CREDAL NETWORKS UNDER EPISTEMIC IRRELEVANCE

THEORY AND ALGORITHMS

JASPER DE BOCK

ISBN 978-90-8578-795-2
NUR 916, 984
Wettelijk depot: D/2015/10.500/39

PROMOTOR

prof. dr. ir. Gert de Cooman, UGent

LEDEN VAN DE LEESCOMMISSIE

prof. dr. ir. Fabio Cozman, University of São Paulo
prof. dr. Enrique Miranda, University of Oviedo
prof. dr. Serafín Moral, University of Granada
prof. dr. ir. Joris Walraevens, UGent

OVERIGE LEDEN VAN DE EXAMENCOMMISSIE

prof. dr. ir. Dirk Aeyels (secretaris), UGent
prof. dr. Dries Benoit, UGent
prof. dr. ir. Jan Van Campenhout, UGent
prof. dr. ir. Rik Van de Walle (voorzitter), UGent

ADRES

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Elektrische Energie, Systemen en Automatisering
Onderzoeksgroep SYSTeMS

Technologiepark-Zwijnaarde 914
9052 Zwijnaarde
België

FINANCIERING

Dit proefschrift bevat de resultaten van onderzoek gefinancierd door het Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO).

CONTENTS

PREFACE 7

SUMMARY 10

SAMENVATTING (DUTCH SUMMARY) 14

LIST OF SYMBOLS AND TERMINOLOGY 18

1 INTRODUCTION 25

1.1 Motivation and main results 26
1.2 Finding your way around 30
1.3 A brief overview 31
1.4 Publications 33

2 MODELLING UNCERTAINTY 37

2.1 Preliminaries 38
2.2 Sets of desirable gambles 38
2.3 Lower previsions 40
2.4 Linear previsions 50
2.5 Full conditional probability mass functions 51
2.6 Sets of linear previsions or mass functions 52
2.7 Regular extension versus natural extension 59

3 UPDATING 69

3.1 Narrowing down the problem 70
3.2 Conditioning as an updating strategy 73
3.3 What about other updating strategies? 76
3.4 Justifying updating with regular extension 79
3.A Technical results related to updating 91

3.B Proof of Theorem 18 97

4 MULTIVARIATE MODELS 100

4.1 Reducing everything to the univariate case 101
4.2 Marginalisation, conditioning and updating 101
4.3 Epistemic irrelevance 114
4.4 Various other notions of independence 120
4.5 Generalising the notion of natural extension 124

5 CREDAL NETWORKS UNDER EPISTEMIC IRRELEVANCE 128

5.1 Directed Acyclic Graphs 129
5.2 Adding Local models 130
5.3 Judgements of epistemic irrelevance 132
5.4 The irrelevant natural extension 134
5.5 Other approaches and how they are related 145
5.A Proof of Propositions 38 and 39 148

6 THEORETICAL PROPERTIES 152

6.1 Closed sets and their properties 153
6.2 A fundamental separating hyperplane result 154
6.3 Factorisation and external additivity 157
6.4 Conditioning, marginalisation and irrelevance 160
6.5 Separation properties 162
6.6 The independent natural extension 171
6.7 A connection with marginal extension 175
6.8 What if we apply regular extension? 180
6.A Technical lemmas about properties of DAGs 182
6.B Proof of Theorem 53 188
6.C Proof of Proposition 56 199
6.D Proof of Theorem 62 201
6.E Proof of Theorem 64 203
6.F Proof of Proposition 67 206
6.G Proof of Proposition 71 207
6.H Proof of Theorem 73 208

7 INFERENCE ALGORITHMS 211

7.1 Which kinds of inferences do we consider? 213
7.2 Making the problem smaller 216
7.3 Reduction to the unconditional case 219
7.4 Brute force techniques 221

7.5 Recursively decomposable networks 235
7.6 What about other types of networks? 271
7.A Proof of Theorem 88 278
7.B Proof of Theorem 92 280

8 CONCLUSIONS 291

BIBLIOGRAPHY 296

PREFACE

"Gratitude is one of the least articulate of the emotions, especially when it is deep."

Felix Frankfurter

This has been one hell of a ride! To reside at the very edge of knowledge and to push it forward, ever so slowly, has been a brilliant experience and a dream come true. During the past few years, I have been taught this edge-pushing art form that is called research, and have been given the opportunity to exercise it in an atmosphere that has been nothing short of delightful. I've had the pleasure to call this my job, but I have to admit that it often felt like a paid hobby, be it a very time-consuming one. Therefore, I would like to kick off this preface by thanking all of you, my fellow tax-paying citizens, for financially supporting me during these years. I have enjoyed it tremendously. That being said, I have quite a few other people to thank as well.

First and foremost, I would like to thank Gert, for being such a terrific scientific mentor. He taught me the ins and outs of research, ranging from technical mathematics to linguistic style. He introduced me to the material in References [42, 45], shared his views on the bigger picture behind it, and as such provided me with the central ideas from which this dissertation has grown. I learned from him that good research is about vision as much as it is about solid mathematics. He trained me in both, shared his own views on these aspects, and allowed me to develop my own. Our views sometimes clash; beautiful results tend to emerge out of these clashes, which we can then go and celebrate over a nice beer. It has been a true pleasure to do research with him, and I hope to keep on doing so for a long time to come. Besides research, he also initiated and trained me in various other aspects of academia, thereby preparing me for the many challenges that lie ahead. If I am ever given the chance to stay in academia, I owe it to him, and for that, I cannot thank him enough. Finally, and most importantly, I would like to thank him for being such a close friend.

I conducted the research that led to this dissertation in a wonderful environment, and I am grateful to the people that created it. On a small scale, these have been the members of the SYSTeMS research group of Ghent University. I ride my bike to work with a smile, almost every single day, and I owe that to my colleagues. For that, I would like to thank them. A special place is reserved for Gert, Arthur and Stavros. I will never forget, and would like to thank them for, the memorable road trips we made, the lovely dinners we had, the many laughs we shared, and their enthusiasm at the blackboard. On a larger scale, these have been the members of SIPTA—the Society for Imprecise Probability: Theories and Applications. I thank them for creating such a welcoming international community, full of friendly, helpful and inspiring people, and for making conferences feel like holidays. Perhaps you won't believe me, but mathematicians can be a lot of fun. A special thanks goes out to Alessandro, Cassio, Enrique, Frank, Marco and Matthias, for their hospitality, and for being such lovely people.

Writing this dissertation has been an exercise in perseverance. I have spent countless hours on it, at the cost of many a good night's sleep. I have definitely squeezed myself to a point that can hardly be called healthy, the sole goal being the book in front of you. Although I hope you are enjoying this preface, this book is ultimately about the content that is about to follow, and its main reason of existence is to allow a jury to decide whether or not my research is of sufficient merit. This decision, and the doors it can open, has been the ultimate reason for the squeezing I referred to above. I would therefore like to express my sincere gratitude towards the people that took it upon themselves to make this decision—favourably, to my great delight. To the members of the examination board and the reading committee of this dissertation: thank you for taking the time to thoroughly assess my research, for your constructive criticism, for your thought-provoking questions, and for giving me the opportunity to defend this dissertation in front of my friends and family.

Apart from being the culmination of four years of intense research, this book also marks the end of my education, and the beginning of a new phase in my life. That I have made it this far, and that I have enjoyed it this much, is also—and mainly—due to a number of people that I have not yet mentioned in this preface, for the sole reason that they were not directly involved in the creation of this dissertation. Nevertheless, I would like to thank them here, for all the things they have done for me.

My friends, for long talks about things that matter, for all those relaxing days out, for the joy and piece of mind that only dancing can provide, for their honest opinion, for their help with all sorts of little things, for beer in good company—often very late at night—and for creating the buzzing sound of happiness that convinces me, time and time again, that it is all worthwhile. Guy, for his inspiring view on life and his everlasting enthusiasm, and Kwinten and Evelien, for their heartwarming support during difficult times, when I most needed it.

My family, for the warm cosy feeling they create on various occasions, and for their support and help, during different steps in my life, ranging from the first very little ones, to the big ones further on. My parents, for raising me, for teaching me the values that I now try to live by, and for their financial support during all those years. My father, for passing on his love for science, both genetically and personally, and for teaching me to walk through life with a critical eye, thereby significantly increasing my chances for survival in this jungle that we call society. My mother, for her unconditional love, for the countless hours she spent driving me around, for her lovely cooking and the delicious dinners that resulted from it, for her helping hand whenever needed, and for teaching me to enjoy the little things in life.

Annelien, the love of my life, for taking care of me when this dissertation started to take up far more time than it ever should have and for her unwavering support during the final months of writing. I also want to thank her for the wonderful past seven years, for being such a warm and caring person, for sharing my love for food and all the other great things life has to offer, for expanding my horizon, and for all those mazurka's. It makes me a very happy man that she has chosen to share her life with me, in days of all sorts. May our lives remain intertwined for a long time to come, filled with plenty of good old-fashioned love. Liefde is niet alles, maar zonder liefde is alles niets.

SUMMARY

This dissertation presents a detailed study of credal networks under epistemic irrelevance [16], mathematical models that are capable of representing compactly and intuitively the uncertainty that is associated with complex multivariate domains and that can subsequently be used to answer various domain-specific queries of interest to the user. These models share many of the nice features of Pearl's celebrated Bayesian networks [78], but have the added advantage that they can represent uncertainty in a more flexible and realistic way.

We model uncertainty using imprecise probabilities [106], the underlying theory of which is an extension of probability theory that can represent model uncertainty and probabilistic uncertainty within a single framework. Simply put, whenever it is infeasible to reliably estimate a single probability, this theory allows for the use of a set of probabilities instead, each of whose elements is regarded as a candidate for some ideal 'true' probability. However, this simplified view is only one of the many ways to look at or interpret imprecise probabilities. Uncertainty can also be expressed without any reference to probabilities, using other imprecise-probabilistic frameworks such as sets of desirable gambles, lower previsions and sets of linear previsions. This dissertation starts with a detailed overview of these different frameworks, their interpretation, and how they are connected to each other. We pay special attention to conditional models, which we regard as primitive concepts whose connection with unconditional models is established by means of rationality criteria. The main advantage of the resulting so-called full conditional models is that they do not suffer from the traditional problems that arise when some of the conditioning events have probability zero. This is especially important in the context of imprecise probabilities, where probability zero cannot be ignored because it may easily happen that an event has lower probability zero but positive upper probability. Of course, even if we regard conditional models as primitive concepts, they may not be readily available, or easy to assess. In that case, we often have no choice but to derive them from an unconditional model. In an imprecise-probabilistic setting, two conditioning rules are commonly used for this purpose: regular extension and natural extension. We explain the difference between them and discuss various technical and computational aspects.

Although this overview of imprecise probability theory contains new re-

sults that fill some gaps in the literature, its contribution mainly consists in bringing together results from various existing frameworks and connecting them to each other. The first real contribution of this dissertation is our discussion of updating, which is the act of changing a model based on the information that some event has occurred. In probability theory, it has become standard practice to do this by conditioning on that event using Bayes's rule. Similarly, in an imprecise-probabilistic setting, updating is typically performed by applying a conditioning rule such as regular or natural extension. However, little argumentation is usually given as to why such an approach would make sense. We help address this problem by providing a firm philosophical justification for using natural and regular extension as updating rules. What makes our justification especially powerful is that we derive it directly in terms of sets of desirable gambles. In this way, we avoid making some of the unnecessarily strong assumptions that are traditionally adopted, such as the existence of an ideal 'true' probability mass function.

In order to apply the theory of imprecise probabilities in a multivariate context, we need additional tools such as marginalisation and ways of combining these tools with concepts such as conditioning and updating. This is all well known and relatively easy in terms of probabilities, but it becomes more challenging for some of the imprecise-probabilistic frameworks that we consider. We have gathered the existing tools, have added new ones whenever something was missing and have connected all of them with one another. The result is a complete and well-founded theory of multivariate imprecise probabilities that is, to the best of our knowledge, novel in its completeness, generality and consistency. It allows us to formally introduce one of the most important concepts of this dissertation: epistemic irrelevance, which is an asymmetric imprecise-probabilistic notion of independence. We discuss several existing definitions for this notion, argue why only one of them is really adequate, and compare epistemic irrelevance to other imprecise-probabilistic independence notions. Finally, we explain how the concept of conservative reasoning allows us to combine structural assessments such as epistemic irrelevance with direct or local partial probability assessments to construct a multivariate uncertainty model.

The rest of this dissertation is concerned with one particular type of multivariate model: the irrelevant natural extension of a credal network under epistemic irrelevance. The basic idea is very similar to that of a Bayesian network. The starting point is a collection of domain-specific variables that are connected by means of arrows that reflect how these variables depend on each other. The arrows form a Directed Acyclic Graph (DAG), which simply means that there are no (directed) cycles. The interpretation of the graph is that for any variable, conditional on its parents, its non-parent non-descendants are epistemically irrelevant. For each of the variables, we are given a local imprecise-probabilistic model conditional on the values of its parents in the graph. In combination with the assessments of epistemic irrelevance that correspond to

the DAG, these local models form a credal network under epistemic irrelevance. The most conservative global uncertainty model that is compatible with all these assessments is called the irrelevant natural extension of the network. This concept was first introduced by Cozman [16], who defined it in terms of sets of probabilities under the simplifying assumption that all probabilities are strictly positive. We drop this positivity assumption and provide definitions in terms of three other frameworks as well: sets of desirable gambles, lower previsions and sets of linear previsions. These different definitions turn out to be closely related, which allows us to translate results that are proved in one framework to analogous results in other frameworks.

Credal networks under epistemic irrelevance are not the only imprecise-probabilistic generalisations of Bayesian networks. In fact, they are not even all that popular. At the present moment, most authors prefer to consider credal networks under strong independence, the difference being that the assessments of epistemic irrelevance that we make, are replaced by assessments of strong independence, which is another imprecise-probabilistic notion of independence. We believe that the main reason for this lack of popularity is a profound lack of known theoretical properties. This has severely inhibited the development of tractable inference algorithms, where, simply put, inference is intended as computing lower and upper updated probabilities. In fact, there is currently only one inference algorithm available, and even then, only for a particular type of inference and for networks whose DAG has a tree structure [42]. Nevertheless, due to the remarkable efficiency of this particular algorithm, which is linear in the size of the network, and because that same inference problem is NP-hard in credal networks under strong independence [67], credal networks under epistemic irrelevance are regarded as a promising alternative that requires—and deserves—further research [2, Section 10.6]. This further research is what this dissertation is all about.

One of our main contributions is a detailed study of the theoretical properties of the multivariate uncertainty model that corresponds to a credal network under epistemic irrelevance: the irrelevant natural extension. By focusing on the framework of sets of desirable gambles, we are able to derive some remarkable properties of this model, which we then manage to translate to other frameworks as well. A first important example is a fundamental separating hyperplane result that establishes a connection between the irrelevant natural extension of a complete network and that of its subnetworks. This result leads to various marginalisation, factorisation and external additivity properties. A second important result is that the irrelevant natural extension satisfies a collection of epistemic irrelevancies that is induced by AD-separation, an asymmetric adaptation of d-separation that is proved to satisfy all graphoid properties except symmetry. We also establish connections with the notions of independent natural extension and marginal extension and study the updated models that are obtained by applying regular extension to the irrelevant natural extension.

