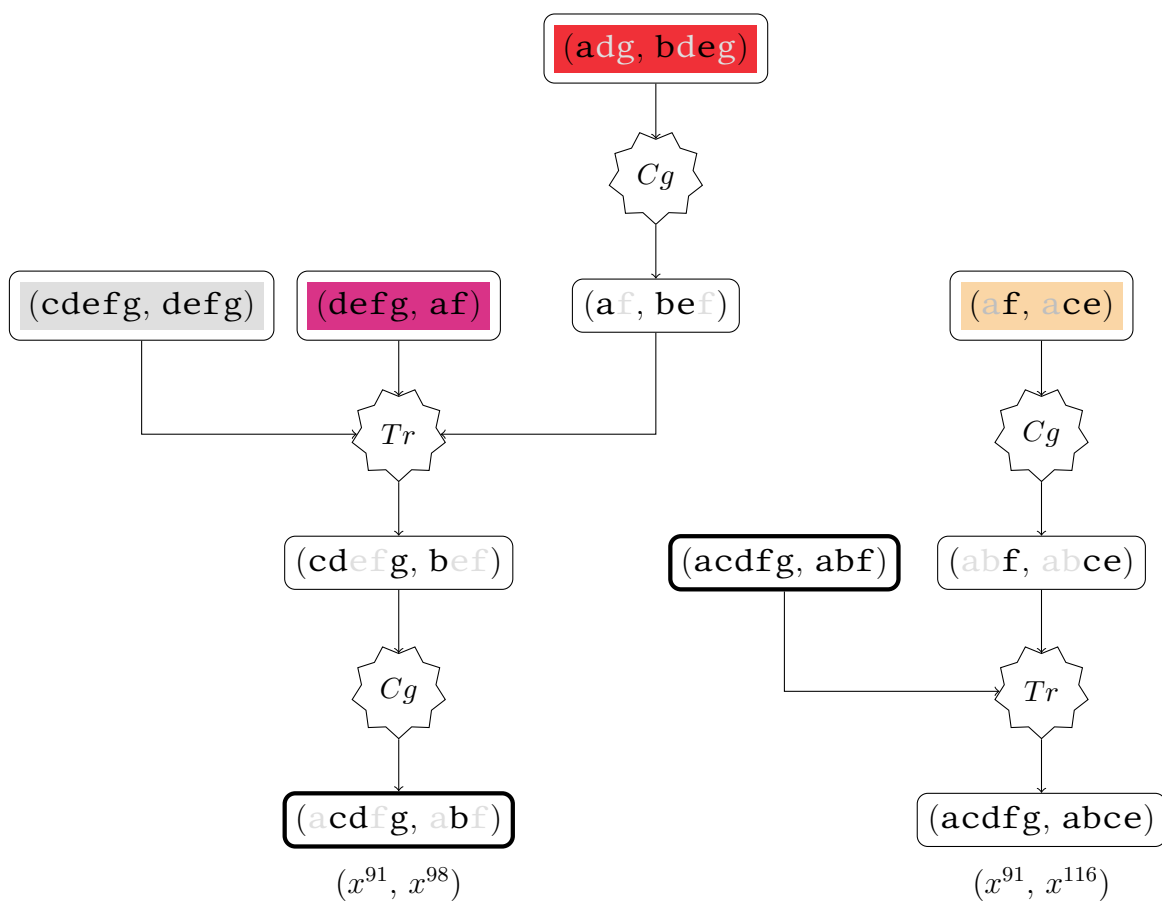


FIGURE 4.8 – Justification du choix nécessaire x^{91} (partie 2 / 3)

FIGURE 4.9 – Justification du choix nécessaire x^{91} (partie 3 / 3)

- Conversion de chaque entier x sous numérotation binaire en alternative : chaque alternative (sous sa forme ensembliste) est composée de l'ensemble des critères i tel que le i -ème bit de x vaut 1.

Il convient de signaler que l'on ne retient une instance que si elle vérifie les deux propriétés suivantes :

- (P1) *Significativité des m critères* : cette propriété garantit que les ensembles \mathbf{A} et \mathbf{A}^R sont bien définis sur m critères. Elle peut s'exprimer mathématiquement comme suit :

$$\forall i \in [m], \exists x, y \in \mathbf{A} \mid i \in x \text{ et } i \notin y \quad (4.7)$$

$$\forall i \in [m], \exists x, y \in \mathbf{A}^R \mid i \in x \text{ et } i \notin y \quad (4.8)$$

- (P2) *Absence de dominance au sens de Pareto* : cette propriété justifie, dans le cas de l'ensemble \mathbf{A} , l'utilité de la fonction de score ω dans la détermination de la recommandation. Pour ce qui est de l'ensemble \mathbf{A}^R , cette propriété garantit d'une certaine façon que chacune des comparaisons par paire fournies par le décideur est informative (par opposition à triviale). Elle peut s'exprimer mathématiquement de la façon suivante :

$$\forall x, y \in \mathbf{A}, \exists i, j \in [m] \mid i \in x \text{ et } j \notin x \Rightarrow i \notin y \text{ et } j \in x \quad (4.9)$$

$$\forall x, y \in \mathbf{A}^R, \exists i, j \in [m] \mid i \in x \text{ et } j \notin x \Rightarrow i \notin y \text{ et } j \in x \quad (4.10)$$

Paramètres retenus et autres éléments de compréhension : Pour chaque valeur $m \in \llbracket 7; 10 \rrbracket$, nous avons – arbitrairement – (comme dans l'Exemple 4.3.1) pris $n = 12$ et $k = 3$. Les valeurs de r correspondant au nombre d'alternatives de référence \mathbf{A}^R ont été choisies dans l'intervalle $m \pm 2^8$. L'information préférentielle fournie par le décideur est, comme dans l'Exemple 4.3.1, le rangement des r alternatives de référence suivant ω et la recommandation consiste à choisir $k = 3$ alternatives parmi les $n = 12$ que compte l'ensemble \mathbf{A} . Ainsi, pour chaque instance, étant donné \mathbb{P} , on peut déduire par calcul de la relation nécessaire entre 0 et 12 conclusions robustes de type **choix nécessaire** et **choix impossible**.

Pour chaque groupe d'instances, on a généré suffisamment d'instances pour obtenir approximativement 1000 **choix nécessaires** ou **choix impossibles** que l'on tentera de justifier dans un premier temps, à l'aide (exclusivement) de *préférence swaps nécessaires* puis, à partir des autres algorithmes que nous avons implémentés et en appliquant la procédure d'évolution dynamique du contexte (sous-section 3.4.1). L'unique mesure que nous réalisons est : la proportion de conclusions robustes explicables.

8. Intervalle défini de façon arbitraire aussi mais dépendant de la dimension m du problème de sorte à ne pas perdre des "chances" de déduction de conclusions robustes au fur et à mesure qu'augmente m .

TABLE 4.6 – Résultats d'analyses descriptives du protocole expérimental

| m | 7 | 8 | 9 | 10 | 11 |
|---------------------------------|------|------|------|------|------|
| # instances générées | 265 | 302 | 346 | 395 | 407 |
| # conclusions robustes déduites | 1001 | 1000 | 1001 | 1002 | 1001 |

Résultats et discussions

Dans la Table 4.6, sont consignés pour chaque valeur $m \in \llbracket 7; 10 \rrbracket$, le nombre d'instances aléatoires qu'il a fallu générer pour obtenir à peu près 1000 conclusions robustes et le nombre exact de conclusions robustes obtenus. On remarque que pour obtenir un nombre constant de conclusions robustes, il faut générer davantage d'instances, au fur et à mesure qu'augmente la valeur de m (de 265 à 407 instances ici).

La Figure 4.10 résume les proportions observées de conclusions robustes que les algorithmes implémentés ont réussi à justifier dans chaque groupe d'instances. Les valeurs numériques correspondant à ces proportions sont inscrites dans la Table 4.7. Par exemple, les algorithmes implémentés parviennent à justifier la quasi-totalité (99,0%) des 1001 **choix nécessaires** et **choix impossibles** déduits dans le groupe d'instances décrites sur $m = 7$ critères. On observe que cette proportion décroît au fur et à mesure qu'augmente m pour tomber à un peu moins de 9 conclusions robustes justifiables sur 10 (88,9%) lorsque $m = 10$. Sur la Figure 4.10, nous avons également pris le soin de représenter (en **bleu**) la part de conclusions robustes justifiables exclusivement à l'aide de *preference swaps nécessaires* afin de pouvoir quantifier l'intérêt de la prise en compte des éléments de la $\mathbb{P}\llbracket$ et du recours à une palette plus variée de schémas déductifs dans la production d'explication à des fins de justification. On observe que seulement 61,1% des 1001 **choix nécessaires** et **choix impossibles** du groupe d'instances décrites sur $m = 7$ critères peuvent être justifiés uniquement à l'aide de *preference swaps nécessaires*. Cette proportion chute en dessous de 50% dès que m passe à 8 et est à peine supérieure à un quart (27%) dans le groupe d'instances décrites sur $m = 10$ critères.

TABLE 4.7 – Valeurs numériques de la Figure 4.10.

| m | <i>Preference swaps</i> | Tous types |
|-----|-------------------------|------------|
| 10 | 27.0% | 88.9% |
| 9 | 38.3% | 95.2% |
| 8 | 49.1% | 98.6% |
| 7 | 61.1% | 99.0% |

Globalement, les expérimentations numériques conduites ici révèlent que les *preference swaps nécessaires* sont, à eux seuls, insuffisants pour permettre la justification de conclusions robustes déduites à partir des déclarations de préférences du décideur. Elles

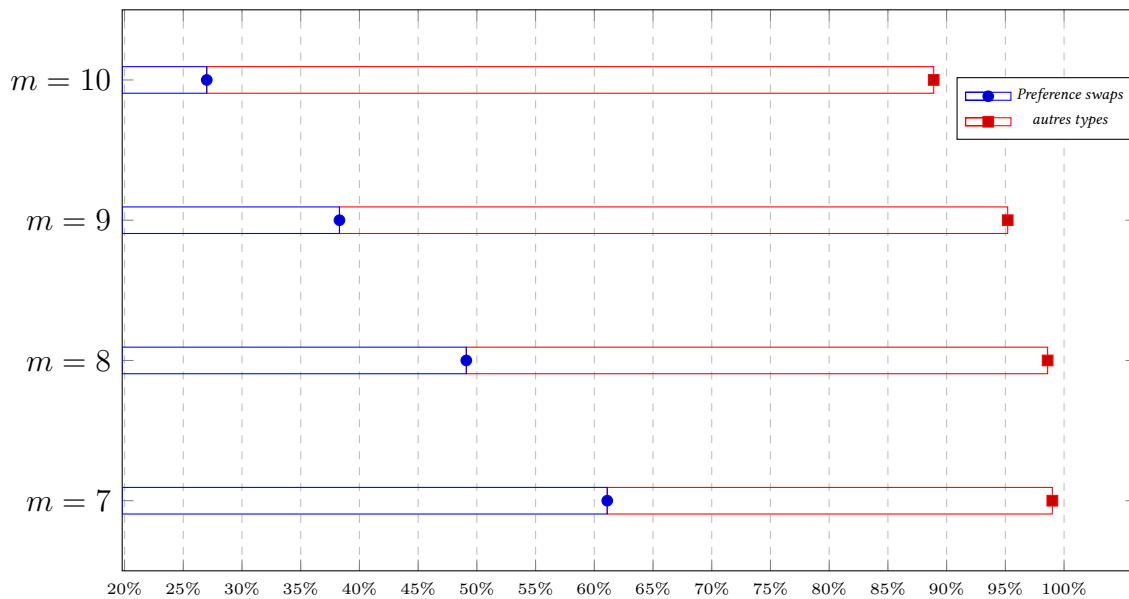


FIGURE 4.10 – Proportions de conclusions robustes justifiables.

montrent qu'en leur adjoignant ces dernières, on arrive à justifier une part beaucoup plus importante de **choix nécessaires** et de **choix impossibles** sans prendre le risque que les explications correspondantes ne paraissent inintelligibles pour le décideur. En effet, elles reprennent ses déclarations de préférences qui sont agencées suivant des schémas déductifs reposant sur les propriétés de transitivité et d'indépendance auxquelles sont a priori familiers la plupart des décideurs. Il convient d'indiquer pour finir que l'utilisation de ces algorithmes dans des situations réelles portant sur des problèmes de tailles similaires à celles des expérimentations, peut être envisagée. En effet, le temps de calcul moyen observé d'une conclusion robuste de type **choix nécessaire** ou **choix impossible** sur des instances de taille $m = 10$ critères est inférieur à 15 secondes.

4.4 Fonction d'élicitation de l'explication

L'objectif de cette section est de montrer que l'explication de conclusions robustes peut permettre à l'analyste, engagé dans un processus interactif avec le décideur, de parvenir à la détermination de la recommandation finale. Une telle explication, à la différence de celle proposée dans la section précédente, ne peut pas être uniquement composée de comparaisons par paire appartenant à la relation nécessaire. En effet, valider un élément de la relation nécessaire n'apporte aucune information préférentielle supplémentaire. Le contester, au contraire remet en cause l'hypothèse de représentabilité des préférences du décideur par un modèle additif en conduisant à une inconsistance. Les explications proposées, à l'image des *explications questionnables* de [Amoussou et al.,

2023a], seront composées de *preference swaps*. À la différence de celles-ci, elles pourront également mentionner un ou plusieurs éléments de la $\mathbb{P}\mathbb{I}$, l'ensemble des déclarations de préférences du décideur. Pour constituer une preuve de la déduction expliquée, les explications seront produites dans le *contexte* $(\succsim_\omega, \mathbb{P}\mathbb{I} \cup \mathbb{S}_{\succsim_\omega}, tr/cg)$ où \succsim_ω est la relation de préférences induite par la fonction de score ω compatible avec la $\mathbb{P}\mathbb{I}$ ($\omega \in \Omega_{\mathbb{P}\mathbb{I}}$).

Dans cette section, nous supposons que les réactions du décideur aux explications qui lui sont proposées sont « fidèles » à son modèle de préférences : il ne se trompe pas dans l'appréciation des explications qui lui sont soumises. Sous cette hypothèse, la collecte d'information préférentielle supplémentaire s'effectue par l'intermédiaire de *preference swap(s)* contesté(s) ou validé(s) par le décideur.

La section est organisée comme suit : elle sera introduite par la présentation d'un exemple de situation d'aide à la décision (sous-section 4.4.1) au travers de laquelle nous illustrerons la fonction d'élicitation que peut remplir l'explication de conclusions robustes (sous-section 4.4.2). Ensuite, nous proposerons une ébauche de protocole interactif centré sur l'utilisation de telles explications à des fins d'élicitation (sous-section 4.4.3) puis achèverons la section par la formalisation et la modélisation du problème de leur calcul (sous-section 4.4.4).

4.4.1 Exemple introductif

Les ensembles \mathbf{A} et \mathbf{A}^R de cet exemple sont respectivement composés de 8 et 5 alternatives (voir Tables 4.8 et 4.9). Ces alternatives sont décrites suivant $m = 6$ critères binaires. L'ensemble des critères est $[m] = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{e}, \mathbf{f}\}$. Les Tables 4.8 et 4.9 permettent d'identifier les alternatives des ensembles \mathbf{A} et \mathbf{A}^R . On remarque que ces deux ensembles ne sont pas disjoints : $\mathbf{A} \cap \mathbf{A}^R = \{x^{50}, x^7\}$. Ici aussi, chaque alternative d'intérêt ($\in \mathbf{A}$) ou de référence ($\in \mathbf{A}^R$) est identifiée à l'aide d'un entier correspondant à la conversion en base décimale de son vecteur d'évaluation.

L'analyste a pour rôle d'aider le décideur à opérer le choix de $k = 1$ alternative parmi les 8 que compte l'ensemble \mathbf{A} .

Le décideur fournit, à cet effet, l'ensemble de préférences suivant :

$$\mathbb{P}\mathbb{I} = \left\{ (r^{43}, r^{24}), (r^{24}, x^{50}), (x^{50}, x^7), (x^7, r^{17}) \right\} \quad (4.11)$$

Il s'agit d'un ordre total sur l'ensemble des alternatives de référence \mathbf{A}^R (que le décideur connaît bien). L'ensemble $\mathbb{P}\mathbb{I}$ est consistant comme le prouve la Table 4.10 qui compile les scores des alternatives de référence calculés à partir de la fonction de score ω suivante :

$$\omega = (\mathbf{a} : 29, \mathbf{b} : 32, \mathbf{c} : 66, \mathbf{d} : 24, \mathbf{e} : 20, \mathbf{f} : 18) \quad (4.12)$$

À partir de l'ensemble $\mathbb{P}\mathbb{I}$ (Equation 4.11), on déduit l'ensemble $\mathbb{N}_{\mathbb{P}\mathbb{I}}^{\mathbf{A}}$ des comparaisons par paire de \mathbf{A} appartenant à la relation nécessaire (Définition 4.3). Sur le graphe

TABLE 4.8 – Description de l'ensemble A .

| | a | b | c | d | e | f | |
|----------|---|---|---|---|---|---|-----|
| x^{56} | 1 | 1 | 1 | 0 | 0 | 0 | abc |
| x^{26} | 0 | 1 | 1 | 0 | 1 | 0 | bce |
| x^{25} | 0 | 1 | 1 | 0 | 0 | 1 | bcf |
| x^{42} | 1 | 0 | 1 | 0 | 1 | 0 | ace |
| x^{11} | 0 | 0 | 1 | 0 | 1 | 1 | cef |
| x^{50} | 1 | 1 | 0 | 0 | 1 | 0 | abe |
| x^{37} | 1 | 0 | 0 | 1 | 0 | 1 | adf |
| x^7 | 0 | 0 | 0 | 1 | 1 | 1 | def |

TABLE 4.9 – Description de l'ensemble A^R .

| | a | b | c | d | e | f | |
|----------|---|---|---|---|---|---|------|
| r^{43} | 1 | 0 | 1 | 0 | 1 | 1 | acef |
| r^{24} | 0 | 1 | 1 | 0 | 0 | 0 | bc |
| x^{50} | 1 | 1 | 0 | 0 | 1 | 0 | abe |
| x^7 | 0 | 0 | 0 | 1 | 1 | 1 | def |
| r^{17} | 0 | 1 | 0 | 0 | 0 | 1 | bf |

TABLE 4.10 – Scores des alternatives de référence (Table 4.9) suivant ω (Équation 4.12)

| | r^{43} | r^{24} | x^{50} | x^7 | r^{17} |
|-----------------|----------|----------|----------|-------|----------|
| $\omega(\cdot)$ | 133 | 98 | 81 | 62 | 50 |

de la Figure 4.11, on a représenté l'ensemble $\mathbb{N}_{\mathbb{P}}^A$: les sommets sont les alternatives de A et l'arc reliant une alternative x à une alternative y symbolise le fait que « x est nécessairement préférée à y ». On a fait le choix d'y omettre, pour des raisons de lisibilité, les arcs de réflexivité et de transitivité. L'ensemble des arcs représentés permet d'identifier les **choix impossibles** (Proposition 4.4) et les recommandations admissibles : les alternatives “recommandables” au regard de l'information préférentielle fournie par le décideur. Lorsque cet ensemble de recommandations admissibles sera réduit à une seule alternative, on aura alors trouvé la meilleure des alternatives selon les préférences du décideur.

Quant à l'ensemble $\mathbb{S}_{\mathbb{N}_{\mathbb{P}}}$ (Équation 4.3) des *preference swaps* déduits des déclarations de préférences \mathbb{P} du décideur, il se présente comme suit :

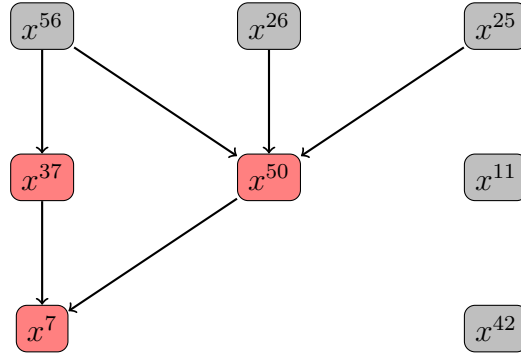


FIGURE 4.11 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}||}^A$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1).

$$\mathbb{S}_{\mathbb{N}_{\mathbb{P}||}} = \{ (c, a), (c, e), (c, f) \} \quad (4.13)$$

Sur la Figure 4.11, on conclut que les recommandations admissibles étant donné $\mathbb{P}||$ (Equation 4.11), sont les (5) suivantes :

$$\{x^{56}\}, \{x^{26}\}, \{x^{25}\}, \{x^{42}\}, \{x^{11}\}$$

La sous-section 4.4.2 simule une interaction à l'issue de laquelle le nombre de recommandations admissibles est réduit à 1. Au cours de cette interaction, l'information préférentielle supplémentaire est collectée à travers l'explication de conclusions robustes de type *choix impossible*.

4.4.2 Illustration de la fonction d'élicitation sur l'exemple introductif

Le déroulé complet de la présente illustration s'étend sur 4 étapes correspondant à différents instants t de l'interaction entre l'analyste et le décideur. À chaque instant, ont successivement lieu :

1. la production de l'explication qui sera soumise au décideur,
2. la collecte des réactions de ce dernier à l'explication proposée,
3. la mise à jour de la $\mathbb{P}||$ et l'observation de ses conséquences sur l'ensemble des recommandations admissibles.

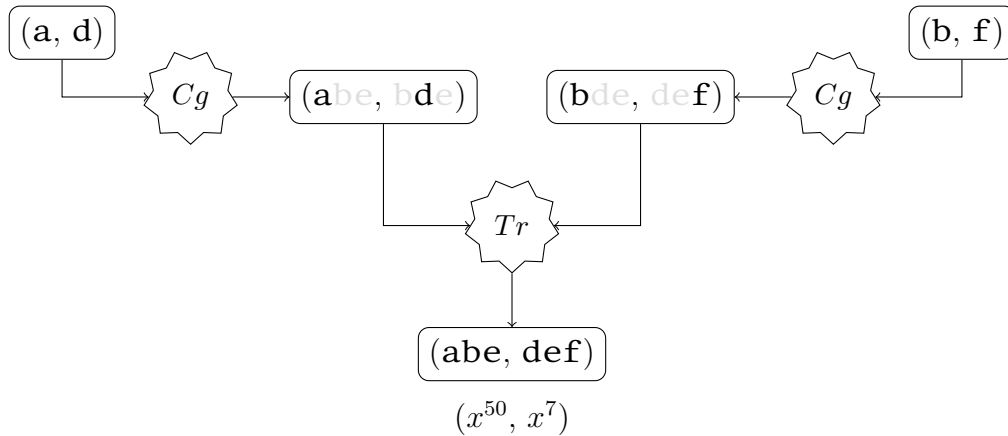
On considère que l'instant $t = 0$ est celui auquel l'analyste a recueilli l'information préférentielle $\mathbb{P}||$ de l'Équation 4.11. On a donc :

$$\mathbb{P}||^0 = \{ (r^{43}, r^{24}), (r^{24}, x^{50}), (x^{50}, x^7), (x^7, r^{17}) \}$$

et la Figure 4.11 correspond à la représentation graphique de l'ensemble $\mathbb{N}_{\mathbb{P}^0}^{\mathbf{A}}$ sans les arcs de réflexivité et de transitivité.