In the final part of this dissertation, we show how the theoretical properties that we have proved can be used to develop efficient inference algorithms for credal networks under epistemic irrelevance. A first important contribution consists of two preprocessing techniques that allow us to simplify inference problems before the actual algorithm is applied. We explain how and when it is possible to translate an inference problem in a large network into a similar problem in a smaller network, and show how solving a conditional inference problem can be reduced to solving a series of unconditional ones. In a second set of results, we rephrase inference as a linear optimisation problem. As was already mentioned by Cozman [16], every unconditional inference can be computed by solving a linear program. However, in order to establish this result, he required a simplifying positivity assumption. We show that this positivity assumption is not needed; unconditional inferences can always be characterised as the solution of a linear program. For conditional inferences, multiple such linear programs need to be solved. Unfortunately, the size of these linear programs is exponential in the size of the network and this in principle generally applicable method is therefore only tractable for small networks. For the specific case of a network that consists of two disconnected binary variables, we are able to solve the corresponding linear program symbolically. In this way, we obtain closed-form expressions for the extreme points of the independent natural extension of two binary models. Fortunately, the intractability of brute force linear programming methods can often be circumvented by developing other, more efficient and often recursive computational techniques. We illustrate this by means of a number of examples. Our most important contribution, and the proverbial icing on the cake, is a collection of recursive algorithms that can efficiently compute various inferences in credal networks under epistemic irrelevance whose graphical structure is a recursively decomposable DAG, a new type of DAG that includes trees as a special case.

SAMENVATTING

Dutch summary

Dit proefschrift legt een gedetailleerde studie voor van credale netwerken onder epistemische irrelevantie [16], wiskundige modellen die de onzekerheid die gepaard gaat met complexe multivariate domeinen compact en intuitief kunnen voorstellen en vervolgens in staat zijn om allerhande domein-specifieke vragen te beantwoorden die relevant zijn voor de gebruiker. Deze modellen delen veel van de gevierde eigenschappen van Pearls Bayesiaanse netwerken [78], en hebben het bijkomende voordeel dat ze onzekerheid flexibeler en realistischer kunnen voorstellen.

We modelleren onzekerheid met imprecieze waarschijnlijkheden, waarvan de onderliggende theorie een uitbreiding is van waarschijnlijkheidsleer die zowel modelonzekerheid als probabilistische onzekerheid in één kader kan voorstellen. Eenvoudig gesteld, telkens als een waarschijnlijkheid niet betrouwbaar geschat kan worden, dan maakt deze theorie het mogelijk om deze enkele waarschijnlijkheid te vervangen door een verzameling waarschijnlijkheden, waarvan elk van de elementen beschouwd wordt als een kandidaat voor de ideale 'correcte' waarschijnlijkheid. Dit is echter maar één van de vele manieren waarop imprecieze waarschijnlijkheden kunnen geïnterpreteerd worden. Onzekerheid kan ook beschreven worden in andere imprecieze waarschijnlijkheidskaders, zoals verzamelingen van begeerlijke gokken, onderprevisies en verzamelingen van lineaire previsies. Dit proefschrift begint met een gedetailleerd overzicht van deze verschillende kaders, hun interpretatie, en hoe ze verbonden zijn. We besteden speciale aandacht aan conditionele modellen, die we als primitieve concepten beschouwen. Rationaliteitscriteria verbinden deze met onconditionele modellen. Het belangrijkste voordeel van deze zogenaamde volledige conditionele modellen is dat ze de traditionele problemen vermijden van het conditioneren op gebeurtenissen met waarschijnlijkheid nul. Dit is extra belangrijk in de context van imprecieze waarschijnlijkheid omdat het vaak voorvalt dat een gebeurtenis met onderwaarschijnlijkheid nul een positieve bovenwaarschijnlijkheid heeft. Conditionele modellen als primitieve concepten beschouwen verhelpt natuurlijk niet het probleem dat deze vaak niet direct beschikbaar, of moeilijk te bepalen kunnen zijn. In dat geval is het nodig om ze af te leiden van onconditionele modellen. Binnen de imprecieze

waarschijnlijkheidsleer zijn er twee conditioneringsregels die hiervoor vaak gebruikt worden: reguliere en natuurlijke extensie. We leggen het verschil uit tussen beide en bespreken allerhande technische en computationele aspecten.

Hoewel dit overzicht van imprecieze waarschijnlijkheidsleer al enkele nieuwe resultaten bevat die de bestaande literatuur waar nodig aanvullen, draagt het vooral bij in de manier waarop we de verschillende bestaande kaders samenbrengen en met elkaar verbinden. De eerste echte bijdrage van dit proefschrift is onze bespreking van updaten: het aanpassen van een onzekerheidsmodel na geïnformeerd te zijn dat een bepaalde gebeurtenis is opgetreden. In waarschijnlijkheidsleer is conditioneren met behulp van de regel van Bayes de standaardmethode om dit probleem aan te pakken. Analoog, binnen de imprecieze waarschijnlijkheidsleer, worden conditioneringsregels zoals reguliere en natuurlijke extensie vaak als updateregels beschouwt. Dit gebruik van conditioneren als een updatemethode wordt echter vaak niet beargumenteerd en meestal als vanzelfsprekend beschouwd. We verhelpen dit probleem door een filosofische verantwoording te ontwikkelen voor het gebruik van reguliere en natuurlijke extensie als updateregels. Wat onze verantwoording bijzonder sterk maakt is dat we ze afleiden in het kader van verzamelingen van begeerlijke gokken en zo traditionele—maar onnodige—aannames zoals het bestaan van een ideale 'correcte' waarschijnlijkheid vermijden.

Om imprecieze waarschijnlijkheidsleer toe te passen in een multivariate context hebben we bijkomende technieken nodig, zoals marginalisatie, en methodes om die te combineren met concepten zoals conditioneren en updaten. Dit is allemaal vrij voor de hand liggend voor waarschijnlijkheden, maar wordt heel wat uitdagender voor sommige van de imprecieze waarschijnlijkheidskaders die we beschouwen. We brengen de verschillende bestaande technieken samen, voegen er nieuwe aan toe waar nodig en verbinden dit alles met elkaar. Het resultaat is een goed gefundeerde theorie van multivariate imprecieze waarschijnlijkheidsleer die, voor zover wij weten, nieuw is in haar volledigheid, algemeenheid en samenhang. Deze theorie stelt ons in staat om één van de belangrijkste concepten van dit proefschrift formeel in te voeren: epistemische irrelevantie, een asymmetrische notie van onafhankelijkheid voor imprecieze waarschijnlijkheden. We bespreken verschillende bestaande definities ervan, beargumenteren waarom slechts één van ze adequaat is, en vergelijken epistemische irrelevantie met andere noties van onafhankelijkheid voor imprecieze waarschijnlijkheden. Tot slot leggen we ook uit hoe het concept van conservatief redeneren ons in staat stelt structurele aannames zoals epistemische irrelevantie te combineren met lokale partiële informatie over waarschijnlijkheden, om zo een multivariaat onzekerheidsmodel op te bouwen.

Het vervolg van dit proefschrift behandelt één specifiek type multivariaat onzekerheidsmodel: de irrelevante natuurlijke extensie van een credaal netwerk onder epistemische irrelevantie. Het basisidee is gelijkaardig aan dat van een Bayesiaans netwerk. Het vertrekpunt is een collectie van domeinspecifieke veranderlijken die verbonden zijn door pijlen die uitdrukken hoe

deze veranderlijken van elkaar afhangen. De pijlen vormen een Gerichte Acyclische Graaf (GAG), wat simpelweg betekent dat ze geen (gerichte) cycli vormen. De interpretatie van de graaf is dat voor elke veranderlijke, conditioneel op haar ouders in de graaf, de veranderlijken die geen ouder of afstammeling zijn epistemisch irrelevant zijn voor deze veranderlijke. Voor elk van de veranderlijken is er ook een lokaal imprecies waarschijnlijkheidsmodel gegeven, conditioneel op de waarde van haar ouders. In combinatie met de aannames van irrelevantie die overeenstemmen met de graaf vormen deze lokale modellen een credaal netwerk onder epistemische irrelevantie. Het meest conservatieve onzekerheidsmodel dat compatibel is met deze irrelevanties en lokale modellen noemen we de irrelevante natuurlijke extensie van het netwerk. Dit concept werd ingevoerd door Cozman [16], die het definiëerde voor verzamelingen van waarschijnlijkheden, onder de vereenvoudigende aanname dat alle waarschijnlijkheden strikt positief zijn. We laten deze aanname vallen en geven ook definities voor drie andere imprecieze waarschijnlijkheidskaders: verzamelingen van begeerlijke gokken, onderprevisies en verzamelingen van lineaire previsies. We tonen aan dat deze verschillende definities nauw met elkaar verbonden zijn, wat ons in staat stelt om de resultaten in dit proefschrift vlot te vertalen van het ene naar het andere kader.

Credale netwerken onder epistemische irrelevantie zijn niet de enige veralgemening van Bayesiaanse netwerken binnen de imprecieze waarschijnlijkheidsleer. Ze zijn niet eens zo populair. Op dit moment verkiezen de meeste auteurs om met credale netwerken onder sterke onafhankelijkheid te werken, waarbij het verschil is dat de aannames van epistemische irrelevantie die wij opleggen, vervangen worden door aannames van sterke onafhankelijkheid, een andere notie van onafhankelijkheid voor imprecieze waarschijnlijkheden. De hoofdreden voor dit gebrek aan populariteit is volgens ons een fundamenteel gebrek aan gekende theoretische eigenschappen. Dit heeft het bijna onmogelijk gemaakt om efficiënte inferentie-algoritmen te ontwikkelen, waar inferentie min of meer gelijk staat aan het berekenen van geüpdatete onder- en bovenwaarschijnlijkheden. Er is maar één efficiënt inferentie-algoritme beschikbaar en dat kan maar één specifiek soort inferenties berekenen in netwerken waarvan de GAG een boomstructuur heeft [42]. En toch, door de opmerkelijke efficiëntie van dit algoritme, en omdat dat specifieke inferentieprobleem NP-moeilijk is in credale netwerken onder sterke onafhankelijkheid [67], worden credale netwerken onder epistemische irrelevantie beschouwd als een veelbelovend alternatief dat verder onderzoek vereist en verdient [2, paragraaf 10.6]. We verrichten dit onderzoek in dit proefschrift.

Eén van de belangrijkste bijdragen van dit proefschrift is een gedetailleerde studie van de theoretische eigenschappen van het multivariate onzekerheidsmodel dat overeenstemt met een credaal netwerk onder epistemische irrelevantie: de irrelevante natuurlijke extensie. Door te werken in het kader van verzamelingen van begeerlijke gokken zijn we er in geslaagd enkele opmerkelijke eigenschappen af te leiden, die we vervolgens ook naar andere imprecieze

waarschijnlijkheidskaders hebben vertaald. Een eerste belangrijk voorbeeld is een scheidend-hypervlakstelling die een verband legt tussen de irrelevante natuurlijke extensie van een compleet netwerk en die van zijn deelnetwerken. Dit leidt tot verscheidene marginalisatie-, factorisatie- en externe additiviteitseigenschappen. Een tweede belangrijk resultaat is dat de irrelevante natuurlijke extensie aan een verzameling van epistemische irrelevanties voldoet die geïnduceerd is door AD-scheiding, een asymmetrische versie van d-scheiding waarvan we aantonen dat ze aan alle grafoïde eigenschappen voldoet, behalve symmetrie. We leggen ook enkele verbanden met de noties van onafhankelijke natuurlijke extensie en marginale extensie en bestuderen de geüpdatete modellen die verkregen worden door reguliere extensie toe te passen op de irrelevante natuurlijke extensie.

In het laatste deel van dit proefschrift tonen we hoe de bewezen theoretische eigenschappen gebruikt kunnen worden om efficiënte inferentiealgoritmen te ontwikkelen voor credale netwerken onder epistemische irrelevantie. Een eerste belangrijke bijdrage bestaat uit twee technieken die kunnen gebruikt worden om een inferentieprobleem te vereenvoudigen voor een inferentie-algoritme wordt toegepast. We leggen uit hoe en wanneer het mogelijk is om een inferentieprobleem in een groot netwerk om te vormen tot een gelijkaardig inferentieprobleem in een kleiner netwerk, en tonen hoe conditionele inferentie kan gereduceerd worden tot het oplossen van een reeks onconditionele inferentieproblemen. In een tweede groep van resultaten herfraseren we inferentie als een lineair optimalisatieprobleem. Zoals al vermeld werd door Cozman [16] kan elke onconditionele inferentie berekend worden met lineaire programmeertechnieken. Om dit resultaat te verkrijgen maakte hij een vereenvoudigende positiviteitsaanname. We tonen dat deze positiviteitsaanname niet noodzakelijk is; onconditionele inferenties kunnen altijd gekarakteriseerd worden als de oplossing van een lineair programma. Voor conditionele inferenties moeten er meerdere zulke lineaire programma's opgelost worden. Jammer genoeg is de grootte van deze lineaire programma's exponentieel in het aantal veranderlijken van het netwerk en deze in principe algemeen toepasbare techniek is daarom enkel computationeel efficiënt voor kleinere netwerken. Voor een netwerk dat bestaat uit twee niet verbonden binaire veranderlijken hebben we het corresponderende lineaire programma symbolisch opgelost, en zo een expliciete uitdrukking afgeleid voor de extreme punten van de onafhankelijke natuurlijke extensie van twee binaire modellen. Gelukkig kan de computationele inefficiëntie van de aanpak met lineaire programma's vaak omzeild worden door andere, efficiëntere en vaak recursieve technieken te ontwikkelen. We illustreren dit met enkele voorbeelden. Onze belangrijkste bijdrage, en de spreekwoordelijke kers op de taart, is een verzameling van recursieve algoritmen die op efficiënte wijze allerhande inferenties kunnen uitrekenen in credale netwerken onder epistemische irrelevantie waarvan de grafische structuur een recursief decomposeerbare GAG is, een nieuw type GAG dat bomen als een speciaal geval omvat.

LIST OF SYMBOLS AND TERMINOLOGY

This list of symbols and terminology is ordered topically. The locations we provide correspond to a definition, or to a first or important use. We do not list symbols that are used only locally. For symbols that have many variations, we list a generic version and/or focus on the most important ones.

NUMBER SETS

Symbol	Meaning	Location
$\mathbb{R}$	Set of real numbers	Page 38
$\mathbb{R}_{>0}$	Set of positive real numbers	Page 38
$\mathbb{R}_{\geq 0}$	Set of non-negative real numbers	Page 39
$\mathbb{N}$	Set of natural numbers: $\{1,2,3,\dots\}$	Page 39
$\mathbb{N}_0$	Set of natural numbers with zero: $\mathbb{N}\cup\{0\}$	Page 39

EVENTS AND GAMBLES

Symbol	Meaning	Location
X	Variable	Page 37
Ω	State space	Page 37
ω	Element of Ω	Page 37
B, O	Events: subsets of Ω	Section 2.1_{38}
$\mathscr{P}(\Omega)$	Set of all events	Section 2.1_{38}
$\mathscr{P}_{\emptyset}(\Omega)$	Set of all non-empty events: $\mathscr{P}(\Omega)\setminus\{\emptyset\}$	Section 2.1_{38}
$\mathbb{I}_B$	Indicator of B	Section 2.1_{38}
f, g, h	Gambles	Section 2.1_{38}
$\mathscr{G}(\Omega)$	Set of all gambles on Ω	Section 2.1_{38}

$\mathscr{G}(\Omega)_{>0}$	Set of all pointwise non-negative gambles on Ω, excluding zero	Section 2.1_{38}
$\mathbb{I}_B f$	Gamble that coincides with f on B and which is zero outside of B	Section $2.2.2_{40}$
$\mathscr{C}(\Omega)$	Set of all couples (f,B), with $B \in \mathscr{P}_\emptyset(\Omega)$ and $f \in \mathscr{G}(B)$	Section $2.3.1_{42}$
$\mathscr{C}$	Subset of $\mathscr{C}(\Omega)$	Section $2.3.1_{42}$
$\mathscr{C}_B$	Set of all gambles $f \in \mathscr{G}(B)$ for which (f,B) is an element of $\mathscr{C}$	Section $2.3.1_{42}$
$\mathscr{C}_*(\Omega)$	Set of all couples (ω,B), with $B \in \mathscr{P}_\emptyset(\Omega)$ and $\omega \in B$	Section 2.5_{51}

BASIC UNCERTAINTY MODELS

Symbol	Meaning	Location
$\mathscr{D}$	Set of desirable gambles	Section 2.2_{38}
$\mathscr{A}$	Assessment of desirable gambles	Section $2.2.1_{39}$
$\mathscr{E}(\mathscr{A})$	Natural extension of $\mathscr{A}$	Section $2.2.1_{39}$
$\mathscr{D}\rfloor B$	Set of desirable gambles, conditional on B	Section $2.2.2_{40}$
$\underline{P}(f)$	Lower prevision of f	Section 2.3_{40}
$\overline{P}(f)$	Upper prevision of f	Section 2.3_{40}
$\underline{P}(B)$	Lower probability of B	Section $2.3.5_{49}$
$\overline{P}(B)$	Upper probability of B	Section $2.3.5_{49}$
$\underline{P}(\cdot)$	Unconditional lower prevision	Section 2.3_{40}
$\underline{P}(\cdot\rfloor\cdot)$	Conditional lower prevision	Section $2.3.1_{42}$
$P(\cdot)$	Unconditional linear prevision	Section 2.4_{50}
$P(\cdot\rfloor\cdot)$	Conditional linear prevision	Section 2.4_{50}
$\mathbb{P}_B$	All linear previsions on $\mathscr{G}(B)$	Section 2.4_{50}
$\mathbb{P}$	All conditional linear previsions on $\mathscr{C}(\Omega)$	Section 2.4_{50}
$p(\cdot\rfloor\cdot)$	Full conditional probability mass function	Section 2.5_{51}
$\mathscr{M}$	Set of (conditional) linear previsions	Section $2.6.4_{57}$
$\mathscr{M}\rfloor B$	Conditional set of linear previsions	Section $2.6.4_{57}$
$\mathscr{F}$	Set of (full conditional) probability mass functions	Section $2.6.4_{57}$
$\mathscr{F}\rfloor B$	Conditional set of probability mass functions	Section $2.6.4_{57}$

$\mathscr{D}_{\underline{P}}$	Smallest set of desirable gambles that corresponds to an unconditional lower prevision $\underline{P}$	Section 2.3_{40}
$\underline{P}_{\mathscr{D}}$	(Conditional) lower prevision that corresponds to a set of desirable gambles $\mathscr{D}$	Section 2.3_{40}
$\underline{P}_{\mathscr{M}}$	(Conditional) lower prevision that corresponds to a set of (conditional) linear previsions $\mathscr{M}$	Section $2.6.4_{57}$
$\mathscr{M}_{\underline{P}}$	Set of (conditional) linear previsions that corresponds to a (conditional) lower prevision $\underline{P}$	Section $2.6.1_{52}$
$\mathscr{M}_{\mathscr{F}}$	Set of (conditional) linear previsions that corresponds to a set of (full conditional) probability mass functions $\mathscr{F}$	Section $2.6.4_{57}$
$\mathscr{F}_{\underline{P}}$	Set of (full conditional) probability mass functions that corresponds to a (conditional) lower prevision $\underline{P}$	Section $2.6.2_{55}$
$\mathscr{F}_{\mathscr{M}}$	Set of (full conditional) probability mass functions that corresponds to a set of (conditional) linear previsions $\mathscr{M}$	Section $2.6.4_{57}$
$\underline{E}$	Natural extension of a (conditional) lower prevision	Sections $2.3.2_{44}$ and $2.7.2_{61}$
$\underline{R}$	Regular extension of a conditional lower prevision	Sections $2.7.2_{61}$ and $3.4.7_{90}$
$\underline{R}^*$	Regular extension of the unconditional part of a lower prevision	Section $3.4.7_{90}$

MULTIVARIATE CONCEPTS

Symbol	Meaning	Location
X_s	A single variable	Section 4.1_{101}
$\mathscr{X}_s$	State space of X_s	Section 4.1_{101}
x_s, z_s	Generic elements of $\mathscr{X}_s$	Section 4.1_{101}
X_S	A tuple of variables X_s, $s \in S$	Section 4.1_{101}
$\mathscr{X}_S$	State space of X_S: $\times_{s \in S} \mathscr{X}_S$	Section 4.1_{101}
x_S, z_S	Generic elements of $\mathscr{X}_S$	Section 4.1_{101}

$\mathrm{marg}_S(\mathscr{D}_G \rfloor B_I)$	Marginal set of desirable gambles for X_S, conditional on B_I	Section 4.2.3$_{105}$
$\underline{P}_S(\cdot \rfloor B_I)$	Marginal lower prevision for X_S, conditional on B_I	Section 4.2.3$_{105}$
$\underline{P}_S(\cdot \rfloor \cdot \times B_I)$	Marginal conditional lower prevision for X_S, conditional on B_I	Section 4.2.4$_{109}$
$P_S(\cdot \rfloor B_I)$	Marginal linear prevision for X_S, conditional on B_I	Section 4.2.3$_{105}$
$P_S(\cdot \rfloor \cdot \times B_I)$	Marginal conditional linear prevision for X_S, conditional on B_I	Section 4.2.4$_{109}$
$p_S(\cdot \rfloor B_I)$	Marginal probability mass function for X_S, conditional on B_I	Section 4.2.3$_{105}$
$p_S(\cdot \rfloor \cdot \times B_I)$	Marginal full conditional probability mass function for X_S, conditional on B_I	Section 4.2.4$_{109}$
$\mathrm{marg}_S(\mathscr{M}_G \rfloor B_I)$	Marginal set of linear previsions for X_S, conditional on B_I	Section 4.2.3$_{105}$
$\mathrm{marg}_S(\mathscr{F}_G \rfloor B_I)$	Marginal set of probability mass functions for X_S, conditional on B_I	Section 4.2.3$_{105}$
$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G \rfloor \rfloor B_I)$	Marginal set of conditional linear previsions for X_S, conditional on B_I	Section 4.2.4$_{109}$
$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_G \rfloor \rfloor B_I)$	Marginal set of full conditional probability mass functions for X_S, conditional on B_I	Section 4.2.4$_{109}$

GRAPHS AND NETWORKS

Terminology that has no symbol	Location
Directed Acyclic Graph (DAG)	Section 5.1$_{129}$
Closed set	Section 6.1$_{153}$
Ancestral set	Section 6.1$_{153}$
Sub-DAG	Section 6.1$_{153}$
Induced DAG	Section 7.5.1$_{238}$
Recursively decomposable DAG	Section 7.5.1$_{238}$
Recursively decomposable credal network	Section 7.5.2$_{246}$
Comparable set	Section 7.5.4$_{255}$

Symbol	Meaning	Location
$s \to t$	There is an edge from s to t	Section 5.1_{129}
$s \sqsubseteq t$	s precedes t	Section 5.1_{129}
$s \sqsubset t$	s strictly precedes t	Section 5.1_{129}
$P(s)$	Parents of a node s	Section 5.1_{129}
$C(s)$	Children of a node s	Section 5.1_{129}
$D(s)$	Descendants of a node s	Section 5.1_{129}
$PN(s)$	Non-parents non-descendants of a node s	Section 5.1_{129}
$N(s)$	Non-descendants of a node s	Section 5.1_{129}
$A(s)$	Ancestors of a node s	Section 5.1_{129}
$P(S)$	Parents of a set of nodes S	Section 6.1_{153}
...	...	...
$P_K(s)$	Parents of a node s, with respect to a sub-DAG with nodes K.	Section 6.1_{153}
...	...	...
$P_K(S)$	Parents of a set of nodes S, with respect to a sub-DAG with nodes K.	Section 6.1_{153}
...	...	...
$Ro(G)$	Roots of a DAG with nodes G	Section 5.1_{129}
$s \parallel t$	s and t are incomparable	Section $7.5.1_{238}$
$\tilde{P}(s)$	Induced parents of s	Section $7.5.1_{238}$
$\tilde{C}(s)$	Induced children of s	Section $7.5.1_{238}$
$\tilde{D}(s)$	Induced descendants of s	Section $7.5.1_{238}$
K_s	s and its descendants: $D(s) \cup \{s\}$	Section $7.5.1_{238}$
AD	Asymmetric D-separation (AD-separation)	Definition 3_{164}

CREDAL NETWORKS UNDER EPISTEMIC IRRELEVANCE

Symbol	Meaning	Location
$\mathscr{D}_{s\rfloor x_{P(s)}}$	Local set of desirable gambles	Section 5.2_{130}
$\underline{P}_{s\rfloor x_{P(s)}}$	Local lower prevision	Section 5.2_{130}
$\mathscr{M}_{s\rfloor x_{P(s)}}$	Local set of linear prevision	Section 5.2_{130}
$\mathscr{F}_{s\rfloor x_{P(s)}}$	Local credal set	Section 5.2_{130}

$\mathscr{D}_G^{\mathrm{irr}}$	Irrelevant natural extension of a network, in the framework of sets of desirable gambles	Section 5.4.1$_{135}$
$\underline{P}_G^{\mathrm{irr}}$	Irrelevant natural extension of a network, in the framework of lower previsions	Section 5.4.2$_{139}$
$\mathscr{M}_G^{\mathrm{irr}}$	Irrelevant natural extension of a network, in the framework of linear previsions	Section 5.4.3$_{141}$
$\mathscr{F}_G^{\mathrm{irr}}$	Irrelevant natural extension of a network, in the framework of probability mass functions.	Section 5.4.4$_{143}$
$\mathscr{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}$	Irrelevant natural extension of a subnetwork with nodes K, in the framework of sets of desirable gambles	Section 6.2$_{154}$
$\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}$	Irrelevant natural extension of a subnetwork with nodes K, in the framework of lower previsions	Section 6.3$_{158}$
$\mathscr{M}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}$	Irrelevant natural extension of a subnetwork with nodes K, in the framework of sets of linear previsions	Section 6.3$_{158}$
$\mathscr{F}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}$	Irrelevant natural extension of a subnetwork with nodes K, in the framework of probability mass functions	Section 6.3$_{158}$
$\underline{R}_G^{\mathrm{irr}}$	Regular extension of $\underline{P}_G^{\mathrm{irr}}$	Section 6.8$_{180}$
$\underline{E}_G^{\mathrm{irr}*}$	Natural extension of the unconditional part of $\underline{P}_G^{\mathrm{irr}}$	Section 7.3$_{219}$
$\underline{R}_G^{\mathrm{irr}*}$	Regular extension of the unconditional part of $\underline{P}_G^{\mathrm{irr}}$	Section 7.3$_{219}$

OTHER SYMBOLS

Symbol	Meaning	Location
$\mathrm{cl}(\mathscr{D})$	Set of almost desirable gambles that correspond to $\mathscr{D}$	Section 2.3$_{40}$
$\mathrm{int}(\mathscr{D})$	Topological interior of $\mathscr{D}$	Section 2.3$_{40}$
$\mathscr{D}_{\underline{P}}^{\mathrm{r}}$	Set of desirable gambles that results in regular extension	Section 2.7.4
$\mathscr{D}_O$	Updated set of desirable gambles	Section 3.2.1$_{73}$
$\mathscr{D}^O$	Set of O-desirable gambles	Section 3.2.1$_{73}$

$\mathscr{E}_O^{\varepsilon}(\mathscr{D})$	Natural extension of $\mathscr{D} \cup \{\mathbb{I}_O - \varepsilon\}$	Section 3.4.3$_{81}$
$\mathscr{E}_O(\mathscr{D})$	Intersection of all $\mathscr{E}_O^{\varepsilon}(\mathscr{D})$ with $\varepsilon > 0$	Section 3.4.3$_{81}$
$\mathscr{D}_O^{\mathrm{r}}$	Updated set of desirable gambles that is based on $\mathscr{E}_O(\mathscr{D})$	Section 3.4.5$_{87}$
VIR	Epistemic value-irrelevance	Section 4.3.1$_{115}$
SIR	Epistemic subset-irrelevance	Section 4.3.2$_{115}$
HIR	Epistemic h-irrelevance	Section 4.3.3$_{117}$
IR	Epistemic irrelevance	Section 4.3.3$_{117}$
$\otimes$	Independent natural extension	Section 6.6$_{171}$
$\oslash$	Marginal extension	Section 6.7$_{175}$
$\mathrm{ext}(\cdot)$	Operator that returns the extreme points of a set	Page 55

1

INTRODUCTION

"If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is."

John von Neumann

The theory of credal networks [16] extends that of Bayesian networks [78] to allow it to deal with imprecise probability assessments [106] or, loosely speaking, partial probability assessments. As a result, credal networks can represent a wider range of expert knowledge and are able to cope with small data sets robustly, while at the same time keeping many of the features that have helped make Bayesian networks so popular. They achieve this goal by replacing the local probability distributions of a Bayesian network by sets of probability distributions, called credal sets.

This dissertation studies a specific type of credal networks, which are called credal networks under epistemic irrelevance. We build a complete theoretical framework for this previously rather ill-known type of credal network and develop several efficient inference algorithms for them. In this introduction, we provide some general information about credal networks, compare them to Bayesian networks, explain why our type of credal network is especially promising and state our main results. We also discuss the navigational tools in this dissertation and provide a short description of its different chapters. We end this introduction with an overview of our main publications.

1.1 MOTIVATION AND MAIN RESULTS

A Bayesian network [78] is a special type of probabilistic graphical network [64] that is popular in fields such as artificial intelligence, machine learning and statistics. It is able to compactly and intuitively represent the uncertainty that is associated with large multivariate domains and can then be used to answer various queries of interest to the user. It consists of a directed acyclic graph (DAG) whose nodes represent relevant domain variables and whose edges express (in)dependencies between these variables: nodes that are connected by an edge influence each other directly; nodes that are not connected represent variables that are conditionally independent of each other. Attached to each of the nodes is a local probability model for the corresponding variable, conditional on the values of its parents in the DAG. These local models are obtained by eliciting them from experts, by learning them from data, or by means of a combination of both approaches. By combining these local models with the assessments of conditional independence that are expressed by the DAG, it is possible to construct a unique global probabilistic model that can be used for further inferences such as updating, classification and decision making. Efficient algorithms have been developed for performing such inferences, leading to the successful application of Bayesian networks in a multitude of real-life problems [64,82], in various domains such as medicine, bioinformatics, speech recognition, engineering, and so on.