Étape $t = 1$: Explication de $(x^{50}, x^7) \in \mathbb{N}_{\mathbb{P}^0}^{\mathbf{A}}$

1. Détails de l'explication



2. Le décideur valide les deux *preference swaps* (a, d) et (b, f) .

3. On réalise donc la mise à jour suivante : $\mathbb{P}^1 = \mathbb{P}^0 \cup \{(a, d), (b, f)\}$ et la Figure 4.12 représente graphiquement l'ensemble $\mathbb{N}_{\mathbb{P}^1}^{\mathbf{A}}$:

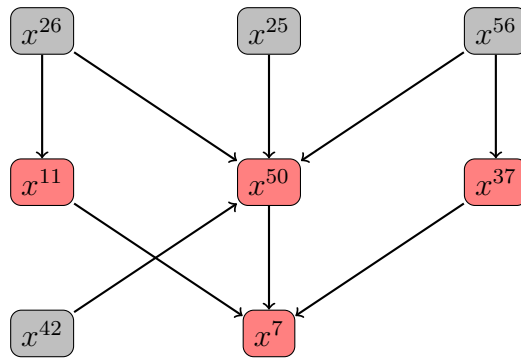


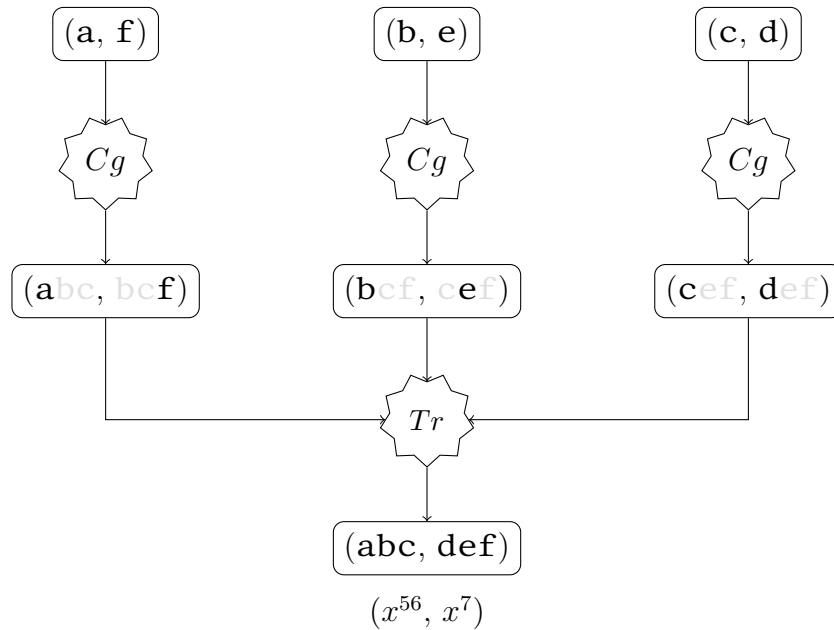
FIGURE 4.12 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}^1}^{\mathbf{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1).

Il s'ensuit que l'ensemble des (4) recommandations admissibles est désormais :

$$\{x^{56}\}, \{x^{26}\}, \{x^{25}\}, \{x^{42}\}$$

Étape $t = 2$: Explication de $(x^{56}, x^7) \in \mathbb{N}_{\mathbb{P}^1}^{\mathbf{A}}$

1. Détails de l'explication



2. Le décideur valide les trois *preference swaps* (a, f) et (b, e) et (c, d) .

3. On réalise la mise à jour suivante : $\mathbb{P}^2 = \mathbb{P}^1 \cup \{(a, f), (b, e), (c, d)\}$ et la Figure 4.13 représente graphiquement l'ensemble $\mathbb{N}_{\mathbb{P}^2}^{\mathbf{A}}$:

Il s'ensuit que l'ensemble des (2) recommandations admissibles est désormais :

$$\{x^{56}\}, \{x^{26}\}$$

Étape $t = 3$: Explication de $(x^{42}, x^7) \in \mathbb{N}_{\mathbb{P}^2}^{\mathbf{A}}$

1. Détails de l'explication

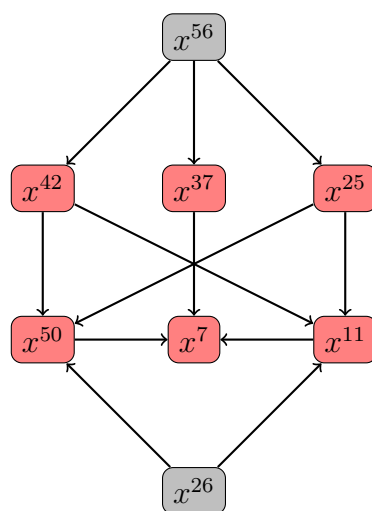
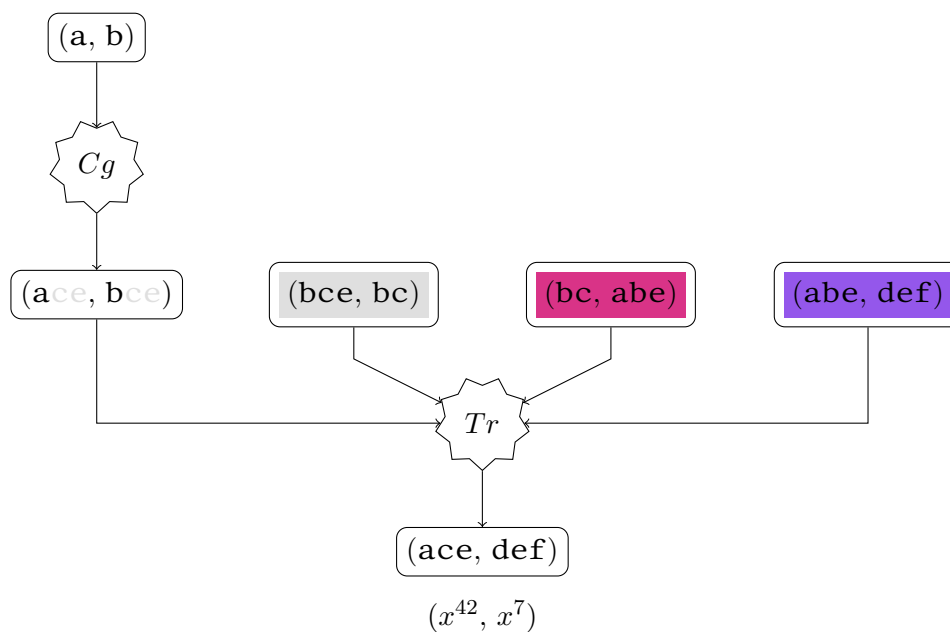


FIGURE 4.13 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}^2}^{\mathbf{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1).



2. Le décideur conteste le *preference swap* (a, b) .
3. On réalise la mise à jour suivante : $\mathbb{P}^3 = \mathbb{P}^2 \cup \{(b, a)\}$ et la Figure 4.14 représente graphiquement l'ensemble $\mathbb{N}_{\mathbb{P}^3}^{\mathbf{A}}$:
Il s'ensuit que l'ensemble des (2) recommandations admissibles reste :

$$\{x^{56}\}, \{x^{26}\}$$

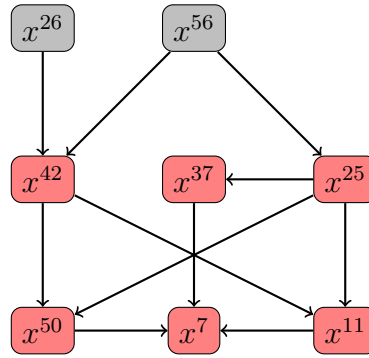
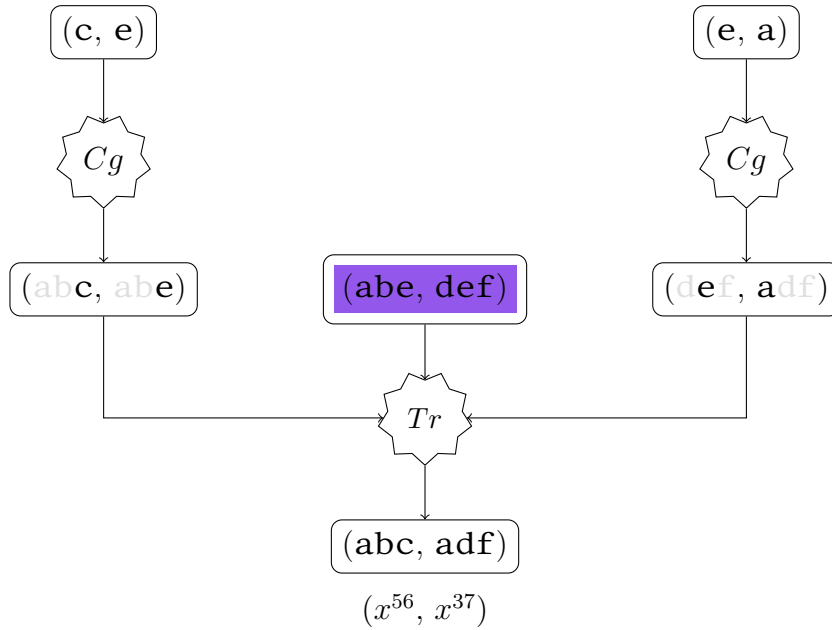


FIGURE 4.14 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}^3}^{\mathbf{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1).

Étape $t = 4$: Explication de $(x^{56}, x^{37}) \in \mathbb{N}_{\mathbb{P}^3}^{\mathbf{A}}$

1. Détails de l'explication



2. Le décideur valide le *preference swap* (c, e) mais conteste le *preference swap* (e, a) .

3. On réalise donc la mise à jour suivante : $\mathbb{P}^4 = \mathbb{P}^3 \cup \{(c, e), (a, e)\}$ et la Figure 4.15 représente graphiquement l'ensemble $\mathbb{N}_{\mathbb{P}^4}^{\mathbf{A}}$:

L'alternative x^{56} devient un **choix nécessaire** et donc la recommandation finale doit être : $\{x^{56}\}$.

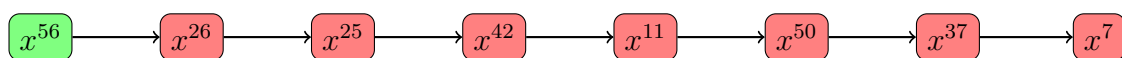


FIGURE 4.15 – Représentation graphique de la relation nécessaire $\mathbb{N}_{\mathbb{P}^4}^{\mathbb{A}}$ sans les arcs de réflexivité et de transitivité (Exemple 4.4.1).

En résumé, cet exemple nous a permis de voir que l'explication de déductions du calcul de la relation nécessaire peut être une occasion de collecte d'information préférentielle supplémentaire auprès du décideur. Ici, la collecte s'effectue par le truchement de *preference swaps* qui peuvent être validés ou contestés. L'évolution des ensembles $\mathbb{N}_{\mathbb{P}^t}^{\mathbb{A}}$ ($t \in \llbracket 1; 4 \rrbracket$) – Figures 4.11 à 4.15 – révèle que chacune des explications soumises au décideur contient au moins un *preference swap* non nécessaire. Comme nous l'avons fait remarquer en fin de sous-section 4.2.3, ces explications répondent à la question (du décideur) « Pourquoi **pourrais-je** préférer l'alternative x à l'alternative y ? » où (x, y) est la conclusion robuste expliquée et non à la question « Pourquoi **devrais-je** préférer l'alternative x à l'alternative y ? ». L'exemple traité ici nous permet (donc) également de constater que l'appartenance d'une comparaison par paire donnée à une relation de préférences (celle représentant les préférences du décideur) admet potentiellement plusieurs explications. C'est le cas par exemple, des comparaisons par paire déduites (x^{50}, x^7) et (x^{56}, x^7) des Étapes $t = 1, 2$ dont les *preference swaps* (non nécessaires) « explicatifs » sont tous validés par le décideur.

Dans la suite, nous esquisserons les grandes lignes d'un protocole interactif qui pourrait permettre à l'explication de conclusions robustes de remplir une fonction d'élicitation (sous-section 4.4.3) puis proposerons une formalisation et une modélisation du problème du calcul d'une telle explication (sous-section 4.4.4).

4.4.3 Ébauche d'un protocole interactif

Dans la sous-section précédente, nous avons montré que l'explication de conclusions robustes peut servir d'occasion de collecte d'information préférentielle additionnelle. Cette collecte s'effectue à partir des réactions du décideur aux explications qui lui sont proposées. Comme on peut aisément le remarquer, celles produites dans l'Exemple 4.4.1 sont conformes au schéma déductif *tr/cg* : elles prennent la forme d'une chaîne transitive d'alternatives reliant x à y où (x, y) est la déduction de la relation nécessaire (Définition 4.3) à expliquer. Cette chaîne transitive met en évidence des comparaisons par paire qui sont soit des déclarations de préférences du décideur soit congruentes à des *preference swaps* qui sont des comparaisons entre paire de critères. De telles explications ont donc de bonnes chances d'être, pour un décideur humain, intelligibles en raison de leur structure (transitive) et du caractère évident (éléments de la \mathbb{P}^4) ou élémentaire (*preference swaps*) des comparaisons par paire qui les composent. En supposant que le décideur ne remette jamais en cause ses propres déclarations de préférences, une fois confronté à une

explication qui en contient, les seules comparaisons par paire permettant la collecte d'information préférentielle supplémentaire sont donc les *preference swaps*. En les validant, les comparaisons par paire correspondantes sont rajoutées à la $\mathbb{P}\mathbb{I}$. En les contestant, on rajoute à la $\mathbb{P}\mathbb{I}$ les comparaisons par paire inverses. Pour que de telles mises à jour de la $\mathbb{P}\mathbb{I}$ puissent faire potentiellement aboutir le processus d'aide à la décision vers la détermination de la recommandation finale, il importe donc que chaque explication contienne au moins une comparaison par paire de type *preference swap* qui soit non nécessaire. Ce faisant, on pourra observer (comme dans l'exemple précédent) que le processus d'aide à la décision converge : le nombre d'éléments de la relation nécessaire sur \mathbf{A} ($|\mathbb{N}_{\mathbb{P}\mathbb{I}}^{\mathbf{A}}|$) croît tandis que le nombre de recommandations admissibles décroît. L'augmentation du nombre d'éléments de la relation nécessaire sur \mathbf{A} offre de nouvelles occasions d'explication qui permettront de capturer de l'information préférentielle additionnelle orchestrant ainsi une sorte d'« effet boule de neige ».

Dans la présente sous-section, nous nous proposons de tenter une formalisation de la simulation présentée précédemment. Le déroulé (en étapes) de l'Exemple 4.4.2 peut être assimilé à l'exécution du protocole interactif décrit ci-dessous :

Squelette d'un protocole interactif

Entrées. \mathbf{A} , \mathbf{A}^R et $\mathbb{P}\mathbb{I}^0$.

$t = 0$

Tant que $|\mathbb{N}_{\mathbb{P}\mathbb{I}^t}^{\mathbf{A}}| \neq 0$ **et** **ConditionArret** non vérifiée

Choisir $(x, y) \in \mathbb{N}_{\mathbb{P}\mathbb{I}^t}^{\mathbf{A}}$

 Tenter d'expliquer (x, y)

 Si l'explication existe :

 Collecter \mathbb{S}^t (Les réactions du décideur)

 Mises à jour :

 ◆ $\mathbb{P}\mathbb{I}^{t+1} = \mathbb{P}\mathbb{I}^t \cup \mathbb{S}^t$

 ◆ $t \leftarrow t + 1$

 Sinon retirer (x, y) de $\mathbb{N}_{\mathbb{P}\mathbb{I}^t}^{\mathbf{A}}$

Dans ce squelette, on voit bien (**Tant que** $|\mathbb{N}_{\mathbb{P}\mathbb{I}^t}^{\mathbf{A}}| \neq 0$) qu'un facteur limitant de la collecte d'information préférentielle additionnelle est l'absence de déductions de comparaisons par paire d'alternatives de l'ensemble \mathbf{A} . Lorsque cette condition est vérifiée, le choix (Procédure **Choisir**) de la paire (x, y) à expliquer peut être aléatoire, conforme à une stratégie bien définie ou encore explicitement désignée par le décideur. Quant à la condition d'arrêt **ConditionArret**, elle pourrait traduire une contrainte sur la valeur de t (avoir t faible pour « ne pas trop fatiguer » le décideur) ou une condition sur l'évolution du nombre de recommandations admissibles (seuil ou convergence).

Différentes instanciations de l'esquisse détaillée ci-dessus peuvent donc être envisagées. Celles-ci dépendront des conditions dans lesquelles se déroule l'interaction, des contraintes du décideur et des objectifs poursuivis par l'analyste. Enfin, il convient de

souligner qu'il peut exister pour une même déduction $(x, y) \in \mathbb{N}_{\mathbb{P}\mathbb{I}}$ plusieurs explications différentes et conformes au schéma déductif tr/cg . Lorsque c'est le cas, il appartiendra à l'analyste (suivant des critères objectifs ou non) de choisir quelle explication il souhaite présenter. Dans la sous-section qui suit, tenant compte des limites de la capacité (cognitive) d'un décideur humain à traiter plusieurs informations à la fois [Miller, 1956], nous avons fait le choix de retenir le critère de minimisation de la « longueur » de l'explication; ce qui nous a conduit à une formalisation du problème du calcul des explications, consistant à rechercher un plus court chemin dans un graphe que nous définirons.

4.4.4 Formalisation et modélisation

De façon informelle, les explications décrites précédemment sont des séquences transitives de comparaisons par paire qui sont :

- soit des éléments de $\mathbb{P}\mathbb{I}$,
- soit des éléments de \mathbb{D}_m , l'ensemble des comparaisons par paire exprimant une dominance de Pareto,
- soit des comparaisons par paire congruentes à des *preference swaps* dont au moins un qui n'appartient pas à $\mathbb{S}_{\mathbb{N}_{\mathbb{P}\mathbb{I}}}$ (Équation 4.3), l'ensemble des *preference swaps* nécessaires.

Du point de vue de leur structure et au regard de la Définition 4.7 (Explication questionnable), ces séquences peuvent être découpées en sous-séquences de deux types :

- des sous-séquences de longueur 1 qui sont des éléments de $\mathbb{P}\mathbb{I}$,
- des sous-séquences de longueur supérieure ou égale à 2 qui sont des explications questionnables [Amoussou et al., 2023a] de la comparaison par paire composée des deux alternatives les délimitant.

Exemple 4.1 L'explication de la comparaison par paire (x^{42}, x^7) à l'étape $t = 3$ de la sous-section 4.4.2 est la séquence :

$$\overbrace{(\text{ace}, \text{bce}), (\text{bce}, \text{bc})}^{\text{explication questionnable}}, \underbrace{(\text{bc}, \text{abe})}_{\in \mathbb{P}\mathbb{I}}, \underbrace{(\text{abe}, \text{def})}_{\in \mathbb{P}\mathbb{I}}$$

dont la sous-séquence composée des deux premières comparaisons par paire est une explication questionnable de (ace, bc) .

Dans la suite, nous désignerons sous le vocable d'*explications mixtes questionnables* de telles explications; le terme *mixtes* permettant à la fois de rendre compte de la diversité de leurs composants (comparaisons de la $\mathbb{P}\mathbb{I}$, de \mathbb{D}_m et congruentes à des *preference swaps*) et de l'utilisation d'au moins un *preference swap* non nécessaire. De façon formelle et à l'image de la Définition 4.7, nous considérerons :

Définition 4.8 (Explication mixte questionnable) *Il existe une explication mixte questionnable de (x, y) si et seulement si il existe une fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ et une séquence transitive de comparaisons par paire $[(u_1, v_1), \dots, (u_l, v_l)]$ de la relation \succsim_ω induite par ω telle que :*

- $u_1 \equiv x, v_l \equiv y$ et $u_k \equiv v_{k-1}$ pour tout $k \in \llbracket 2; l \rrbracket$,
- il existe $\tilde{k} \in \llbracket 1; l \rrbracket$ tel que $(u_{\tilde{k}}, v_{\tilde{k}}) \notin \mathbb{N}_{\mathbb{P}\mathbb{I}}$,
- les sous-séquences $[(u_{k_1}, v_{k_1}), \dots, (u_{k_2}, v_{k_2})], k_1 \leq k_2$ avec $k_1, k_2 \in \llbracket 1; l \rrbracket$ ne contenant aucune comparaison par paire de $\mathbb{P}\mathbb{I}$ sont des explications questionnables de la comparaison par paire (u_{k_1}, v_{k_2}) .

On dira que l'explication mixte questionnable est supportée par la fonction de score ω .

De la Définition 4.7, on rappelle qu'une condition nécessaire à l'existence d'une explication questionnable de la comparaison par paire (u, v) est l'existence d'une application injective ϕ de $\mathbf{con}(u, v)$ vers $\mathbf{pro}(u, v)$. Une telle application injective met donc en évidence exactement $|\mathbf{con}(u, v)|$ correspondances qui sont des *preference swaps*. Ce constat permet d'introduire la définition suivante :

Définition 4.9 (Longueur de (u, v) en termes de *preference swap(s)*) *Soit $(u, v) \in \mathcal{X}$. On définit comme suit la longueur $L(u, v)$ en termes de *preference swap(s)* de la comparaison par paire (u, v) :*

$$L(u, v) = \begin{cases} +\infty & \text{si } |\mathbf{pro}(u, v)| < |\mathbf{con}(u, v)| \\ |\mathbf{con}(u, v)| + I_{(u, v)} & \text{sinon} \end{cases}$$

avec $I_{(u, v)} = 1$ si $|\mathbf{pro}(u, v)| > |\mathbf{con}(u, v)|$ et $I_{(u, v)} = 0$ si $|\mathbf{pro}(u, v)| = |\mathbf{con}(u, v)|$.

Par ailleurs, la Proposition 4.6 rappelle que si une comparaison par paire (u, v) admet une explication questionnable (supportée par la fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$) alors (u, v) est explicable dans le contexte $(\succsim_\omega, \mathcal{S}_{\succsim_\omega}, tr/cg)$ lorsque $|\mathbf{pro}(x, y)| \geq 2$ et explicable dans le contexte $(\succsim_\omega, \mathcal{S}_{\succsim_\omega}, cg)$ lorsque $|\mathbf{pro}(x, y)| = 1$. Soit la définition suivante :

Définition 4.10 (Graphe d'explication mixte questionnable) *Soit un ensemble $\mathbb{P}\mathbb{I}$ de k comparaisons par paires ($\mathbb{P}\mathbb{I} = \{(u^1, v^1), \dots, (u^k, v^k)\}$) tel que $\Omega_{\mathbb{P}\mathbb{I}} \neq \emptyset$. Soient $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ une fonction de score et x, y deux alternatives.*

Le graphe d'explication mixte questionnable de (x, y) étant donné $\mathbb{P}\mathbb{I}$ et ω est le graphe orienté pondéré $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)} = (V, E)$ défini comme suit :

- $V = \{x, y\} \cup \{u^1, \dots, u^k\} \cup \{v^1, \dots, v^k\}$
Les sommets V de $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)}$ représentent l'ensemble des alternatives impliquées dans $\mathbb{P}\mathbb{I}$ complétées (éventuellement) de x et y .
- $E = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$ avec

- $E_1 = \{(x, u^t) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, sch) : \forall t \in \{1, \dots, k\}\}$
- $E_2 = \{(v^t, y) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, sch) : \forall t \in \{1, \dots, k\}\}$
- $E_3 = \{(v^t, u^{t'}) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, sch) : \forall t, t' \in \{1, \dots, k\} \text{ où } t \neq t' \text{ et } v^t \not\equiv u^{t'}\}$
- $E_4 = \{(u^1, v^1), \dots, (u^k, v^k)\} = \mathbb{P}\mathbb{I}$
- $E_5 = \{(x, y) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, sch)\}$

où $sch \in \{cg, tr/cg\}$.

Les arcs E_4 représentent exactement les comparaisons par paire de la $\mathbb{P}\mathbb{I}$.

Les arcs E_1 (respectivement E_2) symbolisent les liens directs entre le sommet-alternative x et les sommets u^t (respectivement v_t) des comparaisons par paire $\{(u^t, v^t)\}$ de E_4 ($1 \leq t \leq k$) lorsque ceux-ci correspondent à des comparaisons par paire explicables dans le contexte $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, tr/cg)$ ou $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, cg)$.

Quant aux arcs E_3 , ils représentent les connexions directes entre toute paire d'arcs distincts de E_4 lorsque celles-ci correspondent à des comparaisons par paire explicables dans le contexte $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, tr/cg)$ ou $(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, cg)$.

Enfin, l'ensemble E_5 contient uniquement l'arc direct (x, y) si et seulement si $(x, y) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, tr/cg)$ ou $(x, y) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, cg)$.

- La fonction g de pondération des arcs est définie comme suit :

$$g((u, v)) = \begin{cases} 1 & \text{si } (u, v) \in E_4 \text{ c'est-à-dire à } \mathbb{P}\mathbb{I} \\ L(u, v) & \text{sinon} \end{cases} \quad (4.14)$$

où $L(\cdot)$ est le longueur de \cdot en termes de préférence swaps (Définition 4.9).

Exemple 4.2 Le graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}^0, \omega}^{(x^{42}, x^7)}$ d'explication mixte questionnable de la comparaison par paire (x^{42}, x^7) étant donné l'ensemble $\mathbb{P}\mathbb{I}^0$ des déclarations de préférences du décideur (Équation 4.11) et de la fonction de score ω de l'Équation 4.15 est représentée ci-dessous :

$$\omega = (\mathbf{a} : 29, \mathbf{b} : 24, \mathbf{c} : 66, \mathbf{d} : 32, \mathbf{e} : 20, \mathbf{f} : 18) \quad (4.15)$$

Sur cette figure, les arcs de l'ensemble E_4 (la $\mathbb{P}\mathbb{I}$) sont en **noir**.

Comme la $\mathbb{P}\mathbb{I}$ est un rangement des alternatives de référence, l'ensemble E_3 est vide.

L'ensemble E_5 contient l'arc (**rouge**) direct (x^{42}, x^7) car comme on peut le voir sur la Figure 4.17, $(x^{42}, x^7) \in \mathcal{E}(\succsim_\omega, \mathbb{S}_{\succsim_\omega}, tr/cg)$

L'ensemble E_1 est composé des arcs en **bleu** (x^{42}, r^{24}) , (x^{42}, x^{50}) et de l'arc direct **rouge** direct (x^{42}, x^7) ; les explications questionnables correspondantes sont également fournies à la Figure 4.17.

L'unique arc de E_2 est (x^{50}, x^7) qui appartient aussi à E_4 (la $\mathbb{P}\mathbb{I}$). La comparaison par paire (r^{24}, x^7) – constituée à partir $(r^{43}, r^{24}) \in \mathbb{P}\mathbb{I}$ et de $(x^7, r^{17}) \in \mathbb{P}\mathbb{I}$ – n'appartient pas à E_2 car $\mathbf{pro}(r^{24}, x^7) = 2 < 3 = \mathbf{con}(r^{24}, x^7)$.

une explication mixte questionnable de (x, y) si et seulement si il existe une fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ telle que dans le graphe d'explication mixte questionnable $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)}$, il existe un chemin de x à y empruntant au moins un arc (u, v) n'appartenant pas à la relation nécessaire $\mathbb{N}_{\mathbb{P}\mathbb{I}}$.

Exemple 4.3 Sur la Figure 4.16, il existe 3 chemins du sommet-alternative x^{42} au sommet-alternative x^7 . Il s'agit de :

- $\mathcal{C}_1 = [x^{42}, r^{24}, x^{50}, x^7]$
- $\mathcal{C}_2 = [x^{42}, x^{50}, x^7]$
- $\mathcal{C}_3 = [x^{42}, x^7]$

A la fin de l'étape $t = 1$ de la sous-section 4.4.2, l'information préférentielle additionnelle capturée auprès du décideur à travers la validation de l'explication mixte questionnable qui lui a été soumise a permis de déduire que $(x^{42}, x^{50}) \in \mathbb{N}_{\mathbb{P}\mathbb{I}^1}^{\mathbf{A}}$ et $(x^{42}, x^7) \in \mathbb{N}_{\mathbb{P}\mathbb{I}^1}^{\mathbf{A}}$ (Voir Figure 4.12) : les chemins \mathcal{C}_2 et \mathcal{C}_3 ne correspondent dès lors plus à des explications mixtes questionnables contrairement au chemin \mathcal{C}_1 qui sera proposé en explication à l'étape $t = 3$.

Sur la base de la Proposition 4.7, nous pouvons comme suit, formaliser le problème de l'existence d'une explication mixte questionnable.

Formalisation du problème

Entrées :

- Un ensemble de comparaisons par paire $\mathbb{P}\mathbb{I}$
- Une comparaison par paire (x, y)
- Un entier naturel L_{max}

Question :

Existe-t-il une fonction de score $\omega \in \Omega_{\mathbb{P}\mathbb{I}}$ telle que dans le graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)}$, on ait un chemin $\mathcal{C} = [z^0, \dots, z^{t-1}, z^t, \dots, z^l]$ de longueur l tel que :

- $l \leq L_{max}$
- reliant x à y c'est-à-dire : $z^0 \equiv x$ et $z^l \equiv y$
- il existe $\tilde{k} \in \llbracket 1; l \rrbracket$ tel que $(z^{\tilde{k}-1}, z^{\tilde{k}}) \notin \mathbb{N}_{\mathbb{P}\mathbb{I}}$.

Comme évoqué précédemment, nous avons fait le choix de modéliser le problème de décision décrit ci-dessus dans sa version optimisation : celle conduisant à retenir l'explication la plus courte. Cette modélisation s'effectuera sous la forme d'un programme linéaire en nombres entiers⁹.

9. Il convient de signaler que le calcul des explications questionnables dans [Amoussou et al., 2023a] résulte de la résolution d'un programme linéaire en nombres entiers. La PLNE nous a donc semblé être un choix naturel de modélisation, d'autant plus qu'elle permet également la modélisation de problèmes de recherche de plus court chemin dans un graphe.

Modélisation à l'aide de la PLNE

La modélisation du problème du calcul de la plus courte explication mixte questionnable va s'appuyer sur l'ensemble noté $\bar{\mathcal{S}}_{N_{\mathbb{P}\mathbb{I}}}$ de *preference swaps* non nécessaires :

$$\bar{\mathcal{S}}_{N_{\mathbb{P}\mathbb{I}}} = \bigcup_{\omega \in \Omega_{\mathbb{P}\mathbb{I}}} \mathcal{S}_{\sim\omega} \setminus \mathcal{S}_{N_{\mathbb{P}\mathbb{I}}} \quad (4.16)$$

afin de garantir la collecte d'information préférentielle additionnelle.

Les variables On distinguera trois catégories de variables :

1. Variables continues w_i

Il s'agit des m variables w_i continues (strictement) positives dont les valeurs correspondent aux évaluations des niveaux "fort" des critères binaires. On rappelle (Définition 4.1) que nous avons fait l'hypothèse d'une évaluation nulle des niveaux "faible" des m critères. Les valeurs des variables w_i à l'optimum correspondront aux composantes de la fonction de score ω supportant l'explication mixte questionnable calculée.

2. Variables binaires $\text{arc}_{(u,v)}$

Pour chaque arc (u, v) du graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x,y)}$, on définit une variable binaire $\text{arc}_{(u,v)}$ valant 1 si et seulement si (u, v) appartient au chemin solution du problème.

3. Variables binaires $s_{(u,v)}^{(i,j)}$

Pour chaque arc $(u, v) \notin E_A$, on définit $|\text{pro}(u, v)| \times |\text{con}(u, v)|$ variables binaires qui permettront de déterminer (lorsqu'elle existe) l'explication questionnable de la comparaison par paire (u, v) utilisée dans l'explication de la déduction (x, y) . $s_{(u,v)}^{(i,j)}$ vaut 1 si et seulement si le *preference swap* (i, j) appartient à l'application injective caractérisant l'explication questionnable de (u, v) calculée.

La fonction objectif Étant donné la fonction de pondération des arcs (Équation 4.14), la fonction objectif s'écrit (linéairement) comme suit :

$$\sum_{(u,v) \in E} \text{arc}_{(u,v)} \times g((u, v))$$

où E représente l'ensemble des arcs du graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x,y)}$.

Les contraintes On distinguera six grandes catégories de contraintes :

1. Normalisation de la fonction de score et (stricte) positivité de ses composantes

$$\sum_{i \in [m]} w_i = 1 \quad (4.17)$$

$$w_i \geq \epsilon \quad \forall i \in [m] \quad (4.18)$$

avec ϵ un réel positif arbitrairement petit.

2. Prise en compte de la $\mathbb{P}\mathbb{I}$

$$\sum_{i \in \mathbf{pro}(u, v)} w_i \geq \sum_{j \in \mathbf{con}(u, v)} w_j \quad \forall (u, v) \in \mathbb{P}\mathbb{I} \quad (4.19)$$

3. Contraintes de « chemin »

$$\sum_{(x, u) \in E} \text{arc}_{(x, u)} = 1 \quad (4.20)$$

$$\sum_{(u, z) \in E} \text{arc}_{(u, z)} = \sum_{(z, v) \in E} \text{arc}_{(z, v)} \quad \forall z \in V \setminus \{x, y\} \quad (4.21)$$

avec E (resp. V) l'ensemble des arcs (resp. sommets) du graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)}$.

4. Contraintes liant les variables $s_{(u, v)}^{(i, j)}$ et $\text{arc}_{(u, v)}$

Pour chaque arc $(u, v) \notin \mathbb{P}\mathbb{I}$,

$$\sum_{i \in \mathbf{pro}(u, v)} s_{(u, v)}^{(i, j)} = \text{arc}_{(u, v)} \quad \forall j \in \mathbf{con}(u, v) \quad (4.22)$$

$$\sum_{i \in \mathbf{con}(u, v)} s_{(u, v)}^{(i, j)} \leq \text{arc}_{(u, v)} \quad \forall i \in \mathbf{pro}(u, v) \quad (4.23)$$

5. « Au moins un swap de $\bar{\mathbb{S}}_{\mathbb{N}\mathbb{P}\mathbb{I}}$ utilisé »

$$\sum_{(u, v) \in E \setminus E_4} \sum_{[i, j] \in \bar{\mathbb{S}}_{\mathbb{N}\mathbb{P}\mathbb{I}}} s_{(u, v)}^{(i, j)} \geq 1 \quad (4.24)$$

avec E l'ensemble des arcs du graphe $\mathcal{G}_{\mathbb{P}\mathbb{I}, \omega}^{(x, y)}$ et E_4 ceux représentant les éléments de la $\mathbb{P}\mathbb{I}$.

6. Contraintes liant les variables $s_{(u, v)}^{(i, j)}$ à w_i et w_j

$$w_i - w_j \geq (1 + \epsilon) \times s_{(u, v)}^{(i, j)} - 1 \quad (4.25)$$

avec ϵ un réel positif arbitrairement petit.

Récupération de la solution Le programme mathématique décrit ci-dessus peut admettre ou non une solution. Si son exécution échoue, cela signifie qu'il n'existe pas d'explication mixte questionnable de la déduction $(x, y) \in \mathbb{N}_{\mathbb{P}}$. Par contre, si elle réussit, la valeur optimale fournit la longueur de l'explication. Les valeurs des variables $\text{arc}_{(u,v)}$ indiquent les arcs empruntés et celles des variables $s_{(u,v)}^{(i,j)}$ lorsque $(u, v) \notin \mathbb{P}$ décrivent les explications questionnables correspondantes.

Une implémentation du programme linéaire en nombres entiers décrit ci-dessus est accessible ici¹⁰. Elle a été réalisée à l'aide du module Python `gurobipy` du solveur Gurobi. C'est cette implémentation qui a permis la construction de l'exemple détaillé dans les sous-sections 4.4.1 et 4.4.2.

4.5 Conclusion et perspectives

Dans les Chapitres 2 et 3, nous avons présenté une nouvelle approche de l'explication de comparaisons par paire d'alternatives décrites sur plusieurs critères et construite autour du concept de *contexte*. Dans ce chapitre, nous avons tenté de montrer que cette approche permet d'attribuer à l'explication, en aide multicritère à la décision, deux fonctions : celle de **justification** et celle d'**élicitation**. Le cadre de décision considéré est celui dans lequel le processus d'aide doit aboutir au choix d'un sous-ensemble d'alternatives parmi un ensemble fini (plus grand) d'alternatives. À cette fin, le décideur est invité à exprimer ses préférences sous la forme d'un ensemble \mathbb{P} de comparaisons par paire d'alternatives. Chaque alternative est décrite sur un ensemble fini de m critères. Lorsque l'on suppose que ses préférences sont représentables par un modèle additif, l'ensemble \mathbb{P} induit sur l'ensemble X des alternatives décrites sur m critères, une relation binaire réflexive et transitive appelée *relation nécessaire* [Greco et al., 2008]. Il s'agit des comparaisons par paire vérifiées par toutes des fonctions de score compatibles avec les déclarations de préférences \mathbb{P} du décideur. De cette relation, peuvent donc découler des conclusions robustes [Roy, 2009] susceptibles d'intéresser le décideur.

Nous nous sommes intéressés dans ce chapitre à l'explication de telles conclusions robustes tant dans une optique de justification que d'élicitation, étendant de ce fait les travaux de [Belahcène et al., 2017] et de [Amoussou et al., 2023a]. Cette extension a notamment consisté à intégrer aux explications produites les éléments de la \mathbb{P} ; ce qui, dans une perspective de justification, conduit logiquement à un accroissement du potentiel explicatif et dans une perspective d'élicitation, offre des chances de collecte plus importante d'information préférentielle additionnelle. Dans le premier cas, des expérimentations numériques ont été conduites et nous ont permis d'apprécier, en comparaison avec les explications à l'aide de *preference swaps* nécessaires de [Belahcène et al., 2017], la proportion supplémentaire de conclusions robustes que les types d'explication

10. <https://github.com/tetras92/SchemasDeductifsEnAMCD.git>

retournées par les algorithmes implémentés permettent d'expliquer. Dans le second cas, une ébauche de protocole interactif a été proposée qui permet la capture d'information préférentielle additionnelle par l'intermédiaire de *preference swaps* validés ou contestés par le décideur. Simulé sur un exemple fictif de situation d'aide à la décision, nous avons pu observer que ce protocole interactif peut aboutir à la détermination de la recommandation finale.

Dans tout ce chapitre, nous avons d'une part supposé que les déclarations du décideur sont fidèles à ses préférences réelles et d'autre part que celles-ci sont représentables par un modèle additif; ce qui dans la réalité peut ne pas s'avérer. Il convient dès lors, d'étudier les propriétés du protocole interactif proposé et de réfléchir à en anticiper voire en traiter les situations éventuelles de blocage dues à des réponses incohérentes fournies par le décideur. Pour ce faire, l'explication à des fins de justification pourraient servir. En effet, les situations d'inconsistance sont dues à la présence d'informations contradictoires dans la base des préférences exprimées par le décideur. Détecter la ou les paires d'alternatives sur lesquelles porte cette incohérence et expliquer les deux comparaisons par paire contradictoires ($x \succ y$ et $y \succ x$ par exemple) en mentionnant, comme nous le proposons ici, les déclarations de préférences du décideur, pourraient lui permettre d'identifier ses erreurs et de réviser conséquemment ses préférences. Pour y arriver, encore faut-il disposer d'une palette riche d'algorithmes permettant d'expliquer de façon intelligible toute déduction du calcul de la relation nécessaire. À défaut, on pourrait déjà commencer à mesurer expérimentalement la capacité des algorithmes proposés dans ce chapitre à contribuer (via l'explication) à la résolution de situations d'inconsistance ou à considérer comme nous le verrons dans le prochain chapitre d'autres "formes" d'explication qui diffèrent essentiellement de celles proposées ici, de par l'ensemble des comparaisons par paire considérées comme admises qui par définition, n'ont pas besoin d'être expliquées.

Chapitre 5

Explications par couverture de recommandations

Résumé

L'objet de ce chapitre est l'explication de la recommandation, aboutissement du processus d'aide à la décision. Plus précisément, nous nous sommes intéressé à la problématique du choix qui repose, à l'image du rangement, sur une évaluation relative des alternatives. Nous avons considéré qu'expliquer le choix d'une alternative revient à expliquer l'ensemble des comparaisons par paire sur laquelle elle repose. L'approche d'explication adoptée est celle que nous avons défendue tout au long de ce travail et qui est centrée autour de la définition d'un *contexte* : le schéma déductif utilisé est celui de couverture et les comparaisons par paire admises sont des éléments structurellement non décomposables de la relation binaire induite par une fonction de score représentant les préférences d'un décideur. Nous avons montré que le problème de décision associé au calcul d'une telle explication est NP-complet et en avons proposé une modélisation sous la forme d'un programme linéaire en nombres entiers.

Ayant assimilé l'explication de la recommandation à celle des comparaisons par paire confrontant la meilleure alternative à toutes les autres, nous concluons à son inexplicabilité dès que l'une de ces comparaisons par paire n'est pas explicable dans le contexte considéré. Dans ce cas, nous suggérons que soit tentée la mise en œuvre de mesures pouvant garantir la délivrance d'une explication potentiellement partielle – le cas de la mesure de *relaxation de la recommandation* – au décideur. Les deux autres mesures proposées sont le *recours à la transitivité* et la *translation de la fonction de score*. Pour chacune de ces “stratégies” de traitement de l'inexplicabilité, nous avons implémenté des algorithmes de mise en œuvre. Au travers d'expérimentations numériques, nous avons enfin “mesuré” à quel point chacune d'elles peut s'avérer intéressante en pratique.

Sommaire

| | | |
|------------|---|------------|
| 5.1 | Introduction | 138 |
| 5.2 | Description du cadre de décision et du contexte d'explication . . | 139 |
| 5.2.1 | Cadre de décision | 139 |
| 5.2.2 | Contexte d'explication | 141 |
| 5.3 | Explication du choix de la meilleure alternative à l'aide de la PLNE | 143 |
| 5.3.1 | NP-complétude du problème de l'explication par couverture . | 143 |
| 5.3.2 | Explication par couverture d'une comparaison par paire . . . | 145 |
| 5.3.3 | Explication par couverture d'une recommandation | 147 |
| 5.4 | Remédier à l'inexplicabilité de la recommandation | 151 |
| 5.4.1 | Recours à la transitivité | 153 |
| 5.4.2 | Relaxation de la recommandation : du 1-best au k -best . . . | 155 |
| 5.4.3 | Translation de la fonction de score : de ω à ω' | 157 |
| 5.5 | Expérimentations numériques | 160 |
| 5.5.1 | Description du protocole expérimental | 161 |
| 5.5.2 | Résultats et discussions | 162 |
| 5.6 | Conclusion et perspectives | 165 |

5.1 Introduction

Dans les Chapitres 2 et 3, nous avons présenté et décrit différents schémas déductifs suivant lesquels sont produites des explications de préférences additives. Dans le Chapitre 4, nous avons montré comment ces schémas peuvent servir en aide multicritère à la décision, à justifier des conclusions robustes d’une part et contribuer à l’élicitation des préférences du décideur d’autre part.

Dans ce chapitre, nous nous intéressons à l’explication de la recommandation finale :

- recommandation finale délivrée au décideur à l’issue du processus d’aide ayant abouti à la détermination d’un ensemble Ω de fonctions de score, soutenant toutes le choix d’une même alternative ou tout simplement
- recommandation admissible, soutenue par un sous-ensemble de Ω , et qui, est portée à la connaissance du décideur bien avant l’issue du processus d’aide qui, pour des raisons diverses, ne pourra pas être conduit à son terme.

Dans l’un ou l’autre des cas, l’explication produite portera sur un ensemble de comparaisons par paire compatibles avec une fonction de score précise ω . La relation de préférences décrivant le *contexte* d’explication est donc celle induite par ω . Quant au(x) schéma(s) déductif(s) employé(s), nous considérerons ici qu’il s’agit essentiellement du schéma de couverture (Définition 2.5) qui, parmi l’ensemble des schémas déductifs présentés, occupe une place centrale. En effet, il s’attaque de front à la principale cause de difficulté en décision multicritère (la multiplicité et la conflictualité des critères) en décomposant la comparaison par paire expliquée en sous-comparaisons atomiques telles que les *preference swaps* ([Belahcène et al., 2017]). Pour rappel, les *preference swaps* sont des comparaisons par paire d’alternatives composées chacune d’un unique critère. Il en existe un nombre polynomiale en m , le nombre total de critères; ce qui, notamment, rend “algorithmiquement facile” le problème du calcul d’une telle explication ([Belahcène et al., 2017]). Le recours aux *preference swaps* est principalement motivé par le fait qu’ils ont de bonnes chances d’être compris du décideur qui reçoit l’explication. Cependant, ils ont comme inconvénient (structurel) de ne pas permettre l’explication de certaines préférences : celles d’une alternative composée de strictement moins de critères sur une autre qui en contient strictement plus.