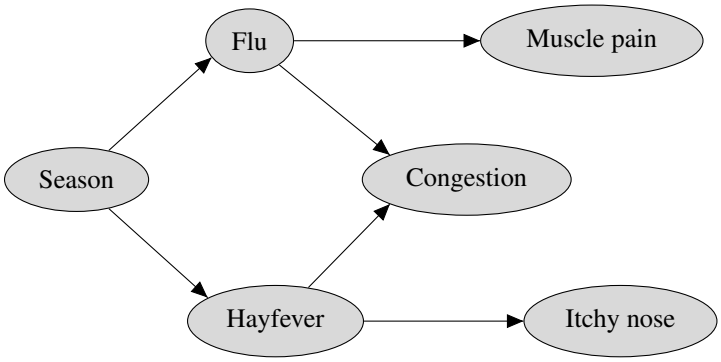


Figure 1.1: A simple example of a Bayesian network

Figure 1.1 depicts a simple hypothetical example of the graphical structure of a medical Bayesian network. Realistic examples typically consist of many more variables, making them far more challenging from a computational point of view. However, the general idea is the same. In this case, for example, the season has a direct influence on your chances for having the flu, and whether or not you have the flu in turn directly influences your chances of having muscle

pain. However, 'Season' and 'Muscle pain' are not directly related: conditional on the value of 'Flu', 'Muscle pain' is independent of 'Season'. The local uncertainty model for 'Muscle pain' consists of two conditional probability distributions, one for each of the two possible values ('yes' and 'no') of its parent variable 'Flu'. Once all the local models are specified, this Bayesian network can be used to answer queries such as the following: given that it is winter and that a patient suffers from congestion but has no muscle pain, what is the probability that she has the flu? By adding decision-theoretic tools such as utility functions, other types of queries can also be answered: for example, given the same—or other—evidence, we can use this Bayesian network to decide whether the patient has the flu or hayfever, or both.

Despite their success, Bayesian networks have an important limitation: the construction of a Bayesian network requires the exact specification of a local conditional probability distribution for every variable in the network. In case of limited data or disagreeing and/or partial expert opinions, this requirement is clearly unrealistic and arguably renders the resulting model fairly arbitrary. Credal networks [16] drop this requirement by allowing for the use of imprecisely specified local models [106]. Although the theory of credal networks has by now adopted various imprecise-probabilistic frameworks, these local models were initially taken to be credal sets, which are sets of probability distributions—hence the terminology. Bayesian networks correspond to the special case where each of these credal sets consists of only a single probability distribution. For example, in order to specify a credal network that has the DAG in Figure $1.1_{\frown}$ as its graphical structure, it is not necessary—although of course allowed—to provide an exact value for the local probability that a patient has muscle pain given that he has the flu. Instead, it is possible to only specify an interval. An expert may for example assess that this probability lies somewhere between 60% and 80%. Partial probability assessments that are more involved than intervals can also be considered. Alternatively, the local credal sets can be learned from data by means of imprecise statistical models such as the IDMM [8, 107] or, if the goal is to perform a sensitivity analysis of some underlying Bayesian network, by considering neighbourhoods of the local probability distributions of that Bayesian network [106, Section 4.6.5].

Credal networks can be used to answer the same queries as Bayesian networks. In a credal network, loosely speaking, updating corresponds to computing a probability interval rather than a single probability. For example, in the case of Figure $1.1_{\frown}$, we can compute the updated lower and upper probability that a patient has the flu, conditional on the information that it is winter and that she suffers from congestion but has no muscle pain. Decision making can be performed in multiple ways [95]. A typical feature is that a credal network may sometimes return multiple answers among which it remains undecided. For example, in the case of Figure $1.1_{\frown}$, deciding whether a patient has the flu may result in three answers: 'yes', 'no' and 'no idea' ('yes' or 'no'). If a credal network provides a single answer, this answer is guaranteed to be

robust with respect to variations within the partial probability constraints that define the credal network. If it remains undecided between some subset of the set of all possible answers, this indicates a lack of robustness of the underlying (precise) probability models. The set-valued solution of a credal network avoids this lack of robustness and, at the same time, often remains informative; it may for example select two answers from a set of, say, twenty. In this way, credal networks can be used to solve the same type of problems as Bayesian networks. The added advantage of the 'credal' approach is that it produces inferences that are more reliable, in the sense that they are robust with respect to the available information or, to put it differently, more honest because they only use the information that is actually available. This approach has been successfully applied to various problems, ranging from military decision support [1] to medical diagnosis [119]. More examples can be found in—among others—References [3, 30, 36].

The main existing problem in the field is that the inferences that are based on credal networks tend to be much harder to compute than their counterparts for Bayesian networks, to the point that they are often intractable even when the corresponding Bayesian network inferences are not. Therefore, the main challenge, and the focus of state of the art research on credal networks, is the development of efficient inference algorithms. The availability of such algorithms, and the extent to which they are efficient, crucially depends on the specific type of credal network that is considered; see Cozman's pioneering work [16] for an overview of different types of credal networks. The most popular type are credal networks under strong independence. This dissertation focuses on a different type: credal networks under epistemic irrelevance. The only difference is the notion of independence they adopt: strong independence or epistemic irrelevance. Both are reasonable notions of independence that generalise stochastic independence to an imprecise probability setting. The former is essentially an assessment of elementwise stochastic independence and takes a sensitivity analysis point of view. The latter is a behavioural notion that can be stated directly in terms of the belief model—the credal set—itself, without having to refer to its individual elements. Consequently, epistemic irrelevance naturally has a wider scope because it can also be imposed in imprecise probability frameworks that go beyond the context of credal sets.

As mentioned above, credal networks under strong independence are by far the most popular type of credal networks. The reason for their popularity is their close connection to Bayesian networks. A credal network under strong independence is essentially a collection of Bayesian networks. Therefore, it is quite intuitive to consider results and techniques from Bayesian networks and try to adapt them to credal networks under strong independence. This has led to an abundance of theoretical results and algorithms for credal networks under strong independence. In fact, the potential for fundamentally new research on this topic seems to slowly saturate. Nevertheless, despite all this attention, only very few efficient inference algorithms are available for this type

of credal network. For exact inference, the only really efficient algorithm is the 2U algorithm [55], which can efficiently compute inferences in singly connected networks that consist of binary variables. Applied researchers usually resort to approximate algorithms, of which there are many; a recent overview of state of the art algorithms can be found in Reference [2]. Of course, as the approximation becomes better, computing it typically becomes less tractable.

Credal networks under epistemic irrelevance are an alternative type of credal network that is rising in popularity. Although they have been around for a while [16], it is only recently—more or less when I started my doctoral research—that they have come to be regarded as a serious alternative. The reason for this recent rise in 'popularity' is the development of an efficient exact algorithm [42] that can compute inferences in credal trees—credal networks whose DAG has a tree structure—whose variables may be non-binary. This result is regarded as very promising because the same type of inference is NP-hard in credal networks under strong independence [67]. However, besides this promising development, very little is known about credal networks under epistemic irrelevance. There is a serious shortage of theoretical results and, consequently, almost no algorithms are available. There are no approximate algorithms and the available exact algorithms are either very inefficient [16] or can only be applied to credal trees [6, 30, 42].

The main contribution of this dissertation is the development of a complete theory of credal networks under epistemic irrelevance, including a number of efficient exact inference algorithms for them. We define this type of credal network in terms of four different imprecise probability frameworks—sets of desirable gambles, lower previsions, sets of linear previsions and credal sets—and connect the resulting models with one another. We prove many previously unknown theoretical properties, including factorisation, external additivity and marginalisation. We also present a result that is analogous to the classical d-separation result in Bayesian networks and establish close connections with the notions of marginal extension and independent natural extension. We use these properties to develop two types of algorithms. The first type can compute any kind of inference, regardless of the topology of the network. However, this generality comes at a price: these algorithms become computationally intractable for larger networks. The second type of algorithms is recursive and can efficiently compute various types of exact inferences in very large networks. As in Bayesian networks, efficiency can only be achieved for a restricted class of topologies. However, as our discussion and our algorithms illustrate, this class is definitely more general than trees.

In order to allow us to obtain and present these results, the availability of powerful tools for modelling uncertainty using imprecise probabilities was crucial, especially tools that are tailored towards a multivariate context. A number of these tools were already available; we present, discuss and connect them to make this dissertation more self-contained. Other tools were not available; our development of them can be regarded as a contribution of its own, which

should be useful also outside of the context of credal networks, whenever there is a need to model uncertainty in a multivariate context.

Finally, on a more philosophical note, we present an extensive discussion on updating. After being informed that an event has occurred, one often wishes to update—or change—the original model to obtain a new model that takes this information into account. For example, in the case of Figure 1.1_{26}, upon receiving the information that a patient has muscle pain, we might want to update our model to take this information into account. The most popular approach for doing so is to condition on this information. In a precise-probabilistic context, this amounts to applying Bayes's rule. Similarly, in an imprecise-probabilistic context, one can apply imprecise conditioning rules such as natural and regular extension. This is also the approach taken in this dissertation: we update a credal network by means of conditioning. However, at first sight, updating and conditioning are two very different things. Conditioning is only a mathematical concept, expressed solely in terms of current beliefs. Updating, on the other hand, is concerned with how to change these beliefs, after being informed that some event has occurred. A claim that these two concepts should somehow be related to one another—let alone that they should coincide—is by no means trivial. Nevertheless, as we will argue extensively, under specific conditions, it does indeed make perfect sense to update by means of a conditioning rule.

1.2 FINDING YOUR WAY AROUND

In this dissertation, as in most manuscripts of considerable length, there are plenty of references, both internal and external.

The main internal references consist of chapters, sections, subsections, appendices, theorems, propositions, corollaries, lemmas and equations. We refer to them by providing their number, which is parenthesized in the case of equations. The reader will often be required to navigate back and forth between these references. In order to try and make this task less frustrating, we have added a subscript to every reference that is not located on the same double-page spread. This subscript provides either the page number or a clue to look at the recto page ($\curvearrowleft$) or the verso page ($\curvearrowright$). Theorem 53_{156}—one of the most important results in this dissertation—can for example be found on page 156 in Section 6.2_{154}, which is the second section of Chapter 6_{152}. Its proof is long and complicated and has therefore been moved to Appendix $6.\mathrm{B}_{188}$, which is the second appendix of that same chapter. Also, we would like to apologise for Equation $(6.20)_{207}$.

External references are listed at the end of this dissertation. We refer to them by means of numbers in square brackets. Interested readers with plenty of time on their hands could for example stroll through Reference [106], which is Walley's seminal book on imprecise probabilities, or read Cozman's pioneering paper on credal networks [16].

1.3 A BRIEF OVERVIEW

Excluding this introduction, the rest of this dissertation consists of seven chapters. The first three are concerned with various general aspects of modelling uncertainty. In the subsequent three chapters, we use these general results to develop a theory of credal networks under epistemic irrelevance. The final chapter presents our conclusions and discusses avenues for further research that we consider to be promising.

We give a brief overview of the results that are discussed in the six main chapters. The introductions to these individual chapters often provide additional information.

We start off in Chapter 2_{37} by connecting the theory of sets of desirable gambles with that of conditional lower previsions, sets of conditional linear previsions and sets of full conditional probability mass functions, restricting ourselves to the case where the uncertain variables of interest take values in a finite space. These are the four main imprecise-probabilistic frameworks that we consider in this dissertation. We provide an overview of some of the most important results in the literature and fill in some gaps as we go along. For readers who are new to imprecise probabilities, this can be regarded as an introduction to the field. The more advanced reader will notice that our approach differs from some of the more conventional ones. We build the theory from scratch, using sets of desirable gambles as our starting point, in a way that closely resembles—but nevertheless differs slightly from—the approach that is advocated by Williams [112]. This chapter also discusses the two main imprecise-probabilistic conditioning rules: natural and regular extension. We define these rules in terms of each of the four frameworks we consider, compare them to one another, provide pointers to the literature, and discuss various technical and computational aspects.

In Chapter 3_{69}, we shift the focus from conditioning to updating and study the problem of updating directly in terms of sets of desirable gambles. We introduce an asymmetric—and arguably improved—version of Walley's updating principle, discuss the conditions under which it makes sense to use it, and explain how it leads to a justification for updating by means of natural extension. It turns out that our approach leaves room for other updating rules as well, including more informative ones such as regular extension. However, in order to justify them, our asymmetric version of Walley's updating principle is not sufficient, and needs to be combined with additional arguments. This is exactly what we do in the remainder of the chapter, for the particular case of updating by means of regular extension. The basic idea is that, since we are looking for an updated model that is meant to be used after some event has occurred, we are—in the process of coming up with such an updated model—making an implicit assumption that this event can occur. This assumption allows us to include an extra assessment that, when combined with our asymmetric version of Walley's updating principle, leads to conditional models that coincide with

regular extension. We provide two versions of this justification for updating by means of regular extension. A simple version, which requires an assumption of ideal precision, and a more involved one, which does not.

Chapter 4_{100} explains how to model uncertainty in a multivariate context. We study the concept of marginalisation and its interplay with conditioning, discuss and compare various definitions of epistemic irrelevance, explain why one of these definitions is to be preferred and compare epistemic irrelevance with other imprecise-probabilistic notions of independence. Finally, we discuss the concept of conservative reasoning and explain how it allows us to extend the notion of natural extension in such a way that it can deal with a combination of direct and structural assessments.

Chapter 5_{128} marks the beginning of the second part of this dissertation, which specifically focuses on credal networks under epistemic irrelevance. We explain how the assessments of epistemic irrelevance that correspond to such a network can be combined with its local models to construct a unique most conservative global uncertainty model, which we call the irrelevant natural extension of the credal network. In contrast with Cozman, who invented this concept [16, Section 8.3], we do not restrict ourselves to the framework of credal sets, but consider other imprecise-probabilistic frameworks as well, including sets of desirable gambles and lower previsions. Our approach also has the advantage that it does not require the simplifying positivity assumptions imposed by Cozman. We end this chapter by comparing credal networks under epistemic irrelevance with other types of credal networks.

Chapter 6_{152} develops some remarkable theoretical properties of the irrelevant natural extension of a credal network. The starting point, and perhaps the main technical achievement of this dissertation is a very strong separating hyperplane result. From it, we are able to derive various theoretical properties of the irrelevant natural extension, including factorisation, external additivity and marginalisation properties. We also show that the irrelevant natural extension satisfies separation properties that are similar to the ones that are induced by d-separation in Bayesian networks. We introduce an asymmetric version of d-separation, called AD-separation, and prove that it implies epistemic irrelevance. Furthermore, since AD-separation is shown to satisfy all asymmetric graphoid properties—all graphoid properties except symmetry—the induced set of epistemic irrelevancies does so as well. We end this chapter by developing connections with the notions of independent natural extension and marginal extension, and by proving marginalisation properties for the updated models that are obtained when we condition the irrelevant natural extension of a credal network using regular extension.