Les comparaisons par paire admises (complétant le *contexte* d’explication) que nous considérerons dans ce chapitre sont des comparaisons par paire “structurellement non décomposables” : elles sont constituées de deux alternatives dont au moins une qui est composée d’un unique critère. Il s’agit donc d’un ensemble incluant celui des *preference swaps*. Sa taille est exponentielle; ce qui rend “algorithmiquement plus difficile” le problème de calcul d’explication associé. Ce sont ces comparaisons par paire qui, combinées suivant le mécanisme de couverture, serviront à l’explication (par couverture) de la recommandation du choix d’une alternative parmi un ensemble fini A d’alternatives

décrits suivant m critères binaires. Nous avons opté pour le choix comme type de recommandation (plutôt que le tri) car il repose essentiellement sur une évaluation relative des alternatives – leur comparaison les unes avec les autres – que nos schémas déductifs tentent justement de rendre intelligible. Nous nous focalisons d'autre part sur la meilleure alternative non seulement parce qu'il s'agit de l'ensemble de choix le plus étudié en pratique mais également à des fins de généralisation : sélectionner k fois successivement la meilleure alternative parmi l'ensemble des alternatives non encore sélectionnées permet de construire l'ensemble des k meilleures alternatives et le faire $k = |\mathbf{A}| - 1$ fois permet de dégager un rangement sur l'ensemble \mathbf{A} .

Le chapitre est organisé en cinq sections. La Section 5.2 réprecise le cadre de décision et le *contexte* d'explication considérés. Dans la Section 5.3, après avoir montré que le problème du calcul de l'explication par couverture d'une comparaison par paire est NP-complet, nous en proposerons une modélisation sous forme de programme linéaire en nombres entiers (PLNE). Après quoi, nous expliciterons à travers différents algorithmes comment peut s'envisager le passage de l'explication d'une (unique) comparaison par paire à celle de la recommandation dans sa globalité. Quant à la Section 5.4, elle sera consacrée à la question du traitement de l'inexplicabilité dans le *contexte* d'explication considéré : nous proposerons trois mesures pouvant être mises en œuvre lorsqu'il n'existe pas d'explication par couverture de la recommandation. La Section 5.5 est celle des expérimentations numériques tandis que la Section 5.6 conclura le chapitre en évoquant des perspectives futures.

5.2 Description du cadre de décision et du contexte d'explication

L'objet de cette section est de spécifier le cadre de décision et de décrire le *contexte* d'explication que nous considérerons dans ce chapitre.

5.2.1 Cadre de décision

On s'intéressera dans ce chapitre à une recommandation de type choix ; choix de la meilleure alternative parmi un ensemble fini \mathbf{A} de n alternatives décrites sur m critères binaires c'est-à-dire ayant exactement deux niveaux d'évaluation : un niveau "fort" et un niveau "faible". L'ensemble des critères est noté $[m]$. Pour chaque critère $i \in [m]$, on admettra que le niveau "fort" symbolisé par 1 est strictement préféré au niveau "faible" symbolisé par 0 : $1 \succ_i 0 \forall i \in [m]$. Une alternative peut donc être représentée, soit, par un vecteur de m composantes – valuées à 1 ou 0 – soit sous une forme "ensembliste" : l'ensemble des critères sur lesquels elle vaut 1. On notera \mathbf{X} le produit cartésien des m

échelles binaires ordonnées $X_i = \{0, 1\}$:

$$\mathbf{X} = \prod_{i \in [m]} X_i \quad (5.1)$$

On a donc : $|\mathbf{X}| = 2^m$ et $\mathbf{A} \subset \mathbf{X}$.

Les préférences du décideur sont représentées par une fonction de score $\omega : \langle \omega_i \rangle_{i \in [m]}$ avec $\omega_i : X_i \rightarrow \mathbb{R}^+$ qui attribue à chaque critère une valeur réelle positive (ω_i) et qui assigne à chaque alternative x une valeur globale obtenue en sommant les scores marginaux sur l'ensemble des critères qui la composent ($\omega(x) = \sum_{i \in [m]} \omega_i(x_i)$).

La meilleure alternative de l'ensemble \mathbf{A} est l'alternative ayant le score global le plus élevé ou (en cas d'ex æquo) arbitrairement l'une de celles ayant le score le plus élevé. Elle sera notée x^* .

Exemple 5.1 Soit la table de performance ci-dessous qui décrit un ensemble \mathbf{A}_1 de quatre alternatives évaluées suivant un ensemble de sept critères binaires $[7] = \{a, b, c, d, e, f, g\}$. Les alternatives de l'ensemble $\mathbf{A}_1 = \{x^{35}, x^{28}, x^{72}, x^{21}\}$ sont numérotées à l'aide de l'en-

TABLE 5.1 – Exemple de table de performance

| | a | b | c | d | e | f | g |
|----------|---|---|---|---|---|---|---|
| x^{35} | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| x^{28} | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| x^{72} | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| x^{21} | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

tier correspondant à la conversion en base décimale de leur vecteur d'évaluation.

En supposant que les préférences du DM sont représentables par la fonction de score ω_1 suivante :

$$\omega_1 = (a : 128, b : 126, c : 77, d : 59, e : 52, f : 41, g : 37) \quad (5.2)$$

la Table 5.2 reproduit la Table 5.1 en représentant les alternatives de l'ensemble \mathbf{A}_1 sous leur forme ensembliste et fournit leur score global respectif. D'après les calculs réalisés, la meilleure alternative est donc $x^* = x^{35}$ dont le score $-\omega_1(x^*) = 204$ – relativement aux scores des autres alternatives de l'ensemble \mathbf{A}_1 est le plus élevé. Il s'agit de l'alternative à recommander.

Justifier le choix de x^* revient donc à expliquer l'ensemble des $n - 1$ comparaisons par paire (x^*, y) où $y \in \mathbf{A} \setminus \{x^*\}$. La description du contexte dans lequel sera produite chacune de ces explications est détaillée à la sous-section 5.2.2.

TABLE 5.2 – Vue ensembliste des alternatives de la Table 5.1 et leurs scores suivant ω

| | $\omega_1(\cdot)$ |
|------------------------------|-----------------------|
| $x^{35} \equiv \mathbf{bfg}$ | $126 + 41 + 37 = 204$ |
| $x^{28} \equiv \mathbf{cde}$ | $77 + 59 + 52 = 188$ |
| $x^{72} \equiv \mathbf{ad}$ | $128 + 59 = 187$ |
| $x^{21} \equiv \mathbf{ceg}$ | $77 + 52 + 37 = 166$ |

5.2.2 Contexte d'explication

La fonction de score ω représentant les préférences du décideur induit sur l'ensemble \mathbf{X} (Équation 5.1) une relation binaire réflexive et transitive notée \succsim_ω . Il s'agit de l'ensemble des comparaisons par paire (u, v) vérifiant $\omega(u) \geq \omega(v)$. L'ensemble des $n - 1$ comparaisons par paire (x^*, y) où $y \in \mathbf{A} \setminus \{x^*\}$ servant à justifier la recommandation x^* appartiennent par définition à \succsim_ω . Ce sont ces comparaisons par paire qui devront faire l'objet d'explication. Chacune de ces explications sera délivrée dans un *contexte*.

Pour rappel (cf. Chapitre 2), le *contexte* d'explication d'une comparaison par paire (x, y) est composé d'une relation binaire \mathbb{R} à laquelle appartient (x, y) , du sous-ensemble $\mathbb{B} \subset \mathbb{R}$ de comparaisons par paire admises et d'un schéma déductif *sch* suivant lequel la déduction de (x, y) à partir d'un ou plusieurs éléments de \mathbb{B} est explicitée. Ici, l'ensemble \mathbb{R} est (bien évidemment) la relation \succsim_ω . Comme indiqué précédemment, le schéma déductif que l'on considérera est celui de couverture *cov* (Définition 2.5). Quant au sous-ensemble \mathbb{B} de comparaisons par paire admises, il est composé des éléments de l'ensemble Δ_{\succsim_ω} défini comme suit :

Définition 5.1 (Ensemble de comparaisons par paire admises Δ_{\succsim_ω}) Soit ω la fonction de score représentant les préférences du décideur et \succsim_ω la relation binaire induite par ω sur le produit cartésien \mathbf{X} des m échelles binaires de critères. On définit comme suit l'ensemble Δ_{\succsim_ω} :

$$\Delta_{\succsim_\omega} = \{(u, v) \in \succsim_\omega \setminus \mathbb{D}_m : u \cap v = \emptyset \text{ et } |u| = 1 \text{ ou } |v| = 1\} \quad (5.3)$$

où \mathbb{D}_m est l'ensemble des comparaisons par paire exprimant une dominance de Pareto.

On a :

$$|\Delta_{\succsim_\omega}| = \sum_{k=0}^{m-1} (2^{m-1} - k) = \frac{m(2^m - (m-1))}{2}$$

L'ensemble Δ_{\succsim_ω} contient donc un nombre exponentiel de comparaisons par paire en fonction du nombre de critères m .

Dans la suite, on considérera que Δ_{\succsim_ω} est partitionné en deux ensembles notés $\Delta_{\succsim_\omega}^1$ et $\Delta_{\succsim_\omega}^2$ tels que :

$$\Delta_{\succsim_\omega}^1 = \{(u, v) \in \Delta_{\succsim_\omega} : |u| = 1\}$$

$$\Delta_{\succsim_\omega}^2 = \{(u, v) \in \Delta_{\succsim_\omega} : |u| > 1\}$$

Exemple 5.2 Soit ω_2 la fonction de score définie comme suit :

$$\omega_2 = (\mathbf{a} : 4, \mathbf{b} : 2, \mathbf{c} : 1)$$

L'ensemble des (9) comparaisons par paire $\Delta_{\succsim_{\omega_2}}$ est :

$$\{(\mathbf{a}, \mathbf{a}), (\mathbf{a}, \mathbf{b}), (\mathbf{a}, \mathbf{c}), (\mathbf{a}, \mathbf{bc}), (\mathbf{b}, \mathbf{b}), (\mathbf{b}, \mathbf{c}), (\mathbf{ac}, \mathbf{b}), (\mathbf{c}, \mathbf{c}), (\mathbf{ab}, \mathbf{c})\}$$

Du point de vue du mécanisme de couverture duquel dérive le schéma déductif *cov*, les éléments de l'ensemble Δ_{\succsim_ω} sont des comparaisons par paire (u, v) qui ne sont pas explicables suivant le schéma *cov* (c'est le cas de tous les éléments de $\Delta_{\succsim_\omega}^1 : (u, v) \notin \mathcal{E}(\succsim_\omega, \mathbb{B}, \text{cov})$ pour tout $(u, v) \in \Delta_{\succsim_\omega}^1$ et tout $\mathbb{B} \subset \succsim_\omega$) ou dont l'explication suivant le schéma *cov* mobilise essentiellement des comparaisons par paire de Δ_{\succsim_ω} et des comparaisons exprimant une dominance de Pareto (c'est le cas de certains éléments de $\Delta_{\succsim_\omega}^2 : (u, v) \in \mathcal{E}(\succsim_\omega, \mathbb{B}, \text{cov}) \implies \mathbb{B} \cap \Delta_{\succsim_\omega}^2 \neq \emptyset$ pour tout $(u, v) \in \Delta_{\succsim_\omega}^2$).

D'un point de vue purement cognitif, nous estimons que l'utilisation des éléments de l'ensemble Δ_{\succsim_ω} dans l'explication d'une comparaison par paire (x, y) n'entame pas l'intelligibilité de celle-ci. En effet, de par leur structure élémentaire, ces comparaisons par paire symbolisent une sorte de "troc" dans lequel exactement un critère est échangé contre un ou plusieurs autres critères. Ceci pourrait également s'assimiler à la comparaison de ce sous-ensemble de critères à un unique étalon de mesure. Il convient par ailleurs de remarquer que l'ensemble des *preference swaps* $\mathbb{S}_{\succsim_\omega}$ – servant de prémisses dans les explications de [Belahcène et al., 2017] – de la relation \succsim_ω est inclus dans Δ_{\succsim_ω} .

Dans la suite, nous considérerons :

Proposition 5.1 (Justification du choix de x^*) Soit x^* l'une des alternatives de l'ensemble \mathbf{A} dont le score global suivant ω , relativement à celui des autres alternatives, est le plus élevé.

Le choix de x^* est explicable par couverture si et seulement si, pour tout $y \in \mathbf{A} \setminus \{x^*\}$, on a :

- $(x^*, y) \in \Delta_{\succsim_\omega}$ ou
- $(x^*, y) \in \mathcal{E}(\succsim_\omega, \Delta_{\succsim_\omega}, \text{cov})$

Il se dégage de la Proposition 5.1 et de la Définition 2.5 du schéma de couverture que les explications que nous considérerons dans ce chapitre sont sous la forme d'un ensemble fini d'éléments de Δ_{\succsim_ω} "couvrant" la comparaison par paire expliquée. Plus

précisément, si (x, y) est la comparaison par paire à expliquer, l'explication de (x, y) est un ensemble $\{(u_k, v_k)\}_{k \geq 1}$ d'éléments de $\Delta_{\succsim \omega}$ et de \mathbb{D}_m tel que :

$$\bigcap_{k \geq 1} u_k = \bigcap_{k \geq 1} v_k = \emptyset \quad (5.4)$$

$$\bigcup_{k \geq 1} u_k = x \quad (5.5)$$

$$\bigcup_{k \geq 1} v_k = y \quad (5.6)$$

Dans la prochaine section, nous montrerons comment à l'aide de la programmation linéaire en nombres entiers, on peut calculer l'explication par couverture d'une unique comparaison par paire (x, y) puis d'un ensemble de comparaisons par paire, en l'occurrence celles permettant de justifier le choix d'une meilleure alternative x^* de l'ensemble A .

5.3 Explication du choix de la meilleure alternative à l'aide de la PLNE

Cette section est consacrée à la description d'algorithmes réalisant d'une part le calcul de l'explication par couverture d'une comparaison par paire (x, y) étant donné une fonction de score ω représentant les préférences du décideur et d'autre part celui de l'explication d'une recommandation de type *choix de la meilleure alternative*. Comme indiqué précédemment, ces algorithmes reposent sur la résolution de programmes linéaires en nombres entiers (PLNE) que nous allons ici décrire (voir sous-sections 5.3.2 et 5.3.3).

Mais avant, nous montrerons (sous-section 5.3.1) que le problème de décision associé au calcul d'une explication par couverture est un problème (algorithmiquement) difficile.

5.3.1 NP-complétude du problème de l'explication par couverture

Soit le problème de décision *CoveringExplanation* défini comme suit :

Entrées :

- Une fonction de score $\omega : \langle \omega_i \rangle_{i \in [m]}$ avec $\omega_i : X_i \rightarrow \mathbb{R}^+$
- Deux alternatives x et y décrites suivant m critères binaires : $x \subset [m]$ et $y \subset [m]$.

Question : La comparaison par paire (x, y) admet-elle une explication par couverture étant donné ω ?

Autrement dit, existe-t-il un ensemble fini d'éléments de $\Delta_{\succsim \omega}$ et de \mathbb{D}_m vérifiant les Équations 5.4, 5.5 et 5.6 ?

Dans cette sous-section, on va montrer que le problème *CoveringExplanation* est NP-complet.

On considérera donc les deux propositions suivantes :

Proposition 5.2 *Le problème *CoveringExplanation* est dans NP.*

Démonstration 5.1 *Soit un ensemble $\mathbb{E} = \{(u_k, v_k)\}_{k \geq 1}$ d'éléments de $\Delta_{\sim \omega}$ et de \mathbb{D}_m "couvrant" la comparaison par paire (x, y) . On peut vérifier en temps polynomial que \mathbb{E} est une instance positive du problème *CoveringExplanation*. Pour ce faire, il suffit, en parcourant l'ensemble \mathbb{E} :*

- de s'assurer que chaque critère i (respectivement j) de x (respectivement y) apparaît dans une unique comparaison par paire (u_k, v_k) (en $\mathcal{O}(m^2)$)
- de vérifier que $|u_k| = 1$ ou $|v_k| = 1$ (en $\mathcal{O}(m^2)$) et
- de vérifier que $\omega(u_k) \geq \omega(v_k)$ (en $\mathcal{O}(m^2)$).

Proposition 5.3 *Le problème *CoveringExplanation* est NP-difficile.*

Démonstration 5.2 *On démontre la Proposition 5.3 en recourant à une réduction polynomiale depuis le problème NP-complet *Partition* qui se décrit comme suit :*

Entrée :

- Une ensemble fini E d'entiers positifs e_i ($i \in \llbracket 1; n \rrbracket$).

Question : *Existe-t-il une partition de l'ensemble E en deux sous-ensembles E_1 et E_2 telle que $\sum_{i \in E_1} e_i = \sum_{j \in E_2} e_j$?*

On transforme comme suit toute instance $\mathcal{I}_{Partition}$ en une instance $\mathcal{I}_{CoveringExplanation}$:

- La fonction de score ω est définie sur $n + 2$ composantes et est telle que : $\omega_i = e_i$ pour tout $i \in \llbracket 1; n \rrbracket$ et $\omega_{n+1} = \omega_{n+2} = \frac{\sum_{i \in \llbracket 1; n \rrbracket} e_i}{2}$
- Les alternatives y et x sont respectivement composées des critères d'indices $n + 1$ et $n + 2$ ($y = \{n + 1, n + 2\}$) et des critères d'indices 1 à n ($x = \{1, \dots, n\}$)

Cette transformation est polynomiale en la taille de $\mathcal{I}_{Partition}$.

*On montre ensuite que $\mathcal{I}_{Partition}$ est une instance positive du problème *Partition* si et seulement si $\mathcal{I}_{CoveringExplanation}$ est une instance positive du problème d'explication par couverture.*

\Rightarrow *Supposons que $\mathcal{I}_{Partition}$ est une instance positive du problème *Partition* et montrons que $\mathcal{I}_{CoveringExplanation}$ est une instance positive du problème d'explication par couverture.*

$\mathcal{I}_{Partition}$ admet une solution (E_1, E_2) telle que $E_1 \cup E_2 = \llbracket 1; n \rrbracket$, $E_1 \cap E_2 = \emptyset$ et $\sum_{i \in E_1} e_i = \sum_{j \in E_2} e_j$. En prenant $(u_1, v_1) = (E_1, \{n+1\})$ et $(u_2, v_2) = (E_2, \{n+2\})$, on construit une instance positive de $\mathcal{I}_{CoveringExplanation}$ car vérifiant l'ensemble des Équations 5.4, 5.5 et 5.6. De plus, $\omega(u_1) = \omega(v_1) = \omega(u_2) = \omega(v_2)$.

⇐ Supposons que $\mathcal{I}_{\text{CoveringExplanation}}$ est une instance positive du problème d'explication par couverture et montrons que $\mathcal{I}_{\text{Partition}}$ est une instance positive du problème Partition.

Soit $\mathbb{E} = \{(u_k, v_k)\}_{k \geq 1}$ l'explication correspondante. Comme :

$$|y| = |\{n+1, n+2\}| = 2 = \left| \bigcup_{k \geq 1} v_k \right|$$

et que

$$\omega_{n+1} = \omega_{n+2} = \frac{\sum_{i \in \llbracket 1; n \rrbracket} e_i}{2} \quad \text{et} \quad \bigcap_{k \geq 1} v_k = \emptyset$$

alors $k = 2$ et sans perte de généralité prenons $v_1 = \{n+1\}$ et $v_2 = \{n+2\}$.

Par définition, $\omega(u_1) \geq \omega(v_1) = \frac{\sum_{i \in \llbracket 1; n \rrbracket} e_i}{2}$ et $\omega(u_2) \geq \omega(v_2) = \frac{\sum_{i \in \llbracket 1; n \rrbracket} e_i}{2}$. On a donc : $\omega(u_1) = \omega(v_1) = \omega(u_2) = \omega(v_2)$ car $u_1 \cap u_2 = \emptyset$ (Équation 5.4).

En prenant E_1 (respectivement E_2) l'ensemble des entiers correspondant aux critères de u_1 (respectivement u_2), on a bien $E_1 \cup E_2 = \llbracket 1; n \rrbracket$, $E_1 \cap E_2 = \emptyset$ et $\sum_{i \in E_1} e_i = \sum_{j \in E_2} e_j$.

Il résulte des Propositions 5.2 et 5.3 que le problème *CoveringExplanation* est NP-complet.

Signalons pour finir que le problème *CoveringExplanation* restreint aux *preference swaps* $\mathbb{S}_{\sim \omega}$ ($\subset \Delta_{\sim \omega}$) et aux comparaisons par paire de \mathbb{D}_m appartient à la classe P (cf. corollaires du [Belahcène et al., 2017, Théorème 6]).

5.3.2 Explication par couverture d'une comparaison par paire

Dans cette sous-section, nous décrivons un programme linéaire en nombres entiers dont la résolution fournit, lorsqu'elle existe, une explication par couverture de la comparaison par paire (x, y) étant donné une fonction de score ω représentant les préférences du décideur.

Pour rappel, une telle explication est un ensemble $\mathbb{E} = \{(u_k, v_k)\}_{k \geq 1} \subset \Delta_{\sim \omega}$ vérifiant les Équations 5.4, 5.5 et 5.6. Nous rappelons également que $\Delta_{\sim \omega}$ est partitionné en deux sous-ensembles $\Delta_{\sim \omega}^1$ et $\Delta_{\sim \omega}^2$ (voir Définition 5.1). Dans la suite, nous désignerons par \mathbb{E}^1 (respectivement \mathbb{E}^2) les ensembles $\mathbb{E} \cap \Delta_{\sim \omega}^1$ (respectivement $\mathbb{E} \cap \Delta_{\sim \omega}^2$) où \mathbb{E} est une explication par couverture de (x, y) . Les comparaisons par paire de l'ensemble \mathbb{E}^1 sont de la forme $(\{i\}, J)$ où $i \in x$ et $J \subseteq y$ avec $|J| \geq 1$ tandis que celles de \mathbb{E}^2 sont de la forme $(I, \{j\})$ où $j \in y$ et $I \subseteq x$ avec $|I| \geq 2$.

Soit le programme linéaire en nombres entiers $\mathcal{P}(x, y, \omega)$ décrit comme suit :

Les variables de décision : Deux types de variables binaires sont à distinguer :

1. Variables binaires b_{ij}^1 et b_{ij}^2

Pour tout $i \in x$ et tout $j \in y$, b_{ij}^1 (respectivement b_{ij}^2) vaut 1 si et seulement si il existe dans \mathbb{E}^1 (respectivement \mathbb{E}^2) une comparaison par paire de la forme $(\{i\}, J)$ avec $j \in J$ (respectivement de la forme $(I, \{j\})$ avec $i \in I$).

Les contraintes : Les quatre groupes de contraintes ci-dessous permettent de garantir la vérification des Équations 5.4, 5.5 et 5.6 et d'assurer que les comparaisons par paire de \mathbb{E}^1 et \mathbb{E}^2 appartiennent bien à l'ensemble $\Delta_{\sim\omega}$.

$$b_{ij}^1 + \sum_{j' \in y} b_{ij'}^2 \leq 1 \quad \forall i \in x \quad \forall j \in y \quad (5.7)$$

$$b_{ij}^2 + \sum_{i' \in x} b_{i'j}^1 \leq 1 \quad \forall i \in x \quad \forall j \in y \quad (5.8)$$

$$\sum_{j \in y} b_{ij}^1 \times \omega_j \leq \omega_i \quad \forall i \in x \quad (5.9)$$

$$\sum_{i \in x} (b_{ij}^1 + b_{ij}^2) \times \omega_i \geq \omega_j \quad \forall j \in y \quad (5.10)$$

Le programme $\mathcal{P}(x, y, \omega)$, pour lequel il n'est pas nécessaire de définir une fonction objectif, peut ou peut ne pas admettre de solution. Lorsque sa résolution (à l'aide d'un solveur par exemple) réussit, les valeurs des variables b_{ij}^1 et b_{ij}^2 permettent de reconstruire les comparaisons par paire de l'ensemble \mathbb{E} expliquant (x, y) .

Si l'on souhaite minimiser la "longueur" de l'explication c'est-à-dire la taille de l'ensemble \mathbb{E} , on pourrait rajouter au programme $\mathcal{P}(x, y, \omega)$ la fonction objectif (5.11) et les contraintes (5.12) et (5.13) suivantes :

$$\min z = \sum_{i \in x} t_i^1 + \sum_{j \in y} t_j^2 \quad (5.11)$$

$$\sum_{j \in y} b_{ij}^1 \leq M \times t_i^1 \quad \forall i \in x \quad (5.12)$$

$$\sum_{i \in x} b_{ij}^2 \leq M \times t_j^2 \quad \forall j \in y \quad (5.13)$$

où les variables binaires t_i^1 et t_j^2 vont respectivement servir à comptabiliser le nombre de comparaisons par paire des ensembles \mathbb{E}^1 et \mathbb{E}^2 utilisées dans l'explication de (x, y) et où M est une constante arbitrairement grande ($M \geq \max(|x|, |y|)$).

5.3.3 Explication par couverture d'une recommandation

Disposant d'une méthode de calcul de l'explication d'une comparaison par paire (x, y) étant donné une fonction de score ω (voir sous-section 5.3.2), on peut à présent calculer celle d'une recommandation. La Proposition 5.1 nous rappelle que la justification du choix de la meilleure alternative x^* revient à expliquer les $|\mathbf{A}| - 1$ comparaisons par paire (x^*, y) où $y \in \mathbf{A} \setminus \{x^*\}$.

Dans cette sous-section, nous allons proposer deux moyens de déterminer l'explication complète d'une recommandation.

L'Algorithme 3 détaille l'un de ces moyens. Il exécute de façon indépendante et au plus $n - 1$ fois le PLNE $\mathcal{P}(x, y, \omega)$ retournant l'explication de la comparaison par paire (x, y) lorsque celle-ci existe (Ligne 5).

Cet algorithme retourne une table de correspondance qui associe à chaque comparaison par paire (x^*, y) son explication par couverture – sous la forme d'un ensemble \mathbb{E} d'éléments de $\Delta_{\succsim \omega}$ – à condition que les $n - 1$ comparaisons par paire (x^*, y) soient toutes explicables.

Algorithme 3 Explication par couverture d'une recommandation

Entrée(s) \mathbf{A} et ω

Sortie(s) un booléen `explainable` indiquant l'existence d'une explication de la recommandation et une table TX contenant les explications des comparaisons par paire (x^*, y) sur laquelle elle repose.

- 1: Trier par ordre décroissant de score (suivant ω) les alternatives de l'ensemble \mathbf{A} : $x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(n)}$
 - 2: Initialiser TX
 - 3: $i \leftarrow 2$
 - 4: **tant que** $(i \leq n)$ **faire**
 - 5: $\mathbb{E}_i \leftarrow \mathcal{P}(x^{(1)}, x^{(i)}, \omega)$
 - 6: **si** \mathbb{E}_i vide (le PLNE n'a pas de solution) **alors**
 - 7: **retourner** False, []
 - 8: **sinon**
 - 9: TX[i] $\leftarrow \mathbb{E}_i$
 - 10: **fin du si**
 - 11: $i \leftarrow i + 1$
 - 12: **fin du tant que**
 - 13: **retourner** True, TX
-

Cette approche du calcul de l'explication de la recommandation, bien que valide, a l'inconvénient de ne pas la considérer comme un tout destiné à un même décideur. Comme une même comparaison par paire (x, y) peut admettre plusieurs explications différentes, cette approche fait abstraction des similitudes qui pourraient exister entre

les explications par couverture des différentes paires (x^*, y) . La prise en compte de ces similitudes peut limiter l'effort cognitif requis pour le traitement intégral de l'explication de la recommandation comme on peut le remarquer sur les Figures 5.1 et 5.2 de l'Exemple 5.3.

Exemple 5.3 Soit la table de performance ci-dessous qui décrit un ensemble A_2 de trois alternatives et fournit leurs scores suivant la fonction ω_1 de l'Équation 5.2 :

$$\omega_1 = (a : 128, b : 126, c : 77, d : 59, e : 52, f : 41, g : 37)$$

TABLE 5.3 – Table de performance A_2 et scores suivant ω_1 (Équation 5.2)

| | A_2 | $\omega_1(\cdot)$ |
|---|----------------------|------------------------|
| * | $x^{100} \equiv abe$ | $128 + 126 + 52 = 306$ |
| | $x^{26} \equiv cdf$ | $77 + 59 + 41 = 177$ |
| | $x^{25} \equiv cdg$ | $77 + 59 + 37 = 173$ |

L'alternative à recommander est x^{100} .

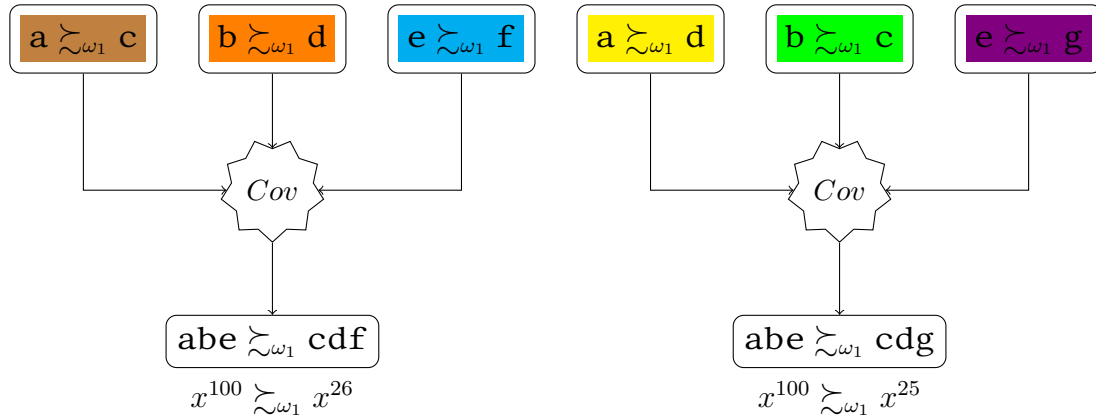
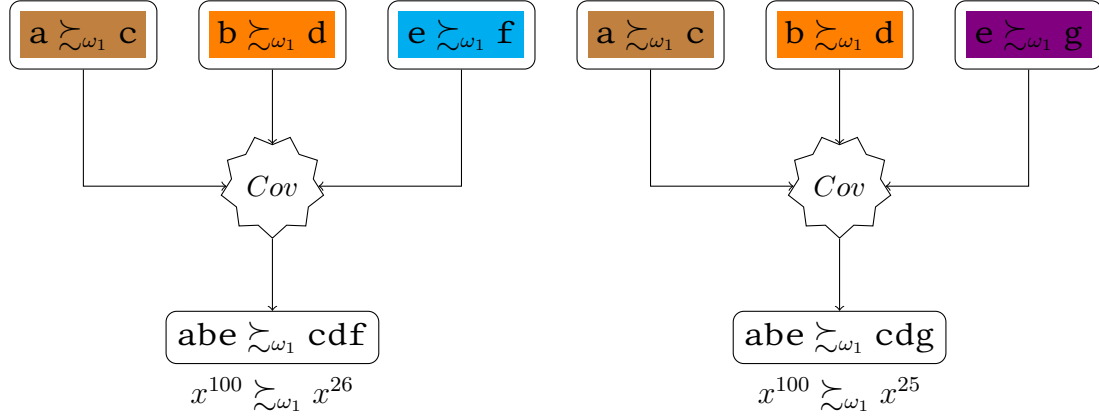


FIGURE 5.1 – Exemple 1 d'explication de la recommandation x^{100}

Les Figures 5.1 et 5.2 sont deux sorties différentes de l'exécution de l'Algorithme 3 sur l'instance de l'exemple.

On remarque (Figure 5.1) que l'explication de la recommandation x^{100} utilise en tout six éléments différents de $\Delta_{\sim_{\omega_1}}$. Ce sont : (a, c), (a, d), (b, d), (b, c), (e, f) et (e, g). Quant à l'explication de la Figure 5.2, elle factorise les "arguments" (a, c) et (b, d) que l'on retrouve à la fois dans l'explication de (x^{100}, x^{26}) et de (x^{100}, x^{25}) , portant ainsi le nombre total d'éléments différents de $\Delta_{\sim_{\omega_1}}$ utilisés à quatre.

En toute vraisemblance, un décideur pourrait préférer cette deuxième explication à la première parce qu'étant globalement « plus facile à traiter cognitivement ».


 FIGURE 5.2 – Exemple 2 d’explication de la recommandation x^{100}

Garantir la production d’une explication de recommandation avec un nombre minimal d’arguments différents (ou un nombre maximal d’arguments identiques) est inenvisageable lorsque le calcul s’effectue à partir de l’Algorithme 3. C’est la raison pour laquelle nous proposons ici un second moyen d’effectuer ce calcul. Celui-ci va s’appuyer sur la résolution d’un PLNE unique dont la fonction objectif permettra de limiter le nombre de comparaisons par paire explicatives différentes mobilisées.

Soit le programme linéaire en nombres entiers $\mathcal{P}_r(\mathbf{A}, \omega)$ décrit comme suit :

Les variables de décision : On distinguera deux grandes catégories de variables binaires

1. Variables binaires $b_{ij}^{1,y}$ et $b_{ij}^{2,y}$ qui à l’image des variables b_{ij}^1 et b_{ij}^2 de la sous-section 5.3.2 vont permettre de reconstruire l’explication de chaque comparaison par paire (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$.
2. Variables binaires β_{ij} : pour tout $i \in x$ et tout $j \in y$ ($y \in \mathbf{A} \setminus \{x^*\}$), la variable β_{ij} vaut 1 si et seulement si il existe au moins une alternative $y \in \mathbf{A} \setminus \{x^*\}$ telle que la $b_{ij}^{1,y}$ ou $b_{ij}^{2,y}$ vaut 1.

Les contraintes : On considérera deux catégories de contraintes :

1. La première catégorie de contraintes reprend les contraintes du programme $\mathcal{P}(x, y, \omega)$ de la sous-section 5.3.2, appliquées à chacune des comparaisons par paire (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$:

$$b_{ij}^{1,y} + \sum_{j' \in y} b_{ij'}^{2,y} \leq 1 \quad \forall i \in x^* \quad \forall j \in y \quad (5.14)$$

$$b_{ij}^{2,y} + \sum_{i' \in x^*} b_{i'j}^{1,y} \leq 1 \quad \forall i \in x^* \quad \forall j \in y \quad (5.15)$$

$$\sum_{j \in y} b_{ij}^{1,y} \times \omega_j \leq \omega_i \quad \forall i \in x^* \quad (5.16)$$

$$\sum_{i \in x^*} (b_{ij}^{1,y} + b_{ij}^{2,y}) \times \omega_i \geq \omega_j \quad \forall j \in y \quad (5.17)$$

2. Contraintes reliant les deux catégories de variables :

$$\sum_{y \in \mathbf{A} \setminus \{x^*\}} b_{ij}^{1,y} \mathbb{1}_{\{i\}, \{j\}}^y \leq (|\mathbf{A}| - 1) \times \beta_{ij} \quad \forall i \in [m] \quad \forall j \in [m] \quad (5.18)$$

$$\sum_{y \in \mathbf{A} \setminus \{x^*\}} b_{ij}^{2,y} \mathbb{1}_{\{i\}, \{j\}}^y \leq (|\mathbf{A}| - 1) \times \beta_{ij} \quad \forall i \in [m] \quad \forall j \in [m] \quad (5.19)$$

avec $\mathbb{1}_{\{i\}, \{j\}}^y$ valant 1 si et seulement si $i \in x^*$ et $j \in y$.

Ici aussi, la constante $(|\mathbf{A}| - 1)$ pourrait être remplacée par une quelconque valeur entière M qui lui est supérieure.

La fonction objectif :

$$\sum_{i \in [m]} \sum_{j \in [m]} \beta_{ij} \quad (\text{à minimiser}) \quad (5.20)$$

Lorsque l'exécution du programme $\mathcal{P}_r(\mathbf{A}, \omega)$ décrit ci-dessus réussit, les valeurs des variables $b_{ij}^{1,y}$ et $b_{ij}^{2,y}$ vont permettre de reconstruire l'explication de chaque comparaison par paire (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$.

Intuitivement, la fonction objectif (5.20) qui minimise le nombre d'associations binaires distinctes de critères au sein des sous-comparaisons explicatives, aura tendance à minimiser aussi le nombre de sous-comparaisons explicatives différentes utilisées. C'est ce que l'on observera chaque fois que les éléments de $\Delta_{\sim\omega}$ utilisés dans l'explication de la recommandation seront essentiellement des *preference swaps* et/ou des éléments de l'ensemble \mathbb{D}_m (comme c'est le cas de l'Exemple 5.3).

En résumé, dans cette sous-section, nous avons proposé deux algorithmes de calcul de l'explication par couverture d'une recommandation : l'un itératif, l'autre s'appuyant sur la résolution d'un programme linéaire en nombres entiers unique. Pour rappel, l'explication de la recommandation x^* revient à expliquer chacune des comparaisons par paire (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$ (Proposition 5.1). La recommandation x^* est donc inexplicable dès que la décomposition d'au moins une de ces comparaisons par paire est irréalisable à partir des comparaisons par paire "atomiques" de l'ensemble $\Delta_{\sim\omega}$; ceci se traduisant

par le fait que le PLNE $\mathcal{P}_r(\mathbf{A}, \omega)$ n'admette pas de solution ou (de façon équivalente) que l'Algorithme 3 retourne `False`. L'objet des deux prochaines sections est de proposer des moyens de remédier aux situations d'inexplicabilité de la recommandation et de mesurer expérimentalement leur intérêt.

5.4 Remédier à l'inexplicabilité de la recommandation

Considérons les deux exemples suivants :

Exemple 5.4 (Suite de l'Exemple 5.1) *On souhaite expliquer à un décideur dont les préférences sont représentées par la fonction de score ω_1 de l'Équation 5.2 (rappelée ci-dessous) pourquoi l'alternative $x^{35} \equiv \text{bfg}$ est préférable aux alternatives $x^{28} \equiv \text{cde}$, $x^{72} \equiv \text{ad}$ et $x^{21} \equiv \text{ceg}$.*

$$\omega_1 = (a : 128, b : 126, c : 77, d : 59, e : 52, f : 41, g : 37)$$

La Figure 5.3 détaille les explications par couverture des comparaisons par paire $x^{35} \succ_{\omega_1} x^{28}$ et $x^{35} \succ_{\omega_1} x^{21}$.

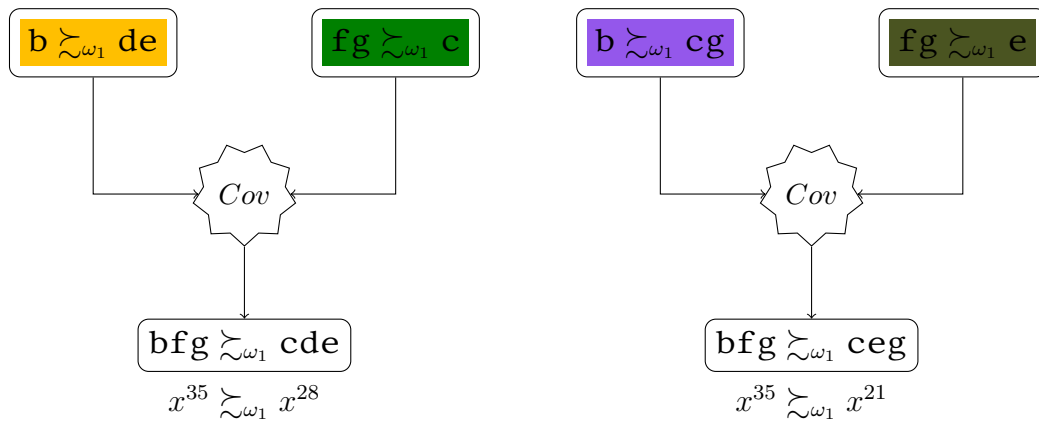


FIGURE 5.3 – Recommandation de x^{35} : explications de $x^{35} \succ_{\omega_1} x^{28}$ et $x^{35} \succ_{\omega_1} x^{21}$

Quant à la comparaison par paire $x^{35} \succ_{\omega_1} x^{72}$, on remarquera qu'elle est non décomposable à partir des éléments de Δ_{ω_1} car comme le montre la Table 5.4, il n'existe pas de sous-ensemble de Δ_{ω_1} vérifiant simultanément les Équations 5.4, 5.5 et 5.6. En effet, cette table qui inventorie les 15 sous-ensembles $\{(u_1, v_1), (u_2, v_2)\}$ de Δ_{ω_1} tels que : $v_1 \cup v_2 = \text{ad} \equiv x^{72}$ (Équation 5.6), permet de remarquer que pour chacun d'eux l'Équation 5.4 (disjonction de u_1 et u_2) est violée.

Il s'ensuit que la recommandation x^{35} est inexplicable.

TABLE 5.4 – Éléments de preuve de l’inexplicabilité de $x^{35} \succ_{\omega_1} x^{72}$

| | | | | | | | | | |
|--------------|---------|---------|----------|---------|---------|----------|----------|----------|----------|
| (u_1, v_1) | (b, d) | (b, d) | (b, d) | (bf, d) | (bf, d) | (bf, d) | (bfg, d) | (bfg, d) | (bfg, d) |
| (u_2, v_2) | (bf, a) | (bg, a) | (bfg, a) | (bf, a) | (bg, a) | (bfg, a) | (bf, a) | (bg, a) | (bfg, a) |

| | | | | | | |
|--------------|---------|---------|----------|---------|---------|----------|
| (u_1, v_1) | (fg, d) | (fg, d) | (fg, d) | (b, g) | (b, g) | (b, g) |
| (u_2, v_2) | (bf, a) | (bg, a) | (bfg, a) | (bf, a) | (bg, a) | (bfg, a) |

TABLE 5.5 – Table de performance A_3 et scores suivant ω_1 (Équation 5.2)

| | A_3 | $\omega_1(\cdot)$ |
|---|-----------------------------|---------------------------|
| * | $x^{74} \equiv \text{adf}$ | $128 + 59 + 41 = 228$ |
| | $x^{29} \equiv \text{cdeg}$ | $77 + 59 + 52 + 37 = 225$ |
| | $x^{22} \equiv \text{cef}$ | $77 + 52 + 41 = 170$ |
| | $x^{21} \equiv \text{ceg}$ | $77 + 52 + 37 = 166$ |

Exemple 5.5 Soit la table de performance ci-dessous qui décrit un ensemble A_3 de quatre alternatives et fournit leurs scores suivant la (même) fonction ω_1 de l’Équation 5.2 :

L’alternative à recommander est x^{74} . Les explications des comparaisons par paire $x^{74} \succ_{\omega_1} x^{21}$ et $x^{74} \succ_{\omega_1} x^{22}$ sont fournies à la Figure 5.4 et la comparaison par paire $x^{74} \succ_{\omega_1} x^{29}$ (entre la meilleure alternative et la deuxième) est inexplicable.

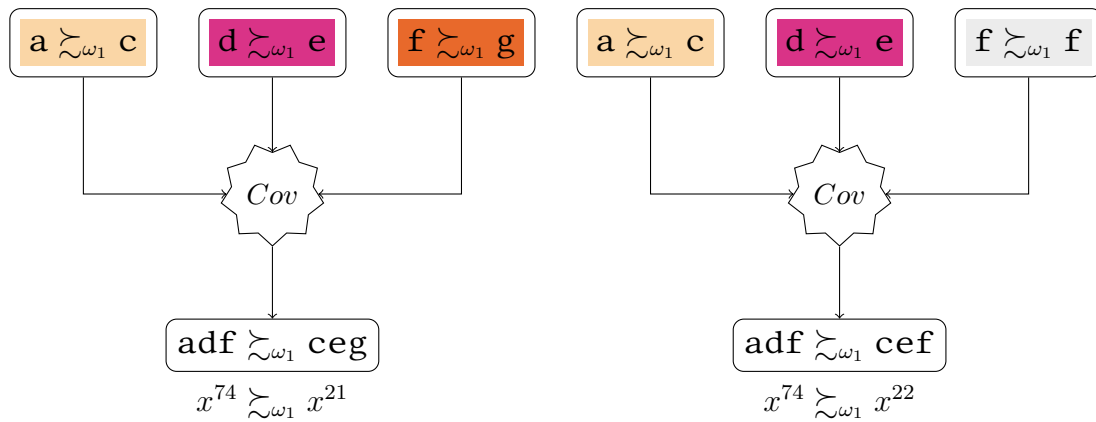


FIGURE 5.4 – Recommandation de x^{74} : explications de $x^{74} \succ_{\omega_1} x^{21}$ et $x^{74} \succ_{\omega_1} x^{22}$

La recommandation x^{74} est donc aussi inexplicable.

« Que faire ? » lorsqu’une recommandation (comme celles des Exemples 5.4 et 5.5) est inexplicable. L’objet de cette section est de proposer trois mesures qui peuvent être prises pour remédier aux situations d’inexplicabilité de la recommandation.