Chapter 7_{211} is the last main chapter of this dissertation and is devoted to the development of efficient inference algorithms. We begin by presenting some techniques that can be used to simplify inference problems beforehand, in a preprocessing step, before any actual inference algorithm is applied. This includes techniques for removing barren nodes and AD-separated evidence as

well as ways of reducing a conditional inference problem to an unconditional one. In the next part of this chapter, we show that inference can be reformulated as a linear programming problem, even in the presence of conditioning events with probability zero. However, the size of the linear programs that need to be solved is exponential in the size of the credal network and, therefore, this approach is only tractable for small networks. In the case of two binary variables, we use this linear programming description to obtain elegant closed-form expressions for the extreme points of the independent natural extension of two binary models. In the final part of this chapter, we develop efficient recursive algorithms for exact inference. We start by focusing on credal networks of which the underlying DAG is recursively decomposable—a new type of DAG that includes trees as a special case. For these networks, we develop several efficient inference algorithms, for various types of inference problems, including inference problems that deal with multiple query variables at once. For credal networks under epistemic irrelevance of which the underlying DAG is not recursively decomposable, it is still possible to develop efficient inference algorithms, but only in specific cases. We illustrate this by means of examples. A particularly interesting example are inferences about a single query node in case of complete evidence.

1.4 PUBLICATIONS

The research that led to this dissertation has resulted in fourteen publications. Five of them have been published in international journals [30–33, 40]. The other nine have been presented at international conferences and were subsequently published in their proceedings [24–29, 34, 39, 41]. The results in this dissertation represent only a small subset of these publications. In order to turn this dissertation into a coherent story that focuses on a single line of research, I have decided to only include results that are directly related to credal networks under epistemic irrelevance. In order to paint a more complete picture of my research, I end this introduction with a brief overview of my main results, focusing especially on the results that are not discussed in this dissertation. For the convenience of the reader, I explictly mention the authors, title and journal—or conference proceedings—of each of my publications; the bibliography contains more detailed information.

The three main publications that did make it into this dissertation are References [27, 28, 31]:

- Jasper De Bock and Gert de Cooman. Allowing for probability zero in credal networks under epistemic irrelevance. Published in the proceedings of ISIPTA '13 [27].
- Jasper De Bock and Gert de Cooman. Credal networks under epistemic irrelevance using sets of desirable gambles. Published in the proceedings

of ISIPTA '13 [28].

- Jasper De Bock and Gert de Cooman. Credal networks under epistemic irrelevance: the sets of desirable gambles approach. Published in the International Journal of Approximate Reasoning [31].

These publications formed the basis for many of the theoretical results in Chapters 5_{128} and 6_{152} and Section 7.4_{221}. The main difference with our former exposition of these results is that we now present them in terms of four different imprecise-probabilistic frameworks, thereby making them more accessible to various audiences. Most of our algorithms in Chapter 7_{211}—especially the efficient recursive ones—are very recent and have therefore not been published yet. This is also the case for the results in Chapters 2_{37}, 3_{69} and 4_{100}.

For imprecise hidden Markov models that adopt epistemic irrelevance as their notion of independence—a specific type of credal network under epistemic irrelevance—I have developed a robust version of the Viterbi algorithm that is capable of robustly estimating the value of a hidden sequence of variables based on a corresponding sequence of—possibly incorrect—observations of these variables, and I have used it to automatically correct the errors that are made by Optical Character Recognition (OCR) software [25, 30, 41]:

- Jasper De Bock and Gert de Cooman. State sequence prediction in imprecise hidden Markov models. Published in the proceedings of ISIPTA '11 [25].

- Jasper De Bock and Gert de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. Published in the Journal of Artificial Intelligence Research [30].

- Gert de Cooman, Jasper De Bock and Arthur Van Camp. Recent advances in imprecise-probabilistic graphical models. Published in the proceedings of ECAI 2012 [41].

These results are closely related to the material in this dissertation and are briefly mentioned in Section $7.5.7_{269}$. I do not discuss these results at length because most of them were already reported in my master dissertation, be it in a less developed form.

Besides credal networks under epistemic irrelevance, I have also worked on credal networks under complete independence, which are very similar—and for most inference problems even equivalent—to credal networks under strong independence. For the corresponding notion of a hidden Markov model, I have recently designed an algorithm that can be regarded as the 'complete independence'-version of the robust Viterbi algorithm that was mentioned above [34]:

- Cedric De Boom, Jasper De Bock, Arthur Van Camp and Gert de Cooman. Robustifying the Viterbi algorithm. Published in the proceedings of PGM 2014 [34].

My other work on credal networks under complete independence has been concerned with using them to efficiently perform a global sensitivity analysis in Bayesian networks and, by extension, in more general graphical models such as Markov random fields [24]:

- Jasper De Bock, Cassio P. de Campos and Alessandro Antonucci. Global sensitivity analysis for MAP inference in graphical models. Published in the proceedings of NIPS 2014 [24].

Finally, during my research on credal networks, I have often not been able to restrain myself from wandering off into other parts of the vast world that is imprecise probability theory. This has led to a number of additional results that are not directly related to credal networks. Some of this material has not been published yet, most notably a number of results in game-theoretic probability, including a game-theoretic ergodic theorem for imprecise Markov chains; a paper on this topic is currently under review. Other results were published in References [26, 29, 32, 33, 39, 40].

A first set of these results originated from my research on an imprecise-probabilistic version of the concept of exchangeability, a structural assessment of symmetry that can be imposed on multivariate models. This has led to a behavioural justification for the use of imprecise Bernoulli processes [26], imprecise-probabilistic representation theorems for partially exchangeable random variables [33], and the development of new predictive inference models [39, 40]:

- Jasper De Bock and Gert de Cooman. Imprecise Bernoulli processes. Published in the proceedings of IPMU 2012 [26].
- Jasper De Bock, Arthur Van Camp, Márcio Alves Diniz and Gert de Cooman. Representation theorems for partially exchangeable random variables. Accepted for publication in Fuzzy Sets and Systems [33].
- Gert de Cooman, Jasper De Bock and Márcio Alves Diniz. Predictive inference under exchangeability, and the imprecise Dirichlet multinomial model. Published in the proceedings of EBEB 2014 [39].
- Gert de Cooman, Jasper De Bock and Márcio Alves Diniz. Coherent predictive inference under exchangeability with imprecise probabilities. Published in the Journal of Artificial Intelligence Research [40].

The remaining result is a connection between the geometrical concept of Minkowski decomposability and the problem of finding the extreme points of

the set of all coherent lower previsions on the linear space of all gambles on some finite state space [29, 32], which are called extreme lower previsions:

- Jasper De Bock and Gert de Cooman. Extreme lower previsions and Minkowski indecomposability. Published in the proceedings of ESQARU 2013 (winner of the best student paper award) [29].
- Jasper De Bock and Gert de Cooman. Extreme lower previsions. Published in the Journal of Mathematical Analysis and Applications [32].

2

MODELLING UNCERTAINTY

"Free yourself from the rigid conduct of tradition and open yourself to the new forms of probability."

Hans Bender

Consider a variable X that takes values ω in a non-empty finite state space Ω. This could be the number of days it will rain in Ghent next year, the name of the first person that you will meet after reading this very sentence, or simply the outcome of some coin flip. As you can gather from these examples, the actual value of such a variable X may be unknown. A subject's—for example your—uncertainty about the value of X can then be represented by means of a belief model. The most common example of such a belief model is a single probability mass function on Ω. However, it is far from the only one, and definitely not the most general one.

In this chapter, we introduce four alternative frameworks for constructing a belief model that captures a subject's uncertainty about the value of X: sets of desirable gambles, (conditional) lower previsions, sets of (conditional) linear previsions and sets of (full conditional) probability mass functions. We build these theories from the ground up starting from basic principles and connect them with each other. They all share two advantages over working with individual probability mass functions. First of all: they allow for imprecision; basically, this means that lower and upper probabilities need not coincide. Secondly, conditioning on events with (lower) probability zero becomes non-problematic and sometimes even trivial.

2.1 PRELIMINARIES

Essential to each of the frameworks that we are about to introduce is the notion of a *gamble* on Ω, which is a real-valued function on Ω that is interpreted as an uncertain payoff. If the actual value of X turns out to be ω, the owner of a gamble f receives the—possibly negative—payoff $f(\omega)$, expressed in units of some predetermined linear utility scale.[1] We denote the set of all gambles on Ω as $\mathscr{G}(\Omega)$. This is a linear space under pointwise addition of gambles and pointwise multiplication of gambles with real numbers. For any two f_1 and f_2 in $\mathscr{G}(\Omega)$, we write '$f_1 \geq f_2$' if $(\forall \omega \in \Omega)\, f_1(\omega) \geq f_2(\omega)$ and '$f_1 > f_2$' if $f_1 \geq f_2$ and $f_1 \neq f_2$. Interesting subsets of $\mathscr{G}(\Omega)$ are denoted by using predicates as subscripts; for example: $\mathscr{G}(\Omega)_{>0} := \{f \in \mathscr{G}(\Omega) \colon f > 0\}$ is the set of all non-negative gambles on Ω, excluding zero.

Events are identified with subsets of Ω. Hence, the set of all events is the power set $\mathscr{P}(\Omega)$ of Ω. We will often consider the set $\mathscr{P}_\emptyset(\Omega) := \mathscr{P}(\Omega) \setminus \{\emptyset\}$ of all non-empty events as well. Since Ω is finite, $\mathscr{P}(\Omega)$ and $\mathscr{P}_\emptyset(\Omega)$ are finite too. With every event $B \in \mathscr{P}(\Omega)$, we associate a special gamble $\mathbb{I}_B$ on Ω, called its *indicator*, that assumes the value 1 on B and 0 elsewhere.

2.2 SETS OF DESIRABLE GAMBLES

Sets of desirable gambles constitute the first framework we consider [85, 109]. The basic idea here is to model a subject's uncertainty about the value of X by means of a set $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ of gambles on Ω—risky transactions whose payoff depends on the value of X—that he considers to be *desirable*. A subject is said to find a gamble $f \in \mathscr{G}(\Omega)$ desirable if he strictly prefers it to the zero gamble or, equivalently, if he strictly prefers ownership of f over ownership of the zero gamble—the status quo.

In order to reflect a rational subject's uncertainty, a *set of desirable gambles* $\mathscr{D}$ should be coherent, meaning that it satisfies the following consistency criteria.

Definition 1 (Coherence for sets of desirable gambles)**.** *A set $\mathscr{D}$ of desirable gambles on Ω is* coherent *if for all $\lambda \in \mathbb{R}_{>0}$ and all $f, f_1, f_2 \in \mathscr{G}(\Omega)$:*[2]

D1. $f = 0 \Rightarrow f \notin \mathscr{D}$; [avoiding null gain]

D2. $f > 0 \Rightarrow f \in \mathscr{D}$; [desiring partial gain]

[1] As long as the amounts of money remain limited, many people perceive the utility of a monetary reward to be a linear function of its monetary value; see Reference [106, Sections 2.2.1 and 2.2.2] for additional discussion, including an example of a well-defined utility scale that is perfectly linear.

[2] We use $\mathbb{R}_{>0}$ as a convenient shorthand notation for $\{\lambda \in \mathbb{R} \colon \lambda > 0\}$, and similarly for $\mathbb{R}_{\geq 0}$.

D3. $f \in \mathscr{D} \Rightarrow \lambda f \in \mathscr{D}$; [positive scaling]

D4. $f_1, f_2 \in \mathscr{D} \Rightarrow f_1 + f_2 \in \mathscr{D}$. [combination]

Criteria D1$_\frown$ and D2$_\frown$ are rationality criteria; they follow directly from our interpretation of desirability. The zero gamble should not be desirable [D1$_\frown$]; gambles without negative payoffs and with the possibility of a positive payoff should always be desirable [D2$_\frown$]. Criteria D3 and D4 follow from the linearity of our utility scale.

Coherence has a number of useful consequences, which can be obtained by combining D1$_\frown$–D4. For example, for any coherent $\mathscr{D}$, and any $f, g \in \mathscr{G}(\Omega)$:

D5. $g \geq f$ and $f \in \mathscr{D} \Rightarrow g \in \mathscr{D}$; [monotonicity]

D6. $f \leq 0 \Rightarrow f \notin \mathscr{D}$. [avoiding non-positive gain]

2.2.1 *Natural extension*

In practice, we cannot expect a subject to specify for each gamble $f \in \mathscr{G}(\Omega)$ whether or not he finds it desirable. Instead, all that is usually obtained from an elicitation procedure is an *assessment* $\mathscr{A} \subseteq \mathscr{G}(\Omega)$, which may be only a subset of a subject's set of desirable gambles. Furthermore, such an assessment is often not coherent. However, by applying D2$_\frown$–D4, we can use $\mathscr{A}$ to infer the desirability of other gambles. The largest set of desirable gambles that can be constructed in this way is[3]

$$\mathscr{E}(\mathscr{A}) := \left\{ \sum_{i=1}^{n} \lambda_i f_i : n \in \mathbb{N}, f_i \in \mathscr{A} \cup \mathscr{G}(\Omega)_{>0}, \lambda_i \in \mathbb{R}_{>0} \right\}. \tag{2.1}$$

By construction, $\mathscr{E}(\mathscr{A})$ satisfies D2$_\frown$–D4. Consequently, $\mathscr{E}(\mathscr{A})$ is coherent if and only if it avoids null gain [D1$_\frown$]. Furthermore, if $\mathscr{E}(\mathscr{A})$ is coherent, then it is the smallest coherent set of desirable gambles that contains $\mathscr{A}$, and we then call $\mathscr{E}(\mathscr{A})$ the *natural extension* of $\mathscr{A}$. Since coherence is trivially preserved under taking intersections, this natural extension $\mathscr{E}(\mathscr{A})$ is then also equal to the intersection of all the coherent supersets of $\mathscr{A}$.

Even after enlarging an assessment by means of natural extension, the resulting set of desirable gambles is not guaranteed to be exhaustive. Further elicitation may result in additional desirable gambles. However, for various reasons, one may be unwilling or incapable of performing such further elicitation; see Reference [106, Section 2.10.3] for numerous examples. Hence, we will not require a set of desirable gambles $\mathscr{D}$ to be exhaustive, nor will we interpret it in this way.

[3] We define the natural numbers $\mathbb{N}$ as the set of all *positive* integers. We use $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ to refer to the version that includes zero.

2.2.2 *Conditional sets of desirable gambles*

One of the advantages of working with sets of desirable gambles is that conditioning is extremely elegant. Consider a set of desirable gambles $\mathscr{D}$ and an event $B \in \mathscr{P}_\emptyset(\Omega)$. Then the corresponding *conditional set of desirable gambles* is given by [47]

$$\mathscr{D}\rfloor B := \{f \in \mathscr{G}(B) \colon \mathbb{I}_B f \in \mathscr{D}\}, \tag{2.2}$$

where, by introducing the convention that $0 \times \text{undefined} := 0$, we let $\mathbb{I}_B f$ be a gamble on Ω that coincides with f on B and is zero elsewhere. The intuition behind this definition is very simple: when B occurs, the gambles $\mathbb{I}_B f$ and f are indistinguishable in practice. Contingent on B occurring, they yield the same payoff; if B does not occur, then $\mathbb{I}_B f$ results in a zero payoff whereas f is not defined. This definition of conditioning preserves coherence: if $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ is coherent, then $\mathscr{D}\rfloor B \subseteq \mathscr{G}(B)$ is clearly coherent as well.