La première intitulée *Recours à la transitivité* (sous-section 5.4.1) consiste à procéder à une modification de la structure de la recommandation : passer d’une structure d’arbre

de hauteur 1 (les comparaisons par paire directes (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$) à une arborescence de hauteur supérieure ou égale à 2. Autrement dit, il s'agira de retenir un ensemble de $|\mathbf{A}| - 1$ comparaisons par paire de $\mathbb{A} = \mathbf{A} \times \mathbf{A}$ compatibles avec ω , différent de l'ensemble $\{(x^*, y), y \in \mathbf{A} \setminus \{x^*\}\}$, qui suffisent (en recourant à la propriété de transitivité) à soutenir la recommandation de x^* et qui sont toutes explicables par couverture.

La deuxième mesure que nous préconisons est la *Relaxation de la recommandation* (sous-section 5.4.2). Elle consiste à fournir l'explication d'une recommandation partielle de type k -best (avec $k \leq k_{max} < |\mathbf{A}|$ et k_{max} une borne limite tolérable). À travers cette proposition, nous défendons l'intérêt de soutenir par une explication, une recommandation partielle à défaut de produire celle (complète) de la recommandation requise.

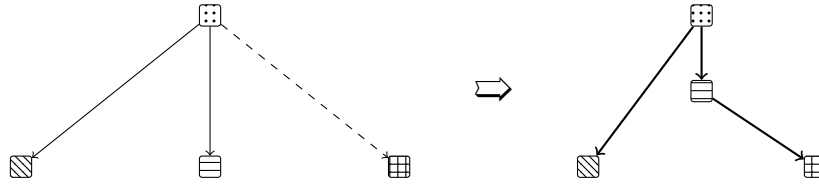
La troisième mesure (voir sous-section 5.4.3) intitulée *Translation de la fonction de score*, consiste à s'autoriser une « légère modification » de la fonction de score ω . Cette modification s'effectue en garantissant que l'alternative x^* recommandée étant donné ω reste celle à recommander au regard de la nouvelle fonction de score ω' . Dans une situation d'aide à la décision, cette façon de remédier à l'inexplicabilité de la recommandation nous semble acceptable pour plusieurs raisons. En effet, la fonction de score ω n'est qu'une représentation des préférences du décideur [Bouyssou et al., 2006b]. C'est un paramètre qui évolue constamment au cours du processus d'élicitation [Belton and Stewart, 2002] et qui n'est utile en réalité qu'à l'analyste qui, à travers cette représentation, s'assure de la compatibilité des préférences exprimées et collectées avec un type de modèle de préférences précis (ici le modèle additif). Ce qui en revanche intéresse le décideur, c'est la recommandation qui lui est faite et qu'il peut interroger d'où la nécessité d'être en capacité de lui en fournir une explication.

Précisons, avant la présentation des algorithmes correspondant à ces différentes propositions de mesures, que ces dernières peuvent être combinées (par exemple, *Recours à la transitivité* et *Relaxation de la recommandation*) pour accroître les capacités d'un analyste à produire à la demande d'un décideur une explication de la recommandation qu'il lui a faite. Dans les sous-sections qui suivent, nous ne donnerons aucune indication sur la façon dont ces combinaisons pourraient être implémentées.

5.4.1 Recours à la transitivité

La Figure 5.5 illustre le *Recours à la transitivité*¹ : on y voit le passage d'une recommandation (en étoile) inexplicable à une recommandation dont toutes les comparaisons par paire sont explicables par couverture. La comparaison par paire non décomposable (en pointillés) dans l'arbre de gauche fait usage, pour son explication dans l'arbre de droite, de la propriété de transitivité et s'appuie sur l'explication de deux autres comparaisons par paire.

1. Le choix de la propriété de transitivité se justifie essentiellement par le fait qu'il s'agit d'une propriété à laquelle sont familiers la plupart des décideurs. Au sujet de cette propriété, [Payne et al., 1999]

FIGURE 5.5 – Inexplicitité : illustration du *Recours à la transitivité*

De façon générale, cette proposition construit (lorsqu'elle existe) une arborescence compatible avec ω et représentant un ensemble de comparaisons par paire (toutes décomposables) qui soutiennent la recommandation de l'alternative x^* . Son implémentation peut s'effectuer à l'aide du programme linéaire en nombres entiers $\mathcal{P}_r^\dagger(\mathbf{A}, \omega)$ décrit ci-dessous. À la différence du PLNE $\mathcal{P}_r(\mathbf{A}, \omega)$ de la sous-section 5.3.3, sont ici prises en compte toutes les $\left(1 + 2 + \dots + (n - 1) = \frac{n \times (n - 1)}{2}\right)$ comparaisons par paire de \mathbf{A} compatibles avec ω au lieu des seules comparaisons directes (x^*, y) . Chaque comparaison par paire sera identifiée à l'aide de la paire d'entiers (k, l) représentant les rangs suivant ω des alternatives qui la composent ($1 \leq k < l \leq |\mathbf{A}| = n$).

Les variables de décision :

1. Variables binaires $p^{(k,l)}$ valant 1 si la comparaison par paire $(x^{(k)}, x^{(l)})$ est retenue dans l'arborescence finale de la recommandation.
2. Variables binaires $b_{ij}^{1,(k,l)}$ et $b_{ij}^{2,(k,l)}$ qui vont permettre de reconstruire l'explication de chaque comparaison par paire $(x^{(k)}, x^{(l)})$ de l'arborescence finale de la recommandation.
3. Variables binaires β_{ij} qui jouent le même rôle que dans le programme $\mathcal{P}_r(\mathbf{A}, \omega)$ de la sous-section 5.3.3.

Les contraintes :

1. Contraintes d'arborescence : chaque alternative (excepté x^*) dans l'arborescence finale a un unique prédécesseur.

$$\sum_{k \in \llbracket 1; l-1 \rrbracket} p^{(k,l)} = 1 \quad \forall l \in \llbracket 2; n \rrbracket \quad (5.21)$$

2. Contraintes garantissant l'explicitité (Équations 5.4, 5.5, 5.6 et appartenance des sous-comparaisons explicatives à $\Delta_{\succsim \omega}$) des comparaisons par paire $(x^{(k)}, x^{(l)})$ retenues.

font remarquer que la plupart des décideurs estiment que leurs préférences devraient être transitives.

$$b_{ij}^{1,(k,l)} + \sum_{j' \in x^{(l)}} b_{ij'}^{2,(k,l)} \leq 1 \quad \forall i \in x^{(k)} \quad \forall j \in x^{(l)} \quad (5.22)$$

$$b_{ij}^{2,(k,l)} + \sum_{i' \in x^{(k)}} b_{i'j}^{1,(k,l)} \leq 1 \quad \forall i \in x^{(k)} \quad \forall j \in x^{(l)} \quad (5.23)$$

$$\sum_{j \in x^{(l)}} b_{ij}^{1,(k,l)} \times \omega_j \leq \omega_i \quad \forall i \in x^{(k)} \quad (5.24)$$

$$\sum_{i \in x^{(k)}} (b_{ij}^{1,(k,l)} + b_{ij}^{2,(k,l)}) \times \omega_i \geq p^{(k,l)} \times \omega_j \quad \forall j \in x^{(l)} \quad (5.25)$$

3. Contraintes reliant les variables $b_{ij}^{1,(k,l)}$, $b_{ij}^{2,(k,l)}$ et β_{ij} :

$$\sum_{y \in \mathbf{A} \setminus \{x^*\}} b_{ij}^{1,(k,l)} \mathbb{1}_{\{i\}, \{j\}}^{(k,l)} \leq M \times \beta_{ij} \quad \forall i \in [m] \quad \forall j \in [m] \quad (5.26)$$

$$\sum_{y \in \mathbf{A} \setminus \{x^*\}} b_{ij}^{2,(k,l)} \mathbb{1}_{\{i\}, \{j\}}^{(k,l)} \leq M \times \beta_{ij} \quad \forall i \in [m] \quad \forall j \in [m] \quad (5.27)$$

avec $\mathbb{1}_{\{i\}, \{j\}}^{(k,l)}$ valant 1 si et seulement si $i \in x^{(k)}$ et $j \in x^{(l)}$ et M une constante arbitrairement grande supérieure à $\frac{n \times (n-1)}{2}$.

La fonction objectif (optionnelle) : La même que celle de l'Équation 5.20.

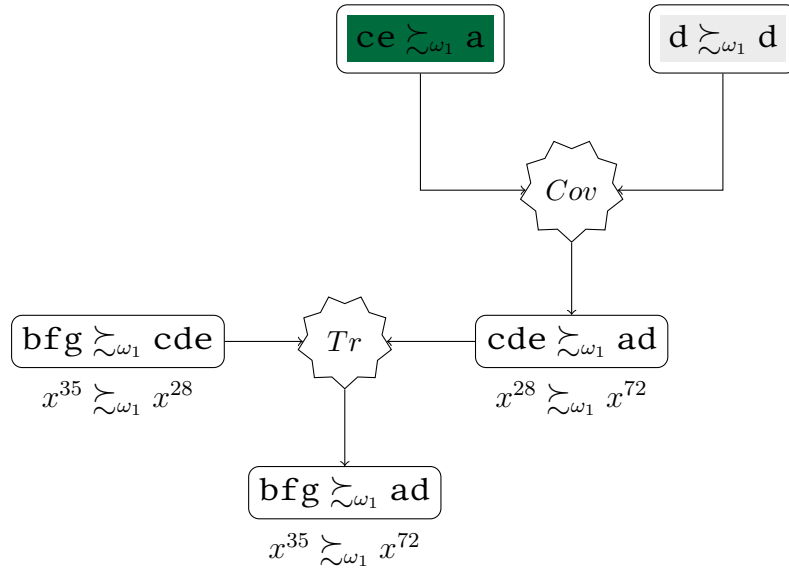
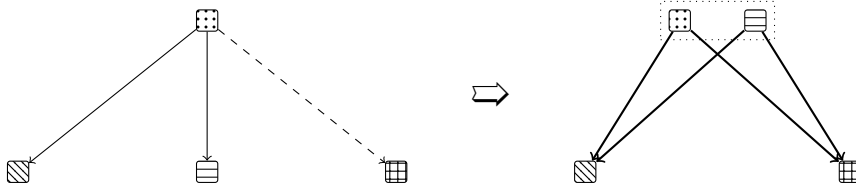
Exemple 5.6 (Suite des Exemples 5.1 et 5.4) Une sortie de l'exécution du programme $\mathcal{P}_r^\dagger(\mathbf{A}_2, \omega_1)$ – appliquée à l'instance (\mathbf{A}_2, ω_1) de l'Exemple 5.1 – produit une explication de la recommandation x^{35} où :

- Les explications des comparaisons par paire $x^{35} \succ_{\omega_1} x^{28}$ et $x^{35} \succ_{\omega_1} x^{21}$ sont représentées à la Figure 5.3.
- L'explication de la comparaison par paire $x^{35} \succ_{\omega_1} x^{72}$ (inexplicable par couverture) est obtenue en recourant à la transitivité (voir Figure 5.6).

5.4.2 Relaxation de la recommandation : du 1–best au k –best

En guise d'illustration de la solution de *Relaxation de la recommandation*, nous proposons la Figure 5.7.

Sur cette figure, on observe le passage d'une recommandation de type 1–best (dont une des comparaisons par paire (x^*, y) n'est pas explicable par couverture) à une recommandation de type 2–best, explicable par couverture. L'ensemble des alternatives \mathbf{A} est donc partitionné en deux sous-ensembles : celui des k meilleures alternatives et celui des

FIGURE 5.6 – Explication de la recommandation de x^{35} via un recours à la transitivitéFIGURE 5.7 – Inexplicabilité : illustration de la *Relaxation de la recommandation*

$n - k$ autres. La *Relaxation de la recommandation* consiste à justifier la préférence globale de chacune des k meilleures alternatives à l'ensemble des $n - k$ autres non choisies.

Nous proposons pour son implémentation, l'Algorithme 4.

Cet algorithme tentera de répondre à la question suivante : « Existe-t-il une recommandation explicable par couverture des k meilleures alternatives où $k \leq k_{max}$? » avec $k_{max} \geq 2$, un paramètre représentant une borne maximale tolérable.

On remarquera que l'Algorithme 4 retourne (lorsqu'elle existe) l'explication d'une recommandation de type k -best avec une valeur k minimale ($k \leq k_{max}$).

Exemple 5.7 (Suite de l'Exemple 5.5) Dans l'Exemple 5.5, on a vu que la recommandation de l'alternative x^{74} ne peut être justifiée car la comparaison par paire $x^{74} \succ_{\omega_1} x^{29}$ (entre la meilleure alternative et la deuxième) est n'est pas explicable par couverture. L'application de l'Algorithme 4 à l'instance (\mathbf{A}_3, ω_1) de cet exemple avec comme paramètre $k_{max} = 2$ retourne :

– pour le sous-ensemble \mathbf{A}_{31}^2 (Ligne 3 de l'Algorithme 4) de \mathbf{A}_3 composé des alternatives

Algorithme 4 Inexplicabilité : implémentation de la *Relaxation de la recommandation***Entrée(s)** \mathbf{A} , ω et k_{max} **Sortie(s)** un booléen explicable indiquant l'existence d'une recommandation de type k -best explicable par couverture avec $k \leq k_{max}$ et une table TX contenant les explications des comparaisons par paire sur laquelle elle repose.

- 1: Trier par ordre décroissant de score (suivant ω) les alternatives de l'ensemble \mathbf{A} : $x^{(1)}, x^{(2)}, x^{(3)}, \dots, x^{(n)}$
- 2: Initialiser TX
- 3: Soit \mathbf{A}_i^k le sous-ensemble de \mathbf{A} composé de $x^{(i)}$ et des $n - k$ « pires » alternatives (de $x^{(k+1)}$ à $x^{(n)}$)
- 4: $k \leftarrow 2$
- 5: **tant que** ($k \leq k_{max}$) **faire**
- 6: **pour** i allant de 1 à k **faire**
- 7: Exécuter $\mathcal{P}(\mathbf{A}_i^k, \omega)$
- 8: **si** le PLNE n'a pas de solution **alors**
- 9: $k \leftarrow k + 1$
- 10: Retourner à la ligne 5
- 11: **sinon**
- 12: Rajouter à TX l'ensemble des explications obtenues
- 13: **fin du si**
- 14: **fin du pour**
- 15: **retourner** True, TX
- 16: **fin du tant que**
- 17: **retourner** False, \square

x^{74} (première suivant ω_1), x^{22} (troisième suivant ω_1) et x^{21} (quatrième suivant ω_1), l'explication représentée à la Figure 5.4 et

- pour le sous-ensemble $\mathbf{A}_{3_2^2}$ de \mathbf{A}_3 composé des alternatives x^{29} , x^{22} et x^{21} , l'explication de la Figure 5.8 relative à la comparaison par paire $x^{29} \succ_{\omega_1} x^{22}$. On remarquera, en ce qui concerne la comparaison par paire $x^{29} \succ_{\omega_1} x^{21}$ qu'il s'agit d'une comparaison par paire de l'ensemble \mathbb{D}_m ($\text{ceg} \subset \text{cdeg}$) c'est-à-dire dispensée d'explication.

5.4.3 Translation de la fonction de score : de ω à ω'

Le schéma de la Figure 5.9 illustre la troisième mesure proposée pour remédier à l'inexplicabilité d'une recommandation. C'est cette solution, qui, conservant la structure de la recommandation (les comparaisons par paire directes (x^*, y)), s'autorise à modifier la fonction de score ω **sans que ne change la recommandation** x^* . En introduction de section, nous avons tenté d'apporter quelques arguments en faveur de cette proposition.

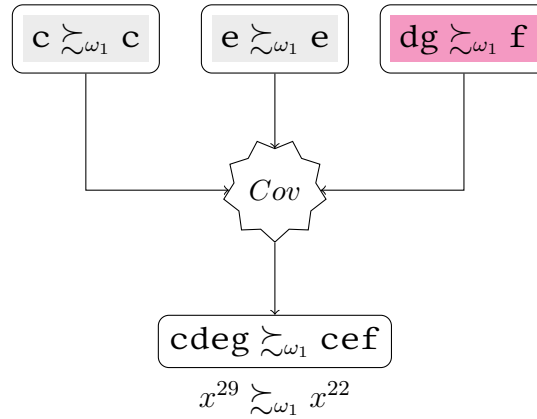
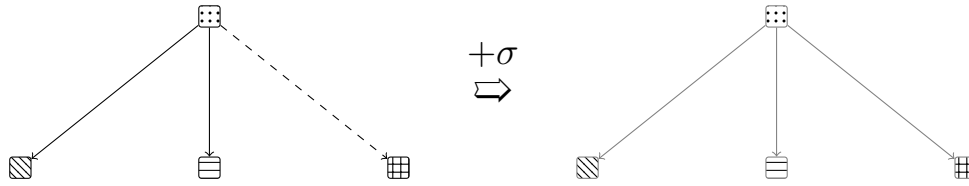


FIGURE 5.8 – Inexplicabilité : application de la relaxation de la recommandation

Ici, nous nous contenterons d'en expliciter le “mode d'action” et les détails d'implémentation.

FIGURE 5.9 – Incomplétude : illustration de la *Translation* de ω

Concrètement, il s'agira d'appliquer aux composantes du vecteur ω une translation « contrainte » d'un vecteur $\sigma = [\sigma_1^+ - \sigma_1^-, \dots, \sigma_m^+ - \sigma_m^-]$ de sorte à avoir $\omega'_i = \omega_i + (\sigma_i^+ - \sigma_i^-)$ où σ_i^+ et σ_i^- représentent des variables de déviation positives. La « contrainte » imposée à σ peut prendre plusieurs formes :

- Prendre σ de sorte que l'ensemble des préférences exprimées lors de la phase d'élicitation restent compatibles avec $\omega' = \omega + \sigma$; ce qui, dans le cadre du présent chapitre ne saurait être appliqué vu que nous ne tenons pas compte desdites préférences.
- Prendre σ de sorte que ω' conserve, en plus des comparaisons par paire directes (x^*, y) (indispensables pour avoir la même recommandation x^*), toutes (ou une majorité) des autres comparaisons par paire induites par ω ; ce qui computationnellement, peut s'avérer coûteux.
- Prendre σ de sorte que ω' et ω induisent le même ordre sur l'ensemble des critères (voir Équation 5.28)

$$\omega_i \geq \omega_j \iff \omega'_i \geq \omega'_j \quad \forall i, j \in [m] \quad (5.28)$$

C'est cette dernière option qui sera retenue dans la présentation de l'algorithme de mise en œuvre de la mesure de translation. Cet algorithme consistera en la résolution d'un programme linéaire mixte (variables entières et continues). Nous supposons² cependant que la fonction de score ω est sous une forme normalisée ($\sum_{i \in [m]} \omega_i = 1$). Aussi, dans le but de faciliter l'écriture de la Contrainte (5.35), nous supposons que : $\omega_i \geq \omega_{i-1}$ pour tout $i \in \llbracket 1; m-1 \rrbracket$.

Soit le programme linéaire mixte $\mathcal{P}'_r(\mathbf{A}, \omega)$ décrit comme suit :

Les variables de décision :

1. Variables binaires $b_{ij}^{1,y}$ et $b_{ij}^{2,y}$ (les mêmes que dans la sous-section 5.3.3).
2. Variables continues positives σ_i^+ et σ_i^- telles que :

$$\omega'_i = \omega_i + (\sigma_i^+ - \sigma_i^-) \quad \forall i \in [m]$$

3. Variables binaires β_{ij} (les mêmes que dans la sous-section 5.3.3).

Les contraintes :

1. Contraintes garantissant l'explicabilité (Équations 5.4, 5.5, 5.6 et appartenance des sous-comparaisons explicatives à $\Delta_{\sim_{\omega'}}$) des comparaisons par paire (x^*, y) , $y \in \mathbf{A} \setminus \{x^*\}$.

$$b_{ij}^{1,y} + \sum_{j' \in y} b_{ij'}^{2,y} \leq 1 \quad \forall i \in x^* \quad \forall j \in y \quad (5.29)$$

$$b_{ij}^{2,y} + \sum_{i' \in x^*} b_{i'j}^{1,y} \leq 1 \quad \forall i \in x^* \quad \forall j \in y \quad (5.30)$$

$$\sum_{j \in y} b_{ij}^{1,y} \times (\omega_j + (\sigma_j^+ - \sigma_j^-)) \leq (\omega_i + (\sigma_i^+ - \sigma_i^-)) \quad \forall i \in x^* \quad (5.31)$$

$$\sum_{i \in x^*} (b_{ij}^{1,y} + b_{ij}^{2,y}) \times (\omega_i + (\sigma_i^+ - \sigma_i^-)) \geq (\omega_j + (\sigma_j^+ - \sigma_j^-)) \quad \forall j \in y \quad (5.32)$$

2. Préservation de la normalisation

$$\sum_{i \in [m]} \sigma_i^+ = \sum_{i \in [m]} \sigma_i^- \quad (5.33)$$

2. ...sans perte de généralité. Cette hypothèse facilite la linéarisation de certaines contraintes dudit programme linéaire mixte.

3. Non négativité des composantes ω'_i

$$\sigma_i^- \leq \omega_i \quad \forall i \in [m] \quad (5.34)$$

4. Préservation de l'ordre sur les critères (Équation 5.28)

$$\omega_i + (\sigma_i^+ - \sigma_i^-) \geq \omega_{i+1} + (\sigma_{i+1}^+ - \sigma_{i+1}^-) \quad \forall i \in \llbracket 1; m-1 \rrbracket \quad (5.35)$$

5. Contraintes reliant les variables $b_{ij}^{1,y}$, $b_{ij}^{2,y}$ et β_{ij} : les mêmes que dans la programme $\mathcal{P}_r(\mathbf{A}, \omega)$ de la sous-section 5.3.3.

On remarquera que les Contraintes (5.31) et (5.32) ne sont pas linéaires. Elles comprennent des produits d'une variable booléenne et d'une variable continue positive inférieure à 1 ($= \sum_{i \in [m]} \omega_i$). C'est le cas par exemple de $b_{ij}^{2,y} \times \sigma_i^-$, $b_{ij}^{1,y} \times \sigma_i^+$, $b_{ij}^{2,y} \times \sigma_i^+$ and $b_{ij}^{1,y} \times \sigma_i^-$ dans le membre de gauche de l'Equation (5.32). Ce type de produits peut être facilement linéarisé grâce à l'introduction d'une nouvelle variable $z = b \times \sigma$ (où σ est la variable continue comprise entre 0 et 1 et b la variable booléenne) et l'ajout des quatre contraintes supplémentaires suivantes $z \leq b$, $z \leq \sigma$, $z \geq \sigma - (1 - b)$ et $z \geq 0$.

La fonction objectif (optionnelle) : La même que celle de l'Équation 5.20.

Exemple 5.8 (Suite de l'Exemple 5.5) *En passant de la fonction de score ω_1 de l'instance (\mathbf{A}_3, ω_1) de l'Exemple 5.5 à la fonction de score ω'_1 (Équation 5.36) obtenue en retirant une unité au critère e que l'on rajoute au critère a , on obtient une sortie possible de l'exécution du programme $\mathcal{P}'_r(\mathbf{A}_3, \omega_1)$ illustrée à la Figure 5.10.*

$$\omega'_1 = (\underline{a} : 129, b : 126, c : 77, d : 59, \underline{e} : 51, f : 41, g : 37) \quad (5.36)$$

5.5 Expérimentations numériques

Dans cette section, nous désignerons sous le vocable d'*instance*, la paire (\mathbf{A}, ω) composée de l'ensemble fini des alternatives sur lequel porte la recommandation et de la fonction de score représentant les préférences du décideur. Les expérimentations numériques que nous conduisons ici visent trois objectifs :

1. Estimer la proportion d'*instances* explicables par couverture (Section 5.3) et la comparer à celles explicables uniquement à l'aide des *preference swaps*.
2. Quantifier les gains apportés par chacune des mesures de traitement de l'inexplicabilité (Section 5.4).
3. Fournir des indications sur les temps d'exécution des différents algorithmes proposés; ceux-ci permettront de savoir si ces algorithmes peuvent être utilisés en situation réelle d'aide à la décision.

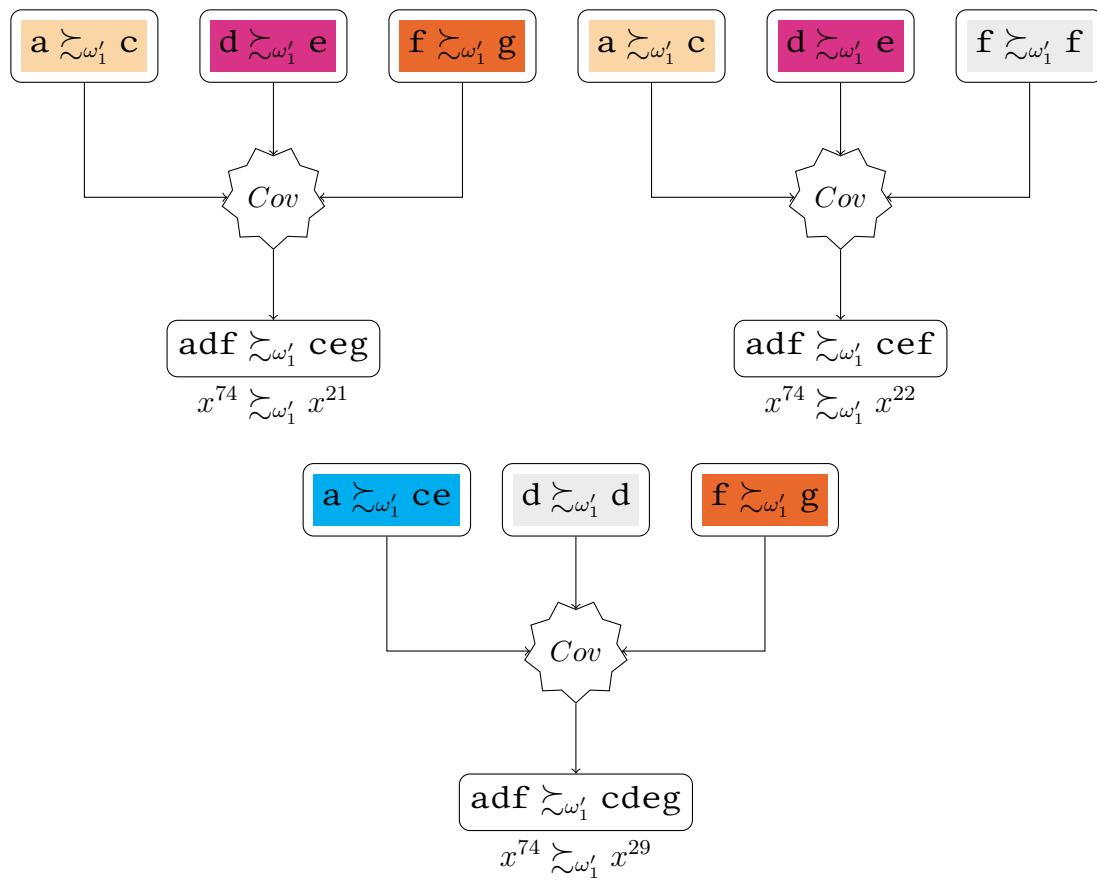


FIGURE 5.10 – Recommandation de x^{74} : explication après translation de la fonction de score

5.5.1 Description du protocole expérimental

Nous avons constitué 9 grands groupes d'instances aléatoires en fonction de la taille du problème qu'elles représentent : $m \in \llbracket 7; 15 \rrbracket$. Étant donné m , chaque instance est obtenue (indépendamment des autres instances du même groupe) en générant aléatoirement $n = |\mathbf{A}|$ alternatives sur lesquelles doit porter la recommandation et une fonction de score ω . Chaque fonction de score est tirée (pour chaque valeur $m \in \llbracket 7; 15 \rrbracket$) suivant la méthode préconisée par [Butler et al., 1997]. Quant aux tirages aléatoires des ensembles \mathbf{A} , ils ont été réalisés (de façon indépendante l'une de l'autre) de la manière suivante :

- Tirage uniforme sans remise de n entiers naturels entre 0 et $2^m - 1$,
- Conversion en binaire des entiers tirés,
- Conversion de chaque entier x sous numérotation binaire en alternative : chaque

alternative (sous sa forme ensembliste) est composée de l'ensemble des critères i tel que le i -ème bit de x vaut 1.

Il convient de signaler que l'on ne retient une instance que si elle vérifie les deux propriétés suivantes :

(P1) *Significativité des m critères* : cette propriété garantit que les ensembles \mathbf{A} générés sont bien définis sur m critères. Elle peut s'exprimer mathématiquement comme suit :

$$\forall i \in [m], \exists x, y \in \mathbf{A} \mid i \in x \text{ et } i \notin y \quad (5.37)$$

(P2) *Absence de dominance au sens de Pareto* : cette propriété justifie l'utilité de la fonction de score ω dans la désignation de la meilleure alternative (à recommander). Elle évite que soit d'office (sans besoin de calculer son score) écartée une alternative dominée par une autre au sens de Pareto. Elle peut s'exprimer mathématiquement de la façon suivante :

$$\forall x, y \in \mathbf{A}, \exists i, j \in [m] \mid i \in x \text{ et } j \notin x \Rightarrow i \notin y \text{ et } j \in x \quad (5.38)$$

Paramètres retenus et autres éléments de compréhension : Pour chaque valeur $m \in \llbracket 7; 15 \rrbracket$, nous avons – arbitrairement – pris $n = 10$. On cherche donc à expliquer la recommandation du choix d'une alternative parmi dix. Chaque groupe d'*instances* en contient exactement 5000 et on mesure (simplement) dans chacun des cas traités, la proportions d'*instances* explicables.

5.5.2 Résultats et discussions

Dans cette sous-section, nous présenterons et commenterons les résultats des expérimentations numériques réalisées.

Estimer la proportion d'*instances* explicables par couverture et la comparer à celles explicables uniquement à l'aide des *preference swaps*. La Figure 5.11 fournit une estimation des proportions de recommandations explicables par couverture en fonction du nombre m de critères lorsque d'une part, les comparaisons par paire admises sont exclusivement des *preference swaps* et d'autre part, lorsque l'on considère tous les types de comparaisons par paire structurellement non décomposables. On observe, dans le cas des *preference swaps*, que le pouvoir explicatif du schéma de couverture décroît strictement entre $m = 7$ et $m = 15$ passant de 42,8% à 18,7%. Quant aux proportions de recommandations explicables dans le *contexte* $(\succsim_\omega, \Delta_{\succsim_\omega}, cov)$, elles semblent constantes entre les valeurs $m = 7$ et $m = 10$ (autour de 92%) et augmentent de 92,4% (lorsque $m = 10$) à 96,3% (quand m vaut 15). Cette augmentation peut s'expliquer par le fait de maintenir la taille de l'ensemble \mathbf{A} constante égale à 10. En effet, lorsque le

nombre de critères augmente, la “distance” entre les différentes alternatives aura tendance à augmenter rendant davantage possible des décompositions du type « un contre plusieurs » et « plusieurs contre un » et ceci, d’autant plus que la taille de l’ensemble $\Delta_{\sim\omega}$ augmente beaucoup plus vite (exponentiellement) en fonction de m (Définition 5.1) que celle de l’ensemble $S_{\sim\omega}$ des *preference swaps* (évolution quadratique).

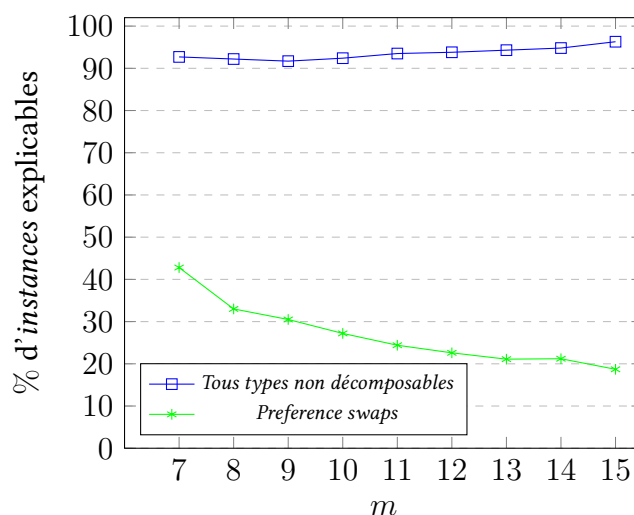


FIGURE 5.11 – Pourcentage d’instances explicables par couverture en fonction du type de comparaisons par paire admises utilisées

TABLE 5.6 – Valeurs numériques de la Figure 5.11

| m | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Tous types non décomposables | 92.7% | 92.2% | 91.7% | 92.4% | 93.5% | 93.8% | 94.3% | 94.8% | 96.3% |
| Preference swaps | 42.8% | 33.0% | 30.5% | 27.2% | 24.4% | 22.6% | 21.1% | 21.2% | 18.7% |

La plus faible proportion relevée est de 91,7% avec $m = 9$. Les expérimentations réalisées permettent donc de conclure que dans 9 cas sur 10, l’utilisation du schéma de couverture à partir de comparaisons par paire structurellement non décomposables permet la justification du choix de la meilleure alternative parmi l’ensemble des dix alternatives considérées.

Quantifier les gains apportés par chacune des mesures de traitement de l’inexplicitabilité La Figure 5.12 nous donne une idée de l’intérêt de chacune des trois mesures de traitement de l’inexplicitabilité proposées à la Section 5.4.

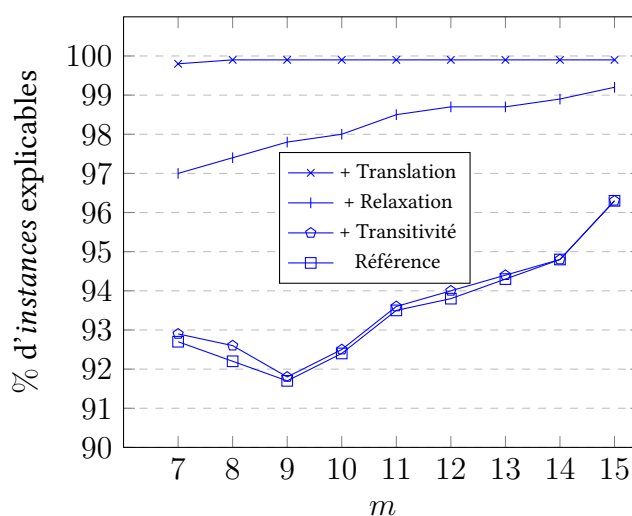


FIGURE 5.12 – Pourcentage d’instances explicables avec mise en œuvre des mesures de traitement de l’inexplicabilité

TABLE 5.7 – Valeurs numériques de la Figure 5.12

| m | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| + Translation | 99.8% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% | 99.9% |
| + Relaxation | 97.0% | 97.4% | 97.8% | 98.0% | 98.5% | 98.7% | 98.7% | 98.9% | 99.2% |
| + Transitivité | 92.9% | 92.6% | 91.8% | 92.5% | 93.6% | 94.0% | 94.4% | 94.8% | 96.3% |
| Référence | 92.7% | 92.2% | 91.7% | 92.4% | 93.5% | 93.8% | 94.3% | 94.8% | 96.3% |

En effet, cette figure recense les proportions d’instances dont la recommandation est explicable par couverture ou pour lesquelles le recours à la transitivité, la relaxation de la recommandation ou la translation de la fonction de score aboutit.

On observe que la tentative consistant à justifier pourquoi les deux meilleures alternatives doivent être préférées aux autres, à défaut d’être en mesure d’expliquer le choix de la meilleure alternative aboutit en moyenne pour 4.7% d’instances supplémentaires.

En comparaison, le recours à la transitivité offre des gains beaucoup plus modestes : au plus 0,4% d’instances supplémentaires expliquées. Si l’on fait l’hypothèse que (très souvent), dans les cas d’inexplicabilité de la recommandation, la comparaison par paire entre la meilleure alternative $x^{(1)}$ et la seconde meilleure $x^{(2)}$ n’est pas explicable par couverture, alors ce faible gain s’explique. En effet, toutes les arborescences “couvrantes” de A suivant ω comportent l’arc $(x^{(1)}, x^{(2)})$. Si la comparaison par paire correspondante n’est pas explicable par couverture, alors le recours à la transitivité ne pourra pas aboutir à la construction d’une arborescence dont tous les arcs sont explicables (voir sous-section 5.4.1). L’observation qui semble soutenir cette hypothèse est celle des gains de la mesure de relaxation de la recommandation dans laquelle on se dispense de l’explication de la

comparaison par paire $(x^{(1)}, x^{(2)})$.

Quant à la mesure de *translation de la fonction de score* ω , on remarque qu'elle surclasse les deux autres : sur l'ensemble des valeurs m considérées, on observe une capacité à expliquer quasiment toutes les recommandations de l'alternative x^* lorsque l'on s'autorise à modifier la fonction de score ω tout en maintenant l'ordre qu'elle induit sur l'ensemble des critères.

Fournir des indications sur les temps d'exécution des différents algorithmes proposés Il faut noter que l'ensemble des résolutions des programmes mathématiques a été réalisé à l'aide du solveur Gurobi³ installé sur une machine Linux Intel Core i7 sous la forme d'une librairie externe de Python (Version 3.7). Nous avons observé qu'aucun calcul d'explication de recommandation n'a nécessité une durée d'exécution supérieure à la seconde (au plus 0,6 s pour le recours à la méthode de translation avec $m = 15$ qui s'implémente à l'aide d'un programme mathématique quadratique dont la linéarisation de certaines contraintes font augmenter la taille du programme linéaire mixte qui en résulte). Il ne fait dès lors aucun doute que l'utilisation des algorithmes proposés, peut être envisagée dans des situations réelles d'aide à la décision.

Le code correspondant aux expérimentations numériques décrites dans cette section est accessible ici⁴.

5.6 Conclusion et perspectives

Ce chapitre a eu pour objet l'explication de la recommandation, aboutissement du processus d'aide à la décision. Nous nous y sommes intéressé à la problématique du choix qui repose, à l'image du rangement, sur une évaluation relative des alternatives. Nous avons considéré qu'expliquer le choix d'une alternative, c'est expliquer l'ensemble des comparaisons par paire sur laquelle elle repose. L'approche d'explication adoptée est celle que nous avons défendue tout au long de ce travail et qui est centrée autour de la définition d'un *contexte* : dans ce chapitre, nous nous sommes limité au schéma déductif de couverture qui explicite l'explication d'une comparaison par paire à partir de sous-comparaisons non décomposables appartenant à la relation binaire induite par une fonction de score représentant les préférences d'un décideur. Nous avons montré que le problème de décision associé au calcul d'une telle explication est NP-complet et en avons proposé une modélisation sous la forme d'un programme linéaire en nombres entiers.

Ayant assimilé l'explication de la recommandation à celle des comparaisons par paire confrontant la meilleure alternative à toutes les autres, nous concluons à l'inexplicabilité

3. <https://www.gurobi.com/>

4. <https://github.com/tetras92/SchemasDeductifsEnAMCD.git>

de la recommandation dès que l'une de ces comparaisons par paire n'est pas explicable dans le *contexte* considéré. Dans ce cas, nous suggérons que soit tentée la mise en œuvre de mesures pouvant garantir la délivrance d'une explication au décideur ; cette dernière pouvant être partielle (le cas de la mesure de la *relaxation de la recommandation*). Les deux autres mesures proposées sont le *recours à la transitivité* et la *translation de la fonction de score*. Pour chacune de ces "stratégies" de traitement de l'inexplicabilité, nous avons implémenté des algorithmes de mise en œuvre. Au travers d'expérimentations numériques, nous avons pu "mesurer" à quel point chacune d'entre elles pourrait s'avérer intéressante en pratique.

Au regard des résultats expérimentaux encourageants observés, il nous semble important d'évaluer dans des situations réelles d'aide à la décision l'intérêt des explications par couverture de recommandation afin de pouvoir objectivement juger de la pertinence des propositions faites dans ce travail et le cas échéant lui apporter des améliorations. Ensuite, estimons-nous qu'il faudra réfléchir à la transposition du schéma de couverture aux cadres de décision – plus courants en pratique – où sont à prendre en compte plus de deux niveaux d'évaluation de critères.

Chapitre 6

Conclusion générale

Ce dernier chapitre sera composé de deux sections : la Section 6.1 récapitule les principales contributions de notre travail tandis que la Section 6.2 indique un petit nombre de directions dans lesquelles celui-ci peut être étendu.

6.1 Synthèse des contributions et des résultats

L'ambition de ce travail de recherche était de répondre à la question : « Comment expliquer une comparaison par paire d'alternatives dans le cadre d'un modèle additif? ». Pour répondre à cette question nous avons mis en place plusieurs contributions pouvant être résumées comme suit :

Un nouveau framework d'explication

La principale contribution de ce travail est la proposition d'un nouveau cadre d'explication adapté aux situations de décision où les alternatives, décrites sur plusieurs critères, sont évaluées les unes comparativement aux autres (problématiques de choix et de rangement). L'explication de la préférence d'une alternative sur une autre y est réalisée à partir d'autres préférences prenant la forme de comparaisons par paire de type atomique et/ou de déclarations de préférences du décideur. L'explication s'effectue suivant des schémas déductifs pour lesquels des algorithmes dédiés ont été implémentés. Cet ensemble d'algorithmes est à considérer comme une sorte de "boîte-à-outils" servant à prouver comment une comparaison par paire se déduit à partir d'une ou plusieurs autres comparaisons par paire. L'explication, telle que nous la concevons, devient dès lors un outil de déduction au même titre que les programmes linéaires dont la résolution permet de vérifier, sous l'hypothèse de préférences additives, si une alternative est nécessairement préférée à une autre. L'avantage de notre approche est qu'elle offre des garanties supérieures d'intelligibilité – pour un décideur quelconque – des raisons de la déduction faite en explicitant les mécanismes qui y sont en jeu. Son inconvénient, c'est (logique-

ment) le coût computationnel que peut requérir le calcul de certaines explications et l'insuffisance des outils proposés à couvrir tous les cas de déduction.

Différentes représentations équivalentes de l'explication

Puisque les explications produites sont destinées à un décideur humain, l'une des contributions de ce travail a été d'en proposer différentes représentations équivalentes. Ainsi, avons-nous envisagé (Chapitre 2) comme supports possibles de délivrance de l'explication : le langage naturel, un ensemble de formules logiques et une représentation graphique sous forme de diagramme de déduction.

Extension de travaux préexistants d'explication de la relation nécessaire

Comme nous avons essayé de le montrer à travers le Chapitre 4, notre approche d'explication basée sur la définition d'un *contexte* englobe des travaux antérieurs d'explication de déductions de la relation nécessaire. Les présentes contributions complètent ces derniers en intégrant aux explications produites les déclarations de préférences du décideur et en diversifiant les types de schémas déductifs employés.

Mise en exergue de la fonction d'élicitation de l'explication

À travers la proposition d'une ébauche de protocole interactif (Chapitre 4), nous avons tenté de montrer qu'en plus de sa fonction (classique) de justification, l'explication de préférences robustes peut contribuer à l'élicitation des préférences du décideur, contribuant ainsi à faire converger le processus d'aide à la détermination exacte de la recommandation qui lui sera faite. En effet, une même comparaison par paire peut s'expliquer de plusieurs façons différentes ; certaines explications pouvant mobiliser des comparaisons par paire de type atomique (cognitivement simples). Leur validation ou contestation constitue de l'information préférentielle additionnelle collectée auprès du décideur.

Alternative aux langages d'explication de la dominance entre coalitions de critères

Les schémas déductifs proposés et illustrés à travers les Chapitres 2 et 3 constituent une alternative aux langages d'explication utilisés pour justifier l'appartenance d'une option à l'ensemble de Smith ; une telle justification s'appuyant sur l'identification d'une *coalition gagnante* de critères (agents) supportant cette option en comparaison aux options n'appartenant pas à l'ensemble de Smith.

Pouvoir déductif des propriétés de transitivité, congruence et couverture

Les expérimentations numériques conduites en fin de Chapitre 3 nous ont permis de découvrir que pour toute comparaison par paire de toute relation d'ordre linéaire additive définie sur $m = 4, 5, 6$ items est explicable à partir d'une combinaison des propriétés de transitivité, d'indépendance et de couverture lorsque celles-ci s'appliquent à l'ensemble des paires critiques de ladite relation ; les algorithmes que nous avons implémentés ayant permis de calculer chacune de ces explications.

Découverte de la diversité des préférences représentables par un modèle additif

Les propriétés de transitivité, d'indépendance et de couverture à partir desquelles ont été définis les différents schémas déductifs considérés dans ce travail correspondent à des modes de raisonnement différents dont la fréquence et les modalités d'utilisation (leur combinaison avec l'un ou les deux autres modes de raisonnement) pour déduire des comparaisons par paire d'une certaine relation d'ordre linéaire additive, pourraient constituer une "signature" de cette dernière. L'important travail d'expérimentations numériques réalisé nous a permis d'observer un certain nombre de ces "signatures" suivant lesquelles il nous semble possible de regrouper plusieurs ordres linéaires additifs et/ou qui pourraient permettre d'établir entre eux des liens de similitude.