Alternative methods for conditioning a set of desirable gambles have also been proposed [14, 76, 106, 109], using the notation '$|$' rather than '$\rfloor$'; all of these alternative methods result in a set of gambles on Ω instead of B. We prefer the present version because we find it more intuitive that conditioning on an event B produces a model for—a set of gambles on—that event. In any case, the choice between these definitions is mainly an aesthetic one, because they are all mathematically equivalent [47, Section 3.2].

2.3 LOWER PREVISIONS

Instead of asking a subject to evaluate the desirability of a gamble directly, one can also ask him at which prices he would be willing to buy or sell that gamble. This is the approach that is taken in Walley's theory of lower previsions [68, 71, 96, 106]. For any gamble f on Ω, the *lower prevision* $\underline{P}(f)$ of f is a subject's supremum buying price for f. Similarly, the upper prevision $\overline{P}(f)$ of f is his infimum selling price for f. Since selling f for a price α is equivalent to buying $-f$ for a price $-\alpha$, lower and upper previsions are related by *conjugacy*: $\overline{P}(f) = -\underline{P}(-f)$. For this reason, it suffices to discuss only one of them. We follow Walley in concentrating on lower previsions.

Due to their interpretation as supremum buying prices, lower previsions can easily be related to sets of desirable gambles. In order to connect both approaches, it suffices to require that a subject considers the gamble $f - \alpha$ to be desirable if and only if he strictly prefers buying f for the price α to the status quo—not buying any gamble at all. Using this connection, a coherent set of desirable gambles $\mathscr{D}$ trivially results in a lower prevision $\underline{P}_{\mathscr{D}}$, defined by

$$\underline{P}_{\mathscr{D}}(f) := \sup\bigl\{\alpha \in \mathbb{R} \colon f - \alpha \in \mathscr{D}\bigr\} \text{ for every } f \in \mathscr{G}(\Omega). \tag{2.3}$$

Alternatively, lower previsions can be assessed directly as well. Any real valued function $\underline{P}$ with arbitrary domain $\mathscr{K} \subseteq \mathscr{G}(\Omega)$ can be interpreted as a lower

prevision. We say that $\underline{P}$ is *coherent* if there is a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ such that $\underline{P}(f) = \underline{P}_{\mathscr{D}}(f)$ for all $f \in \mathscr{K}$.[4]

Lower previsions are less expressive than sets of desirable gambles. For a given coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, there may be multiple coherent sets of desirable gambles $\mathscr{D}$ such that $\underline{P} = \underline{P}_{\mathscr{D}}$, the smallest of which is[5]

$$\mathscr{D}_{\underline{P}} := \{f \in \mathscr{G}(\Omega) \colon \underline{P}(f) > 0 \text{ or } f > 0\}. \tag{2.4}$$

All of these sets have the same associated set of *almost desirable gambles*

$$\begin{aligned} \mathrm{cl}(\mathscr{D}) :=& \big\{f \in \mathscr{G}(\Omega) \colon (\forall \delta \in \mathbb{R}_{>0})\ f + \delta \in \mathscr{D}\big\} && (2.5)\\ =& \{f \in \mathscr{G}(\Omega) \colon \underline{P}_{\mathscr{D}}(f) \geq 0\}. && (2.6) \end{aligned}$$

We write $\mathrm{cl}(\mathscr{D})$ because, if $\mathscr{D}$ is coherent, then the set on the right-hand side of the defining equality in Equation (2.5) is equal to the topological closure of $\mathscr{D}$, with respect to the topology that is induced by the Euclidean metric. Similarly, they will also have the same topological interior

$$\begin{aligned} \mathrm{int}(\mathscr{D}) :=& \big\{f \in \mathscr{G}(\Omega) \colon (\exists \delta \in \mathbb{R}_{>0})\ f - \delta \in \mathscr{D}\big\} && (2.7)\\ =& \{f \in \mathscr{G}(\Omega) \colon \underline{P}_{\mathscr{D}}(f) > 0\}. \end{aligned}$$

It is furthermore easily proved that these conditions are equivalent: if $\mathscr{D}_1$ and $\mathscr{D}_2$ are two coherent sets of desirable gambles, then

$$\underline{P}_{\mathscr{D}_1} = \underline{P}_{\mathscr{D}_2} \Leftrightarrow \mathrm{cl}(\mathscr{D}_1) = \mathrm{cl}(\mathscr{D}_2) \Leftrightarrow \mathrm{int}(\mathscr{D}_1) = \mathrm{int}(\mathscr{D}_2). \tag{2.8}$$

Hence, coherent sets of desirable gambles with the same lower prevision $\underline{P}_{\mathscr{D}}$ differ only in their border $\mathrm{cl}(\mathscr{D}) \setminus \mathrm{int}(\mathscr{D})$. Nevertheless, which part of this border belongs to $\mathscr{D}$—the border structure of $\mathscr{D}$—may be important, for the following two reasons [109]. First of all, in a decision making context, it enables one to distinguish between strict and weak preference. For example, for two gambles $f, g \in \mathscr{G}(\Omega)$, we might say that f is strictly preferred over g if and only if $f - g \in \mathscr{D}$, whereas f is weakly preferred over g if and only if $f - g \in \mathrm{cl}(\mathscr{D}) \Leftrightarrow \underline{P}_{\mathscr{D}}(f - g) \geq 0$. The set of desirable gambles $\mathscr{D}$ is able to distinguish between these two notions, but the lower prevision $\underline{P}_{\mathscr{D}}$ is not.[6] Secondly, as we will illustrate further on, the border structure of a set of desirable gambles may have a significant impact on the conditional models it produces.

[4] The connection with other definitions of coherence for lower previsions is discussed in Section 2.3.3$_{47}$.

[5] In Reference [106, Section 3.8.1], this set is denoted by $\mathscr{D}^+$ and is called the associated set of strictly desirable gambles.

[6] Similar observations can be made for other notions of strict and weak preference for sets of desirable gambles; see for example References [74, 106, 109].

2.3.1 *Conditional lower previsions*

In the theory of lower previsions, conditional models are not merely regarded as derived concepts that are obtained through conditioning. Instead, they are primitive concepts, can be assessed directly, and are related to their unconditional counterparts by coherence. For any event $B \in \mathscr{P}_{\emptyset}(\Omega)$ and any gamble $f \in \mathscr{G}(B)$, we interpret the conditional lower prevision $\underline{P}(f\rfloor B)$ of f given B as a subject's supremum buying price for f, contingent on the occurrence of B.[7,8] When considered as an operator, a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ is a real-valued function whose domain can be any set $\mathscr{C}$ of couples (f,B), with $B \in \mathscr{P}_{\emptyset}(\Omega)$ and $f \in \mathscr{G}(B)$. Hence, if we let

$$\mathscr{C}(\Omega) := \{(f,B) \colon B \in \mathscr{P}_{\emptyset}(\Omega), f \in \mathscr{G}(B)\}$$

be the largest such set, then $\mathscr{C}$ can be any subset of $\mathscr{C}(\Omega)$. If $\mathscr{C}$ contains only couples of the form (f,Ω), with $f \in \mathscr{K} \subseteq \mathscr{G}(\Omega)$, then $\underline{P}(\cdot\rfloor\cdot)$ can be identified with an unconditional lower prevision $\underline{P}$ on $\mathscr{K}$, defined by $\underline{P}(f) := \underline{P}(f\rfloor\Omega)$ for all $f \in \mathscr{K}$. Furthermore, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$, the operator $\underline{P}(\cdot\rfloor B)$ can be regarded as an unconditional lower prevision on

$$\mathscr{C}_B := \{f \in \mathscr{G}(B) \colon (f,B) \in \mathscr{C}\} \subseteq \mathscr{G}(B).$$

Due to the connection between desirability and buying prices, every coherent set of desirable gambles $\mathscr{D}$ has a unique corresponding conditional lower prevision $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$, obtained by letting $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ be the lower prevision that corresponds to $\mathscr{D}\rfloor B$:

$$\underline{P}_{\mathscr{D}}(f\rfloor B) := \sup\big\{\alpha \in \mathbb{R} \colon f-\alpha \in \mathscr{D}\rfloor B\big\} \tag{2.9}$$

$$= \sup\big\{\alpha \in \mathbb{R} \colon \mathbb{I}_B[f-\alpha] \in \mathscr{D}\big\} \text{ for every } (f,B) \in \mathscr{C}(\Omega). \tag{2.10}$$

However, sets of desirable gambles are still more expressive; different $\mathscr{D}$ can lead to the same $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$ [74, Section 6].

A conditional lower prevision is said to be coherent if it can be derived from a coherent set of desirable gambles by means of Equation (2.10).

Definition 2 (Coherence for conditional lower previsions)**.** *A conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ is said to be* coherent *if there is some coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ such that*[9]

$$\underline{P}(f\rfloor B) = \underline{P}_{\mathscr{D}}(f\rfloor B) \textit{ for every } (f,B) \in \mathscr{C}.$$

[7] Other authors consider the conditional lower prevision $\underline{P}(f|B)$ of gambles $f \in \mathscr{G}(\Omega)$ instead; the connection with our approach will be established in Section 2.3.3$_{47}$.

[8] See Equation (2.10) for a definition in terms of sets of desirable gambles.

[9] We will establish a connection with other definitions of coherence for conditional lower previsions in Section 2.3.3$_{47}$.

Since coherence for sets of desirable gambles is preserved under taking intersections, we find that coherence for conditional lower previsions is preserved under taking pointwise infima, or equivalently: the lower envelope of a set of coherent lower previsions is again a coherent lower prevision.

Proposition 1. *Consider an arbitrary index set I and, for every $i \in I$, a coherent conditional lower prevision $\underline{P}_i(\cdot\rfloor\cdot)$ on $\mathscr{C} \subseteq \mathscr{C}(\Omega)$. The conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, defined by*

$$\underline{P}(f\rfloor B) := \inf_{i\in I} \underline{P}_i(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}$$

is then also coherent.

Proof of Proposition 1. For every $i \in I$, the fact that $\underline{P}_i(\cdot\rfloor\cdot)$ is coherent implies the existence of a coherent set of desirable gambles $\mathscr{D}_i \subseteq \mathscr{G}(\Omega)$ such that $\underline{P}_i(f\rfloor B) = \underline{P}_{\mathscr{D}_i}(f\rfloor B)$ for all $(f,B) \in \mathscr{C}$. Now let $\mathscr{D} := \cap_{i\in I}\mathscr{D}_i$. Since coherence for sets of desirable gambles is trivially preserved under taking intersections, we know that $\mathscr{D}$ is coherent. Consider now any $(f,B) \in \mathscr{C}$. In order to prove the result, it clearly suffices to show that $\underline{P}_{\mathscr{D}}(f\rfloor B) = \inf_{i\in I}\underline{P}_i(f\rfloor B)$.

For all $i \in I$, we infer from $\mathscr{D} \subseteq \mathscr{D}_i$ that $\underline{P}_{\mathscr{D}}(f\rfloor B) \leq \underline{P}_{\mathscr{D}_i}(f\rfloor B) = \underline{P}_i(f\rfloor B)$, which in turn implies that $\underline{P}_{\mathscr{D}}(f\rfloor B) \leq \inf_{i\in I}\underline{P}_i(f\rfloor B)$. Consider now any $\varepsilon \in \mathbb{R}_{>0}$. Equation $(2.10)_{\cap}$ then implies that $\mathbb{I}_B[f - \underline{P}_{\mathscr{D}}(f\rfloor B) - \varepsilon] \notin \mathscr{D} = \cap_{i\in I}\mathscr{D}_i$. Hence, there is some $i^* \in I$ such that $\mathbb{I}_B[f - \underline{P}_{\mathscr{D}}(f\rfloor B) - \varepsilon] \notin \mathscr{D}_{i^*}$. Since $\mathscr{D}_{i^*}$ is coherent, this implies that $\mathbb{I}_B[f-\alpha] \notin \mathscr{D}_{i^*}$ for all $\alpha \geq \underline{P}_{\mathscr{D}}(f\rfloor B) + \varepsilon$, which in turn implies that $\underline{P}_{\mathscr{D}_{i^*}}(f\rfloor B) \leq \underline{P}_{\mathscr{D}}(f\rfloor B) + \varepsilon$. Since we also know that $\inf_{i\in I}\underline{P}_i(f\rfloor B) = \inf_{i\in I}\underline{P}_{\mathscr{D}_i}(f\rfloor B) \leq \underline{P}_{\mathscr{D}_{i^*}}(f\rfloor B)$, this allows us to infer that $\inf_{i\in I}\underline{P}_i(f\rfloor B) \leq \underline{P}_{\mathscr{D}}(f\rfloor B) + \varepsilon$. Since this is true for all $\varepsilon \in \mathbb{R}_{>0}$, we find that $\inf_{i\in I}\underline{P}_i(f\rfloor B) \leq \underline{P}_{\mathscr{D}}(f\rfloor B)$. We conclude that $\underline{P}_{\mathscr{D}}(f\rfloor B) = \inf_{i\in I}\underline{P}_i(f\rfloor B)$. □

Coherence of $\underline{P}(\cdot\rfloor\cdot)$ implies that $\underline{P}(\cdot\rfloor\cdot)$ is *separately coherent* [106, Section 6.2.2], by which we mean that, for all $B \in \mathscr{P}_{\emptyset}(\Omega)$, $\underline{P}(\cdot\rfloor B)$ is a coherent lower prevision on $\mathscr{C}_B$.[10] However, separate coherence does not imply coherence: coherence of $\underline{P}(\cdot\rfloor B)$, for all $B \in \mathscr{P}_{\emptyset}(\Omega)$, is not sufficient for $\underline{P}(\cdot\rfloor\cdot)$ to be coherent. Whenever we want to clearly distinguish between separate coherence and coherence, we will refer to the latter as *joint coherence*.

If for every $B \in \mathscr{B}$, with $\mathscr{B}$ a subset of $\mathscr{P}_{\emptyset}(\Omega)$, we have a lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{K}_B \subseteq \mathscr{G}(B)$, we use $\underline{P}(\cdot\rfloor\mathscr{B}) := \{\underline{P}(\cdot\rfloor B) \colon B \in \mathscr{B}\}$ to refer to this collection of lower previsions. It should be clear that such a collection can also be identified—trivially—with a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C} := \{(f,B) \colon B \in \mathscr{B}, f \in \mathscr{K}_B\}$. We call a collection $\underline{P}(\cdot\rfloor\mathscr{B})$ separately coherent if each of its individual elements $\underline{P}(\cdot\rfloor B)$ is coherent; we call it (jointly)

[10] It suffices to realise that there is a coherent set of desirable gambles on $\mathscr{G}(B)$—in particular: $\mathscr{D}\rfloor B$—from which $\underline{P}(\cdot\rfloor B)$ can be obtained by means of Equation $(2.3)_{40}$—in particular: Equation $(2.9)_{\cap}$.

coherent if the corresponding conditional lower prevision is (jointly) coherent. If the collection we are referring to is clear from the context, we do not mention it explicitly. For example, we might say that a lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ is *coherent with* a lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{G}(B)$; by this, we simply mean that the corresponding collection—consisting of the lower previsions $\underline{P}(\cdot\rfloor\Omega) := \underline{P}(\cdot)$ and $\underline{P}(\cdot\rfloor B)$—is jointly coherent.