6.2 Perspectives

Avant d'évoquer les perspectives qu'il nous semble intéressantes d'explorer à la suite de ce travail, il nous importe de rappeler les hypothèses sur lesquelles il repose globalement :

- Les préférences du décideur sont additives.
- Les mécanismes de déduction dérivant des propriétés de transitivité, d'indépendance et de couverture correspondent à des modes de raisonnement auxquels est familier le décideur ou qui sont facilement assimilables par lui.
- Leur composition à travers des schémas déductifs composés ne saurait nuire à l'intelligibilité de l'explication résultant de leur mise en œuvre.
- Les *preference swaps* et plus généralement les comparaisons par paire structurellement non décomposables (Chapitre 5) sont l'expression de préférences cognitivement simples et donc compréhensibles pour le décideur au même titre que celles exprimant une dominance de Pareto.

La première perspective que nous formulerons est de tenter de **vérifier empiriquement ces hypothèses** auprès de décideurs réels et à travers des enquêtes portant sur des problèmes de décision sur lesquels leurs préférences ont de bonnes chances d'être additives. Ce travail de vérification pourrait être grandement facilité si l'on disposait d'un système automatique d'aide à la décision que l'on doterait d'un module d'explication capable de générer via une interface graphique et sous différentes formes les explications de recommandations faites. **Disposer d'un tel outil graphique pourrait contribuer à enrichir la connaissance que nous (en tant que chercheur) avons des préférences représentables par un modèle additif** qui, au vu des observations faites à travers ce travail, sont assez diverses.

Ensuite, pouvons-nous **travailler à viser le "100%" en termes de comparaisons par paire déduites explicables** en complétant notre "boîte-à-outils d'algorithmes" de calcul d'explication. Ce faire pourrait nécessiter la définition de nouveaux schémas déductifs à partir des trois propriétés du modèle additif considérées dans ce travail voire conduire à l'identification de nouveaux mécanismes de déduction dont la mise en œuvre à des fins d'explication peut être envisagée. À ce sujet, il nous a été donné de constater (Chapitre 3) qu'en considérant comme ensemble de comparaisons par paire admises celui des *paires critiques* de chaque relation d'ordre linéaire additive définie sur $m = 4, 5, 6$ items, toute comparaison par paire s'explique à partir d'une combinaison des propriétés de transitivité, d'indépendance et de couverture. Une question de recherche qui mérite donc d'être posée est de savoir si **il existerait une valeur $m > 6$ à partir de laquelle cette observation cesse de se confirmer**. D'un point de vue purement expérimental à présent, et sachant que certaines paires critiques se déduisent à partir d'autres paires critiques, il nous semble intéressant de **voir ce que deviennent les résultats expérimentaux de la fin du Chapitre 3 lorsque l'ensemble des paires admises est réduit aux paires critiques non déductibles**.

Quand bien même les temps d'exécution des différents algorithmes implémentés sont peu importants, il convient de procéder pour chacun d'eux et dans la mesure du possible, à une **analyse rigoureuse de leur complexité** afin de pouvoir, sur une base objective, les comparer aux résultats d'autres travaux similaires de la littérature.

En ce qui concerne la fonction d'élicitation (Chapitre 4) que peut remplir l'explication de préférences robustes et tout particulièrement le protocole interactif que nous avons ébauché, **nous pensons qu'il convient de le spécifier plus précisément**; ce travail consistant entre autres à en étudier les conditions de terminaison. Aussi, pourrait-on réfléchir à y rendre possible la collecte d'information préférentielle supplémentaire à travers des comparaisons par paire atomiques autres que les *preference swaps*. Le choix de la conclusion robuste à expliquer à chaque itération et celui de l'explication à présenter au décideur pourraient par ailleurs être dictés par des stratégies d'élicitation (basées par exemple sur le calcul du regret maximal) avec comme objectif de limiter la durée de l'interaction. D'autre part, une attention particulière devra être portée à la question de l'inconsistance car dans un tel protocole, on n'est pas à l'abri du fait que le déci-

deur exprime des préférences contradictoires vis-à-vis du modèle additif. Dans ce cas, il conviendra d'appliquer des techniques dédiées de résolution de l'inconsistance ou mettre à contribution l'"explication" (de deux propositions contradictoires) pour parvenir à cette fin.

Bibliographie

- [Amgoud and Prade, 2009] Amgoud, L. and Prade, H. (2009). Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3) :413–436.
- [Amoussou et al., 2023a] Amoussou, M., Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2023a). Questionable stepwise explanations for a robust additive preference model. *International Journal of Approximate Reasoning*, page 108982.
- [Amoussou et al., 2022] Amoussou, M., Belahcène, K., Maudet, N., Mousseau, V., and Ouerdane, W. (2022). Des explications par étapes pour le modèle additif. In *16èmes Journées d’Intelligence Artificielle Fondamentale (JIAF 2022), Journées d’Intelligence Artificielle Fondamentale PFIA 2022*, pages 35–48, Saint-Etienne, France. Association Française pour l’Intelligence Artificielle.
- [Amoussou et al., 2023b] Amoussou, M., Belahcène, K., Maudet, N., Mousseau, V., and Ouerdane, W. (2023b). Des explications transitives questionnables au service de l’élicitation de préférences additives. *Journées d’Intelligence Artificielle Fondamentale*.
- [Barberà et al., 2004] Barberà, S., Bossert, W., and Pattanaik, P. K. (2004). Ranking sets of objects. In Barberà, S., Hammond, P. J., and Seidl, C., editors, *Handbook of Utility Theory : Volume 2 Extensions*, pages 893–977. Springer US, Boston, MA.
- [Belahcène, 2018] Belahcène, K. (2018). *Towards accountable decision aiding : explanations for the aggregation of preferences*. Theses, Université Paris Saclay (COMUE).
- [Belahcène et al., 2017] Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2017). Explaining robust additive utility models by sequences of preference swaps. *Theory and Decision*, 82(2) :151–183.
- [Belahcène et al., 2019] Belahcène, K., Labreuche, C., Maudet, N., Mousseau, V., and Ouerdane, W. (2019). Comparing options with argument schemes powered by cancellation. In *28th International Joint Conference on Artificial Intelligence (IJCAI-19)*, pages 1537–1543, Macao, Macau SAR China.
- [Belton and Stewart, 2002] Belton, V. and Stewart, T. (2002). *Multiple Criteria Decision Analysis : An Integrated Approach*. Kluwer Academic, Dordrecht.

- [Bogomolnaia and Jackson, 2001] Bogomolnaia, A. and Jackson, M. (2001). The stability of hedonic coalition structures. Games and Economic Behavior, 38 :201–230.
- [Bohman, 1997] Bohman, T. (1997). A construction for sets of integers with distinct subset sums. Electron. J. Comb., 5.
- [Boixel and de Haan, 2021] Boixel, A. and de Haan, R. (2021). On the complexity of finding justifications for collective decisions. In Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, pages 5194–5201. AAAI Press.
- [Boixel and Endriss, 2020] Boixel, A. and Endriss, U. (2020). Automated justification of collective decisions via constraint solving. In Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '20, page 168–176, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.
- [Boixel et al., 2022] Boixel, A., Endriss, U., and Haan, R. d. (2022). A calculus for computing structured justifications for election outcomes. Proceedings of the AAAI Conference on Artificial Intelligence, 36(5) :4859–4866.
- [Boutilier et al., 2004] Boutilier, C., Brafman, R. I., Domshlak, C., Hoos, H. H., and Poole, D. (2004). Cp-nets : A tool for representing and reasoning with conditional ceteris paribus preference statements. J. Artif. Int. Res., 21(1) :135–191.
- [Bouyssou et al., 2006a] Bouyssou, D., Dubois, D., Pirlot, M., and Prade, H. (2006a). Concepts et méthodes pour l'aide à la décision 3 : analyse multicritère (Traité IC2, série Informatique et Systèmes d'Information), volume 3 of Traité IC2, série Informatique et Systèmes d'Information. Lavoisier.
- [Bouyssou et al., 2006b] Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., and Vincke, P. (2006b). Evaluation and Decision models with multiple criteria : stepping stones for the analyst, volume 86. Springer.
- [Butler et al., 1997] Butler, J., Jia, J., and Dyer, J. (1997). Simulation techniques for the sensitivity analysis of multi-criteria decision models. European Journal of Operational Research, 103(3) :531–546.
- [Cailloux, 2012] Cailloux, O. (2012). Elicitation indirecte de modèles de tri multicritère. PhD thesis. Thèse de doctorat dirigée par Mousseau, Vincent Informatique Châtenay-Malabry, Ecole centrale Paris 2012.
- [Carenini and Moore, 2006] Carenini, G. and Moore, J. D. (2006). Generating and evaluating evaluative arguments. Artificial Intelligence, 170(11) :925–952.
- [Cawsey, 1995] Cawsey, A. (1995). Developing an explanation component for a knowledge-based system : Discussion. Expert Systems with Applications, 8(4) :527–531. Explanation : The Way Forward.

- [Daher et al., 2017] Daher, J. B., Brun, A., and Boyer, A. (2017). A Review on Explanations in Recommender Systems. Technical report, LORIA - Université de Lorraine.
- [Fischhoff, 1991] Fischhoff, B. (1991). Value elicitation : Is there anything in there? American Psychologist, 46 :835–847.
- [Fishburn, 1967] Fishburn, P. C. (1967). Letter to the editor—additive utilities with incomplete product sets : Application to priorities and assignments. Operations Research, 15(3) :537–542.
- [Fishburn, 1977] Fishburn, P. C. (1977). Condorcet social choice functions. SIAM Journal on Applied Mathematics, 33(3) :469–489.
- [Fishburn, 1996] Fishburn, P. C. (1996). Finite linear qualitative probability. Journal of Mathematical Psychology, 40(1) :64–77.
- [Fishburn, 1997] Fishburn, P. C. (1997). Cancellation conditions for multiattribute preferences on finite sets. In Karwan, M. H., Spronk, J., and Wallenius, J., editors, Essays In Decision Making : A Volume in Honour of Stanley Zionts, pages 157–167. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Fishburn, 2001] Fishburn, P. C. (2001). Cancellation conditions for finite two-dimensional additive measurement. Journal of Mathematical Psychology, 45(1) :2–26.
- [Fishburn et al., 2002] Fishburn, P. C., Pekec, A., and Reeds, J. A. (2002). subset comparisons for additive linear orders. Mathematics of Operations Research, 27 :227–243.
- [Greco et al., 2011] Greco, S., Kadziński, M., and Słowiński, R. (2011). Selection of a representative value function in robust multiple criteria sorting. Computers & Operations Research, 38(11) :1620–1637.
- [Greco et al., 2008] Greco, S., Mousseau, V., and Słowiński, R. (2008). Ordinal regression revisited : Multiple criteria ranking using a set of additive value functions. European Journal of Operational Research, 191(2) :416–436.
- [Greco et al., 2010a] Greco, S., Mousseau, V., and Słowiński, R. (2010a). Multiple criteria sorting with a set of additive value functions. European Journal of Operational Research, 207(3) :1455–1470.
- [Greco et al., 2010b] Greco, S., Słowiński, R., Figueira, J. R., and Mousseau, V. (2010b). Robust ordinal regression. In Ehrgott, M., Figueira, J. R., and Greco, S., editors, Trends in Multiple Criteria Decision Analysis, pages 241–283. Springer US, Boston, MA.
- [Greco et al., 2013] Greco, S., Słowiński, R., and Zielniewicz, P. (2013). Putting dominance-based rough set approach and robust ordinal regression together. Decision Support Systems, 54(2) :891–903.
- [Greer et al., 1994] Greer, J. E., Falk, S., Greer, K. J., and Bentham, M. J. (1994). Explaining and justifying recommendations in an agriculture decision support system. Computers and Electronics in Agriculture, 11(2) :195–214.

- [Hammond et al., 1998] Hammond, J., Keeney, R., and Raiffa, H. (1998). Even swaps : A rational method for making trade-offs. Harvard business review, 76 :137–8, 143.
- [Ishizaka and Nemery, 2013] Ishizaka, A. and Nemery, P. (2013). Multi-Criteria Decision Analysis : Methods and Software.
- [Jacquet-Lagrange and Siskos, 1982] Jacquet-Lagrange, E. and Siskos, J. (1982). Assessing a set of additive utility functions for multicriteria decision-making, the uta method. European Journal of Operational Research, 10(2) :151–164.
- [Kadzinski et al., 2014] Kadzinski, M., Corrente, S., Greco, S., and Słowiński, R. (2014). Preferential reducts and constructs in robust multiple criteria ranking and sorting. Operations Research-Spektrum, 36 :1021–1053.
- [Kadziński et al., 2012] Kadziński, M., Greco, S., and Słowiński, R. (2012). Selection of a representative value function in robust multiple criteria ranking and choice. European Journal of Operational Research, 217(3) :541–553.
- [Karp, 1972] Karp, R. M. (1972). Reducibility among combinatorial problems. In Miller, R. E., Thatcher, J. W., and Bohlinger, J. D., editors, Complexity of Computer Computations : Proceedings of a symposium on the Complexity of Computer Computations, held March 20–22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, and sponsored by the Office of Naval Research, Mathematics Program, IBM World Trade Corporation, and the IBM Research Mathematical Sciences Department, pages 85–103. Springer US, Boston, MA.
- [Kass and Finin, 1988] Kass, R. and Finin, T. (1988). The need for user models in generating expert system explanations. 1.
- [Keeney and Raiffa, 1976] Keeney, R. and Raiffa, H. (1976). Decisions with multiple objectives : Preferences and value tradeoffs. J. Wiley, New York.
- [Kim and Kim, 1991] Kim, S. H. and Kim, J. K. (1991). Explanation in a decision-theoretic consulting system : An axiomatic approach. Applied Artificial Intelligence, 5(4) :393–409.
- [Klein, 1994] Klein, D. A. (1994). Decision-Analytic Intelligent Systems : Automated Explanation and Knowledge Acquisition. L. Erlbaum Associates Inc., USA.
- [Labreuche, 2011] Labreuche, C. (2011). A general framework for explaining the results of a multi-attribute preference model. Artificial Intelligence, 175(7) :1410–1448. Representing, Processing, and Learning Preferences : Theoretical and Practical Challenges.
- [Labreuche et al., 2011] Labreuche, C., Maudet, N., and Ouerdane, W. (2011). Minimal and complete explanations for critical multi-attribute decisions. In Algorithmic Decision Theory, pages 121–134.
- [Labreuche et al., 2012] Labreuche, C., Maudet, N., and Ouerdane, W. (2012). Justifying dominating options when preferential information is incomplete. In Proceedings of

- the 20th European Conference on Artificial Intelligence, ECAI'12, page 486–491, NLD. IOS Press.
- [Labreuche et al., 2015] Labreuche, C., Maudet, N., Ouerdane, W., and Parsons, S. (2015). A dialogue game for recommendation with adaptive preference models. In Proceedings AAMAS, pages 959–967.
- [Liffiton and Sakallah, 2008] Liffiton, M. H. and Sakallah, K. A. (2008). Algorithms for computing minimal unsatisfiable subsets of constraints. Journal of Automated Reasoning, 40 :1–33.
- [Luce and Raiffa, 1957] Luce, R. D. and Raiffa, H. (1957). Games and Decisions. Wiley, New York.
- [Lunnon, 1988] Lunnon, W. (1988). Integer sets with distinct subset-sums. Mathematics of Computation - Math. Comput., 50 :297–297.
- [Maclagan, 1998] Maclagan, D. (1998). Boolean term orders and the root system B_n . Order, 15 :279–295.
- [March, 1978] March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. The Bell Journal of Economics, 9(2) :587–608.
- [Miller, 1956] Miller, G. A. (1956). The magical number seven, plus or minus two : Some limits on our capacity for processing information. The Psychological Review, 63(2) :81–97.
- [Mousseau et al., 2003] Mousseau, V., Figueira, J., Dias, L., Silva, C., and Clmaco, J. (2003). Resolving inconsistencies among constraints on the parameters of an mcda model. European Journal of Operational Research, 147 :72–93.
- [Nunes et al., 2014] Nunes, I., Miles, S., Luck, M., Barbosa, S., and Lucena, C. (2014). Pattern-based explanation for automated decisions. In Proceedings of the Twenty-First European Conference on Artificial Intelligence, ECAI'14, page 669–674, NLD. IOS Press.
- [Nunes et al., 2012] Nunes, I., Miles, S., Luck, M., and de Lucena, C. J. P. (2012). Investigating explanations to justify choice. In Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization, UMAP'12, page 212–224. Springer-Verlag.
- [Papamichail and French, 2003] Papamichail, K. and French, S. (2003). Explaining and justifying the advice of a decision support system : a natural language generation approach. Expert Systems with Applications, 24(1) :35–48.
- [Pawlak, 1991] Pawlak, Z. (1991). Rough sets - theoretical aspects of reasoning about data, volume 9 of Theory and decision library : series D. Kluwer.
- [Payne et al., 1999] Payne, J. W., Bettman, J. R., and Schkade, D. A. (1999). Measuring constructed preferences : Towards a building code. Journal of Risk and Uncertainty, 19(1-3) :243–70.

- [Raiffa, 1970] Raiffa, H. (1970). Decision Analysis - Introductory lectures on choices under uncertainty. Addison Wesley, Reading, MA. traduction française : Analyse de la décision : introduction aux choix en avenir incertain, Dunod, 1973.
- [Rooij et al., 2005] Rooij, I., Stege, U., and Kadlec, H. (2005). Sources of complexity in subset choice. Journal of Mathematical Psychology, 49 :160–187.
- [Roy, 1996] Roy, B. (1996). Multicriteria Methodology for Decision Aiding. Kluwer Academic, Dordrecht.
- [Roy, 2009] Roy, B. (2009). Pour mieux répondre à la préoccupation de robustesse en aide à la décision : quatre propositions suite à un double constat. working paper or preprint.
- [Scott, 1964] Scott, D. (1964). Measurement structures and linear inequalities. Journal of Mathematical Psychology, 1(2) :233–247.
- [Simonson and Tversky, 1992] Simonson, I. and Tversky, A. (1992). Choice in context : Tradeoff contrast and extremeness aversion. Journal of Marketing Research, 29(3) :281–295.
- [Siskos et al., 2016] Siskos, Y., Grigoroudis, E., and Matsatsinis, N. F. (2016). Uta methods. In Greco, S., Ehrgott, M., and Figueira, J. R., editors, Multiple Criteria Decision Analysis : State of the Art Surveys, pages 315–362. Springer New York.
- [Slade, 1991] Slade, S. (1991). Generating natural language explanations for decisions.
- [Slade, 1992] Slade, S. (1992). Generating explanations for goal-based decision making*. Decision Sciences, 23(6) :1440–1461.
- [Tintarev and Masthoff, 2007] Tintarev, N. and Masthoff, J. (2007). A survey of explanations in recommender systems. In Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop, page 801–810, USA. IEEE Computer Society.
- [Walton, 1996] Walton, D. (1996). Argumentation schemes for Presumptive Reasoning. Mahwah, N. J., Erlbaum.
- [Wulf and Bertsch, 2017] Wulf, D. and Bertsch, V. (2017). A natural language generation approach to support understanding and traceability of multi-dimensional preferential sensitivity analysis in multi-criteria decision making. Expert Systems with Applications, 83 :131–144.
- [Zwicker and Moulin, 2016] Zwicker, W. S. and Moulin, H. (2016). Introduction to the theory of voting. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., and Procaccia, A. D., editors, Handbook of Computational Social Choice, page 23–56. Cambridge University Press.

Annexe A

Algorithme de génération exhaustive de relations d'ordre linéaire additives

L'Algorithme 5 génère et retourne sous la forme d'une liste de fonctions de score ω à m composantes entières ($m \geq 2$), l'ensemble de toutes les relations d'ordre linéaire additives définies sur $2^{[m]}$ et vérifiant un ordre croissant arbitraire σ de préférences sur les alternatives singletons. Il prend en entrées m (le nombre total de critères/items) et σ et énumère tous les rangements R des alternatives de $2^{[m]} \setminus \emptyset$ compatibles avec σ et représentables par une fonction de score ω vérifiant l'Équation 2.1 (Définition de représentation des ROLA). Cette vérification s'effectue en résolvant le programme linéaire en nombres entiers $\mathcal{P}_{gen}(R, \sigma)$ décrit comme suit :

1. Variables : ω_i pour tout $i \in [m]$ de type entier naturel.
2. Contraintes :
 - $\omega_{\sigma(i)} + 1 \leq \omega_{\sigma(i+1)}$ pour tout $i \in \llbracket 1; m-1 \rrbracket$
 - $1 + \sum_{i \in R[k]} \omega_i \leq \sum_{j \in R[k+1]} \omega_j$ pour tout $k \in \llbracket 1; |R| - 1 \rrbracket$
3. Fonction objectif (optionnelle) : $\sum_{i \in [m]} \omega_i$ (à minimiser)

L'implémentation en Python de cet algorithme est accessible sur le dépôt suivant : <https://github.com/tetras92/SchemasDeductifsEnAMCD.git>

Algorithme 5 Génération exhaustive de relations d'ordre linéaire additives

Entrée(s) m et σ

Sortie(s) La liste de toutes les ROLA représentées chacune par une fonction de score.

- 1: CBTO := [] # une liste vide
 - 2: Soient $\sigma(1)$ et $\sigma(2)$ les alternatives singletons de rangs 1 et 2 suivant σ .
 - 3: $L := [\sigma(1), \sigma(2)]$
 - 4: $\bar{L} \leftarrow$ la liste des alternatives de $2^{[m]} \setminus \emptyset$ autres que $\sigma(1)$ et $\sigma(2)$.
 - 5: Soit S , une pile vide.
 - 6: Empiler dans S le couple (L, \bar{L})
 - 7: **tant que** S non vide **faire**
 - 8: $R, \bar{R} \leftarrow$ Dépiler S
 - 9: **si** $|R| \geq 2^{m-1} - 1$ **alors**
 - 10: Soit ω la fonction de score retournée par la résolution du PLNE $\mathcal{P}_{gen}(R, \sigma)$
 - 11: Rajouter ω à CBTO
 - 12: Retourner à la ligne 7
 - 13: **fin du si**
 - 14: **pour** $x \in \bar{R}$ **faire**
 - 15: $\bar{R}_x \leftarrow$ copie de \bar{R} sans x
 - 16: $R_x \leftarrow R + [x]$
 - 17: **si** l'exécution du PLNE $\mathcal{P}_{gen}(R_x, \sigma)$ réussit **alors**
 - 18: Empiler dans S le couple (R_x, \bar{R}_x)
 - 19: **fin du si**
 - 20: **fin du pour**
 - 21: **fin du tant que**
 - 22: **retourner** CBTO
-