2.3.2 *Natural extension*

Since $\mathscr{D}$ is not required to be exhaustive, $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ is not exhaustive either; the subject's actual supremum buying price for f contingent on B may be higher than $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$. If $\underline{P}(\cdot\rfloor\cdot)$ is assessed directly, then similarly, we do not require it to be exhaustive. A particularly useful advantage of this interpretation is that it allows us to turn a—possibly incoherent—lower prevision $\underline{P}(\cdot\rfloor\cdot)$ into a coherent one by correcting it upwards.

To understand how this comes about naturally, the first step is to realise that a coherent lower prevision is simply an assessment of desirable gambles: due to our interpretation for $\underline{P}(f\rfloor B)$, we know that for any $\varepsilon \in \mathbb{R}_{>0}$, there is some $\alpha \geq \underline{P}(f\rfloor B) - \varepsilon$ such that $\mathbb{I}_B[f-\alpha]$ is a desirable gamble. By combining this with $\mathrm{D2}_{38}$ and $\mathrm{D4}_{39}$, we find that the gambles in

$$\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)} := \left\{\mathbb{I}_B[f - \underline{P}(f\rfloor B) + \varepsilon] \colon (f,B) \in \mathscr{C} \text{ and } \varepsilon \in \mathbb{R}_{>0}\right\}$$

are desirable and therefore also, by $\mathrm{D2}_{38}$–$\mathrm{D4}_{39}$, that each of the gambles in $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} := \mathscr{E}(\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)})$ is desirable. Furthermore, since $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is the natural extension of the assessment $\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$, we know from Section $2.2.1_{39}$ that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is the smallest set of gambles whose desirability is implied by coherence ($\mathrm{D2}_{38}$–$\mathrm{D4}_{39}$) and (the assessment that corresponds to) $\underline{P}(\cdot\rfloor\cdot)$.

The next step is to consider the supremum buying prices that correspond to this set of desirable gambles $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, as given by

$$\underline{E}(f\rfloor B) := \underline{P}_{\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}}(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}(\Omega). \tag{2.11}$$

The resulting operator $\underline{E}(\cdot\rfloor\cdot)$ is defined on $\mathscr{C}(\Omega)$, and its restriction to $\mathscr{C}$ dominates $\underline{P}(\cdot\rfloor\cdot)$, in the sense that

$$\underline{E}(f\rfloor B) \geq \underline{P}(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}. \tag{2.12}$$

Let us begin by taking a look at what happens if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent. In that case, $\underline{E}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision, and we will refer to it as the *natural extension* of $\underline{P}(\cdot\rfloor\cdot)$. Not only does it—if necessary—correct $\underline{P}(\cdot\rfloor\cdot)$ upwards on $\mathscr{C}$ to make it coherent, it also extends the domain of this correction to all of $\mathscr{C}(\Omega)$. Furthermore, out of all such coherent upwards corrections of $\underline{P}(\cdot\rfloor\cdot)$, $\underline{E}(\cdot\rfloor\cdot)$ is the most conservative—most imprecise—one and therefore the only one that can always be inferred from $\underline{P}(\cdot\rfloor\cdot)$ without having to add extra assessments.

Proposition 2. *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ and let $\underline{P}'(\cdot\rfloor\cdot)$ be any coherent conditional lower prevision on $\mathscr{C}' \supseteq \mathscr{C}$ that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Then*

$$\underline{P}'(f\rfloor B) \geq \underline{E}(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}'.$$

Proof of Proposition 2. By Definition 2_{42}, there is some coherent set of desirable gambles $\mathscr{D}$ such that $\underline{P}_{\mathscr{D}}(f\rfloor B) = \underline{P}'(f\rfloor B)$ for all $(f,B) \in \mathscr{C}'$. Furthermore, by an argument similar to the one we provided for $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ in the main text, we know that $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)}$ consists of gambles whose desirability is implied by $\underline{P}'(\cdot\rfloor\cdot)$. Hence, we find that $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)} \subseteq \mathscr{D}$. Also, since $\underline{P}'(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, $\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$ is clearly a subset of $\mathscr{A}_{\underline{P}'(\cdot\rfloor\cdot)}$ and therefore $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is a subset of $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)}$. This implies that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} \subseteq \mathscr{D}$ and therefore also, by Equation $(2.9)_{42}$, that $\underline{P}'(f\rfloor B) = \underline{P}_{\mathscr{D}}(f\rfloor B) \geq \underline{P}_{\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}}(f\rfloor B) = \underline{E}(f\rfloor B)$ for all $(f,B) \in \mathscr{C}'$. □

We conclude that, if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent, the natural extension $\underline{E}(\cdot\rfloor\cdot)$ provides us with the most conservative—lowest—coherent supremum buying prices that are compatible with (the assessment that corresponds to) $\underline{P}(\cdot\rfloor\cdot)$.

So far, so good. But what if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent? As we know from Section $2.2.1_{39}$, the only way for this to happen is if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid null gain [$D1_{38}$], meaning that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ contains the zero gamble. Even worse, as we show in Proposition 3 below, there are $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ for which $f < 0$. In other words, there are gambles whose desirability is implied by $\underline{P}(\cdot\rfloor\cdot)$ and $D2_{38}$–$D4_{39}$, but which are guaranteed never to yield a positive payoff, and in some cases even yield a negative payoff. If this happens, then clearly, there is something wrong with $\underline{P}(\cdot\rfloor\cdot)$. Indeed, it turns out that if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent, then $\underline{P}(\cdot\rfloor\cdot)$ is incoherent as well, and it cannot be made coherent by correcting it upwards. Hence, in that case, it is not possible to construct a coherent lower prevision that is consistent with (the non-exhaustive interpretation of) $\underline{P}(\cdot\rfloor\cdot)$, and the only option is to reassess $\underline{P}(\cdot\rfloor\cdot)$.

Proposition 3. *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. Then the following statements are equivalent:*

(i) *$\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent;*

(ii) *$f < 0$ for some $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$;*

(iii) *$\underline{E}(f\rfloor B) = +\infty$ for some $(f,B) \in \mathscr{C}$;*

(iv) *Every conditional lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$ is incoherent.*

Proof of Proposition 3. It clearly suffices to show that (i)⇒(ii)⇒(i), (i)⇒(iii) and ¬(i)⇒ ¬(iv)⇒ ¬(iii).

We start by proving that (i)⇒(ii) and (i)⇒(iii). So assume that (i) holds: $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent. Then as explained in Section $2.2.1_{39}$, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid

null gain [D1$_{38}$], implying that there are $n \in \mathbb{N}$, $(\forall i \in \{1,\dots,n\})$ $(f_i, B_i) \in \mathscr{C}$, $\varepsilon_i \in \mathbb{R}_{>0}$, $\lambda_i \in \mathbb{R}_{>0}$ such that

$$g := \sum_{i=1}^{n} \lambda_i \mathbb{I}_{B_i}[f_i - \underline{P}(f_i \rfloor B_i) + \varepsilon_i] = 0.$$

Hence, for all $\lambda \in \mathbb{R}_{>0}$, we find that

$$\begin{aligned} -\lambda \mathbb{I}_{B_1} = -\frac{2\lambda}{\lambda_1 \varepsilon_1} \lambda_1 \mathbb{I}_{B_1} \frac{\varepsilon_1}{2} &\geq \frac{2\lambda}{\lambda_1 \varepsilon_1} \left(-\sum_{i=1}^{n} \lambda_i \mathbb{I}_{B_i} \frac{\varepsilon_i}{2} + g \right) \\ &= \frac{2\lambda}{\lambda_1 \varepsilon_1} \left(\sum_{i=1}^{n} \lambda_i \mathbb{I}_{B_i} \left[f_i - \underline{P}(f_i \rfloor B_i) + \frac{\varepsilon_i}{2} \right] \right) =: h \end{aligned}$$

and therefore, since $h \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, and because $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D2$_{38}$ and D4$_{39}$ and therefore also D5$_{39}$, that $-\lambda \mathbb{I}_{B_1} \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. This already implies that (ii)$_{\frown}$ holds. Now assume *ex absurdo* that (iii)$_{\frown}$ does not hold, implying in particular that $\underline{E}(f_1 \rfloor B_1) < \infty$, which in turn implies that there is some $\alpha \in \mathbb{R}$ such that $\mathbb{I}_{B_1}(f_1 - \alpha) \notin \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. Consider any $\lambda \in \mathbb{R}_{>0}$ such that $\lambda > \alpha - \min f_1$. Then $\mathbb{I}_{B_1}(f_1 - \alpha) \geq -\lambda \mathbb{I}_{B_1}$ and hence, since (as we have just shown) $-\lambda \mathbb{I}_{B_1} \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, and because $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D5$_{39}$, we find that $\mathbb{I}_{B_1}(f_1 - \alpha) \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. This is a contradiction, allowing us to conclude that (iii)$_{\frown}$ holds.

(ii)$_{\frown}$⇒(i)$_{\frown}$ is trivial: by combining (ii)$_{\frown}$ with the fact that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D5$_{39}$, we immediately find that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid null gain [D1$_{38}$].

For ¬(i)$_{\frown}$⇒ ¬(iv)$_{\frown}$, it suffices to recall that, as explained earlier on in Section 2.3.2$_{44}$, coherence of $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ implies that $\underline{E}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.

Finally, we prove that ¬(iv)$_{\frown}$⇒ ¬(iii)$_{\frown}$. So assume that ¬(iv)$_{\frown}$: there is a coherent conditional lower prevision $\underline{P}'(\cdot\rfloor\cdot)$ that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Then by Proposition 2$_{\frown}$, for any $(f,B) \in \mathscr{C}$, $\underline{P}'(f\rfloor B) \geq \underline{E}(f\rfloor B)$. Since $\underline{P}'(\cdot\rfloor\cdot)$ is by definition real-valued, this implies that $\underline{E}(f\rfloor B) < \infty$ for all $(f,B) \in \mathscr{C}$. □

By combining this result with Proposition 2$_{\frown}$, we obtain the following alternative characterisation of coherence.

Corollary 4. *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. Then $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if it coincides with $\underline{E}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.*

Proof of Corollary 4. First assume that $\underline{P}(\cdot\rfloor\cdot)$ is coherent. Then trivially, $\underline{P}(\cdot\rfloor\cdot)$ is a coherent lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Hence, by Proposition 2$_{\frown}$, $\underline{P}(\cdot\rfloor\cdot)$ dominates $\underline{E}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. However, by Equation (2.12)$_{44}$, the converse holds as well. Hence, $\underline{P}(\cdot\rfloor\cdot)$ and $\underline{E}(\cdot\rfloor\cdot)$ coincide on $\mathscr{C}$.

Next, assume that $\underline{P}(\cdot\rfloor\cdot)$ and $\underline{E}(\cdot\rfloor\cdot)$ coincide on $\mathscr{C}$. Now assume *ex absurdo* that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent. Then by Proposition 3(i)$_{\frown}$&(iii)$_{\frown}$, there is some $(f,B) \in \mathscr{C}$ such that $\underline{E}(f\rfloor B) = +\infty$ and therefore, by assumption, also

$\underline{P}(f\rfloor B) = +\infty$. This is a contradiction because $\underline{P}(\cdot\rfloor\cdot)$ is a conditional lower prevision and therefore, by definition, real-valued. Hence, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent. This implies that $\underline{E}(\cdot\rfloor\cdot)$ is coherent too, and therefore, since $\underline{P}(\cdot\rfloor\cdot)$ is by assumption the restriction of $\underline{E}(\cdot\rfloor\cdot)$ to $\mathscr{C}$, $\underline{P}(\cdot\rfloor\cdot)$ is coherent as well. □

If $\underline{P}(\cdot\rfloor\cdot)$ is coherent then, as one would intuitively expect, the most conservative coherent upwards correction of $\underline{P}(\cdot\rfloor\cdot)$ is $\underline{P}(\cdot\rfloor\cdot)$ itself. However, even in that case, natural extension is still a important tool, as it allows us to coherently extend $\underline{P}(\cdot\rfloor\cdot)$ to all of $\mathscr{C}(\Omega)$ in the most conservative way possible.

Consider for example the common case where $\mathscr{C} := \{(f,\Omega)\colon f \in \mathscr{G}(\Omega)\}$, which means that $\underline{P}(\cdot\rfloor\cdot)$ is an unconditional lower prevision $\underline{P}(\cdot) := \underline{P}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$. In that case, natural extension can be regarded as a conditioning rule: for any $B \in \mathscr{P}_\emptyset(\Omega)$, it provides us with a corresponding coherent lower prevision $\underline{E}(\cdot\rfloor B)$ on $\mathscr{G}(B)$;[11] it is guaranteed to be coherent with $\underline{P}$, and out of all lower previsions on $\mathscr{G}(B)$ that are coherent with $\underline{P}$, it is the most conservative—most imprecise—one.

2.3.3 *Comparison with other approaches*

Besides Definition 2_{42}, many other definitions for coherence have been proposed as well; see Reference [80] for an overview. As we are about to explain, within our finitary context—recall that Ω is finite—, many of them are mathematically equivalent to our approach.

First of all, in these other definitions, $\underline{P}(\cdot|B)$ is defined for gambles on Ω instead of B. By analogy with conditioning for sets of desirable gambles, we reflect this in our notation by using '$|$' instead of '$\rfloor$'. If we adopt this approach as well, then $\underline{P}_{\mathscr{D}}(f|B)$ is still defined by Equation $(2.10)_{42}$, but with f an element of $\mathscr{G}(\Omega)$ instead of $\mathscr{G}(B)$. Definition 2_{42} remains identical; we just have to replace '$\rfloor$' by '$|$'. We prefer our present version because we find it more intuitive that conditioning on an event B produces a model for—a lower prevision for gambles on—that event. Also, it allows us to use Equation $(2.9)_{42}$, which we think is particularly elegant because it illustrates that Equation $(2.10)_{42}$ follows directly from Equations $(2.2)_{40}$ and $(2.3)_{40}$. Mathematically, both approaches are equivalent. If $\underline{P}(\cdot|\cdot)$ is coherent, then $\underline{P}(f|B)$ depends only on the restriction $f_B \in \mathscr{G}(B)$ of f to B, allowing us to identify $\underline{P}(\cdot|\cdot)$ with a coherent lower prevision $\underline{P}(\cdot\rfloor\cdot)$. Conversely, if $\underline{P}(\cdot\rfloor\cdot)$ is coherent, then the lower prevision $\underline{P}(\cdot|\cdot)$ that is defined by

$$\underline{P}(f|B) := \underline{P}(f_B\rfloor B) \text{ for all } f \in \mathscr{G}(\Omega) \text{ and } B \in \mathscr{P}_\emptyset(\Omega) \text{ such that } (f_B, B) \in \mathscr{C}$$

is coherent. Even stronger: $\underline{P}(\cdot|\cdot)$ is coherent if and only if $\underline{P}(\cdot\rfloor\cdot)$ is coherent. Given this connection, we now compare Definition 2_{42} with a number of alternative definitions for coherence.

[11] We provide an explicit expression for this conditioning rule further on in Section $2.7.2_{61}$.

Formally, the notion of coherence that resembles Definition 2_{42} the most is that of Williams [104, 112] (W-coherence). The main difference is that he considers so-called acceptable gambles rather than desirable ones. As was essentially pointed out by Williams himself—the cone $\mathscr{A}_0'$ in Reference [113, Proposition 3] is identical to our $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$—, this leads to an equivalent definition; see Reference [74] for some results in terms of desirable gambles. Two other differences are that Williams considers upper rather than lower previsions and that he imposes some structure on the domain of $\underline{P}(\cdot|\cdot)$.[12] Reference [80, Section 3.1] explains that this does not make any difference either; structure-free generalisations of W-coherence for lower previsions can be found in References [80, 94, 98] and [96, Chapter 13]. Another, more popular definition of coherence is that of Walley [106] (Walley-coherence); this definition is not structure-free. When it is applicable, Walley-coherence is known to be equivalent to W-coherence if Ω is finite [106, Appendix K]. We conclude that within our finitary context, Definition 2_{42} is equivalent to both Walley- and W-coherence and that therefore, we can import all sorts of useful results that were developed for these other notions of coherence, ranging from theoretical properties [96, 106] to computational techniques [98, 111].

2.3.4 *Properties of coherent conditional lower previsions*

Coherence has many useful consequences; see for example References [96, 106, 112]. We list only a few.

Consider a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with arbitrary domain $\mathscr{C}$ and let $\overline{P}(\cdot\rfloor\cdot)$ be the associated conditional upper prevision, defined by

$$\overline{P}(f\rfloor B) := -\underline{P}(-f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}(\Omega) \text{ such that } (-f,B) \in \mathscr{C}.$$

Then for all $A,B \in \mathscr{P}_{\emptyset}(\Omega)$ such that $B \subseteq A$, all $\lambda \in \mathbb{R}_{\geq 0}$, all $\mu \in \mathbb{R}$ and all $f,g \in \mathscr{G}(B)$, the following properties hold whenever the expressions involved are well-defined:

C1. $\underline{P}(f\rfloor B) \geq \min f$;

C2. $\underline{P}(\lambda f\rfloor B) = \lambda \underline{P}(f\rfloor B)$; [non-negative homogeneity]

C3. $\underline{P}(f+g\rfloor B) \geq \underline{P}(f\rfloor B) + \underline{P}(g\rfloor B)$; [super-additivity]

C4. $\underline{P}(\mathbb{I}_B[f - \underline{P}(f\rfloor B)]\rfloor A) = 0$;

C5. $\overline{P}(f\rfloor B) + \overline{P}(g\rfloor B) \geq \overline{P}(f+g\rfloor B) \geq \overline{P}(f\rfloor B) + \underline{P}(g\rfloor B) \geq \underline{P}(f+g\rfloor B)$;

C6. $\underline{P}(f\rfloor B) \leq \overline{P}(f\rfloor B) \leq \max f$;

[12] He requires the domain to be of the form in Corollary 6_{50}.

C7. $g \geq f \Rightarrow \underline{P}(g\rfloor B) \geq \underline{P}(f\rfloor B)$; [monotonicity]

C8. $\underline{P}(f+\mu\rfloor B) = \underline{P}(f\rfloor B)+\mu$. [constant additivity]

Conditions C1$_\frown$–C4$_\frown$ have a special status because—as the following result by Williams establishes—they can be used to characterise coherence.

Proposition 5 ([96, 106, 112])**.** *Consider a non-empty subset $\mathscr{B}$ of $\mathscr{P}_\emptyset(\Omega)$ and, for all $B \in \mathscr{B}$, a linear subspace $\mathscr{K}_B$ of $\mathscr{G}(B)$. Let $\underline{P}(\cdot\rfloor\cdot)$ be any conditional lower prevision with domain $\mathscr{C} := \{(f,B)\colon B \in \mathscr{B} \text{ and } f \in \mathscr{K}_B\}$. Then $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if it satisfies* C1$_\frown$–C4$_\frown$*. Furthermore, for any $B \in \mathscr{B}$, $\underline{P}(\cdot\rfloor B)$ is coherent if and only if it satisfies* C1$_\frown$–C3$_\frown$*. Hence, in this particular case, $\underline{P}(\cdot\rfloor\cdot)$ is jointly coherent if and only if it is separately coherent and satisfies* C4$_\frown$*.*

2.3.5 *Betting rates*

Before we move on to the connection between lower previsions and probability measures, we would like to draw attention to a particular aspect of sets of desirable gambles and the lower previsions that are associated with them.

Clearly, the notion of a gamble is closely related to betting: similarly to what happens with betting, we either lose or gain utility (money), depending on the uncertain value of a variable X (the outcome of some experiment). This connection is especially clear if we consider indicators of events. For any $B \in \mathscr{P}_\emptyset(\Omega)$ and $\lambda \in \mathbb{R}_{>0}$ and $\alpha \in \mathbb{R}$, the gamble $\lambda(\mathbb{I}_B - \alpha)$ corresponds to paying $\lambda\alpha$ in order to receive λ if B happens. In other words: betting on B, at a betting rate α, and with stakes λ. Due to the linearity of our utility scale [C2$_\frown$], the desirability of such a bet does not depend on the stakes λ, but only on the betting rate α: the gamble $\mathbb{I}_B - \alpha$ is desirable if and only if you are willing to bet on B at a betting rate α. Hence, we find that the supremum betting rate at which you are willing to bet on B, defined by

$$\underline{P}_{\mathscr{D}}(B) := \sup\bigl\{\alpha \in \mathbb{R}\colon \mathbb{I}_B - \alpha \in \mathscr{D}\bigr\} = \underline{P}_{\mathscr{D}}(\mathbb{I}_B), \tag{2.13}$$

is equal to the lower prevision of $\mathbb{I}_B$. Similarly, since $\alpha - \mathbb{I}_B$ is desirable if and only if you are willing to *take* bets on B at a betting rate α—bet against B at a betting rate $1-\alpha$—we find that the upper prevision of $\mathbb{I}_B$ is equal to the infimum betting rate

$$\overline{P}_{\mathscr{D}}(B) := \overline{P}_{\mathscr{D}}(\mathbb{I}_B) = \inf\bigl\{\alpha \in \mathbb{R}\colon \alpha - \mathbb{I}_B \in \mathscr{D}\bigr\} \tag{2.14}$$

at which you are willing to take bets on B. By coherence [C1$_\frown$ and C6$_\frown$], we find that

$$0 \leq \underline{P}_{\mathscr{D}}(B) \leq \overline{P}_{\mathscr{D}}(B) \leq 1, \tag{2.15}$$

as is to be expected for (supremum and infimum) betting rates. For reasons that should become clear shortly, $\underline{P}_{\mathscr{D}}(B)$ and $\overline{P}_{\mathscr{D}}(B)$—or $\underline{P}(B)$ and $\overline{P}(B)$—are often

referred to as the lower and upper probability of B, respectively. However, this should not be taken to imply the existence of an unknown probability $P(B)$ of B, for which $\underline{P}_{\mathscr{D}}(B)$ and $\overline{P}_{\mathscr{D}}(B)$ provide lower and upper bounds; this may be the case in some situations, but in general, the interpretation in terms of betting rates is more fundamental. We discuss this point further in Section 2.6.3$_{56}$.

2.4 LINEAR PREVISIONS

If a subject's supremum rate $\underline{P}(B)$ for betting on B coincides with the infimum rate $\overline{P}(B)$ at which he is willing to take bets on B, then $P(B) := \underline{P}(B) = \overline{P}(B)$ is his fair betting rate for the event B. Similarly, for a gamble $f \in \mathscr{G}(\Omega)$, if $\underline{P}(f) = \overline{P}(f)$, then $P(f) := \underline{P}(f) = \overline{P}(f)$ is the subject's fair price for f, called the prevision of f by de Finetti [50]. If a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ assigns such fair prices to all gambles, in the sense that $(f,B) \in \mathscr{C}$ if and only if $(-f,B) \in \mathscr{C}$ and that

$$\underline{P}(f\rfloor B) = \overline{P}(f\rfloor B) =: -\underline{P}(-f\rfloor B) \text{ for all } (f,B) \in \mathscr{C},$$

then $\underline{P}(\cdot\rfloor\cdot)$ is said to be *self-conjugate*, is referred to as a *conditional prevision*, and we then simply write $P(\cdot\rfloor\cdot)$ instead of $\underline{P}(\cdot\rfloor\cdot)$ or $\overline{P}(\cdot\rfloor\cdot)$.

If a conditional prevision $P(\cdot\rfloor\cdot)$ is *coherent*, then by combining C1$_{48}$–C4$_{48}$ with self-conjugacy, we find that it satisfies the following properties. For all $A,B \in \mathscr{P}_{\emptyset}(\Omega)$ such that $B \subseteq A$, all $\lambda \in \mathbb{R}$, and all $f,g \in \mathscr{G}(B)$, and whenever the expressions involved are well-defined:

P1. $\min f \leq P(f\rfloor B) \leq \max f$;

P2. $P(\lambda f\rfloor B) = \lambda P(f\rfloor B)$; [homogeneity]

P3. $P(f+g\rfloor B) = P(f\rfloor B) + P(g\rfloor B)$; [additivity]

P4. $P(\mathbb{I}_B f\rfloor A) = P(f\rfloor B)P(B\rfloor A)$. [Bayes's rule]

As we can see from conditions P2 and P3, for all $B \in \mathscr{P}_{\emptyset}(\Omega)$, $P(\cdot\rfloor B)$ is a linear operator, and for this reason, coherent conditional previsions are also referred to as conditional *linear* previsions. Proposition 5$_{\frown}$ leads to the following convenient characterisation.

Corollary 6 ([96, 106, 112]). *Consider a non-empty subset $\mathscr{B}$ of $\mathscr{P}_{\emptyset}(\Omega)$ and, for all $B \in \mathscr{B}$, a linear subspace $\mathscr{K}_B$ of $\mathscr{G}(B)$. Let $P(\cdot\rfloor\cdot)$ be any conditional prevision with domain $\mathscr{C} := \{(f,B) \colon B \in \mathscr{B} \text{ and } f \in \mathscr{K}_B\}$. Then $P(\cdot\rfloor\cdot)$ is coherent if and only if it satisfies* P1–P4. *Furthermore, for any $B \in \mathscr{B}$, $P(\cdot\rfloor B)$ is coherent if and only if it satisfies* P1–P3. *Hence, in this particular case, $P(\cdot\rfloor\cdot)$ is jointly coherent if and only if it is separately coherent and satisfies* P4.

We denote the set of all conditional linear previsions on $\mathscr{C}(\Omega)$ by $\mathbb{P}$. Furthermore, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$, we let $\mathbb{P}_B$ be the set consisting of all *unconditional* linear previsions on $\mathscr{G}(B)$. In this way, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$ and $P(\cdot\rfloor\cdot) \in \mathbb{P}$, we have that $P(\cdot\rfloor B) \in \mathbb{P}_B$. $\mathbb{P}_\Omega$ corresponds to an important special case.

2.5 FULL CONDITIONAL PROBABILITY MASS FUNCTIONS

One of the reasons why linear previsions are an important, is because they allow us to link the gamble-orientated approach to modelling uncertainty, which we have just introduced, with the more conventional approach that uses probability measures and probability mass functions.

Indeed, consider a conditional linear prevision $P(\cdot\rfloor\cdot)$ on $\mathscr{C}(\Omega)$ and use $P(C\rfloor B)$ as a shorthand notation for $P(\mathbb{I}_C\rfloor B)$, for all $C, B \in \mathscr{P}(\Omega)$ such that $C \subseteq B$. It then follows from P1$_\cap$–P4$_\cap$ that, for all $A, B \in \mathscr{P}_{\emptyset}(\Omega)$ and $C, D \in \mathscr{P}(\Omega)$ such that $C \cap D = \emptyset$ and $C \cup D \subseteq B \subseteq A$:

F1. $P(B\rfloor B) = 1$;

F2. $P(C\rfloor B) \geq 0$;

F3. $P(C \cup D\rfloor B) = P(C\rfloor B) + P(D\rfloor B)$;

F4. $P(C\rfloor A) = P(C\rfloor B)P(B\rfloor A)$. [Bayes's rule]

Hence, formally, the restriction of $P(\cdot\rfloor\cdot)$ to indicators can be identified with a full conditional probability measure, because F1–F4 are the defining properties for such a measure [54].[13] Conditions F1–F3 assert that, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$, $P(\cdot\rfloor B)$ satisfies the usual axioms of probability, and F4 corresponds to Bayes's rule. Furthermore, the original conditional prevision can be fully recovered from this full conditional probability measure. Indeed, by linearity, for any $f \in \mathscr{G}(B)$, since $f = \sum_{\omega \in B} f(\omega)\mathbb{I}_{\{\omega\}}$, we find that

$$P(f\rfloor B) = P\left(\sum_{\omega \in B} f(\omega)\mathbb{I}_{\{\omega\}}\Big\rfloor B\right) = \sum_{\omega \in B} f(\omega)P(\mathbb{I}_{\{\omega\}}\rfloor B) = \sum_{\omega \in B} f(\omega)p(\omega\rfloor B)$$

is the expected value of the gamble f with respect to the probability mass function $p(\cdot\rfloor B)$ on B that corresponds to $P(\cdot\rfloor B)$, defined for all $\omega \in B$ by $p(\omega\rfloor B) := P(\{\omega\}\rfloor B) = P(\mathbb{I}_{\{\omega\}}\rfloor B)$.

Inspired by these results, we let $\mathscr{C}_*(\Omega) := \{(\omega, B) \colon B \in \mathscr{P}_{\emptyset}(\Omega) \text{ and } \omega \in B\}$ and we call an operator $p(\cdot\rfloor\cdot)$ on $\mathscr{C}_*(\Omega)$ a full conditional probability mass

[13] See Reference [21] for a recent overview of related literature.

function if and only if, for all $A, B \in \mathscr{P}_{\emptyset}(\Omega)$ such that $B \subseteq A$, $p(\cdot\rfloor B)$ is a probability mass function on B and

$$p(\omega\rfloor A) = p(\omega\rfloor B) \sum_{\omega' \in B} p(\omega'\rfloor A) \text{ for all } \omega \in B. \tag{2.16}$$

It follows from the results above that there is a one-to-one correspondence between conditional linear previsions, full conditional probability measures and full conditional probability mass functions.

If every element of Ω has positive probability—if $p(\omega\rfloor\Omega) > 0$ for all $\omega \in \Omega$—then Bayes's rule—or, equivalently, Equation (2.16)—uniquely determines all conditional probabilities. In that case, a full conditional probability mass function and its associated full conditional probability measure and conditional linear prevision are completely characterised by the unconditional mass function $p(\cdot) := p(\cdot\rfloor\Omega)$. However, if $P(B) = \sum_{\omega \in B} p(\omega\rfloor B) = 0$, Bayes's rule imposes no non-trivial restrictions on $p(\cdot\rfloor B)$ and in that case, a full conditional probability mass function allows for $p(\cdot\rfloor B)$ to be specified independently from $p(\cdot\rfloor\Omega)$.

2.6 SETS OF LINEAR PREVISIONS OR MASS FUNCTIONS

The link between (full conditional) probability mass functions and (conditional) lower previsions is not restricted to the special case of (conditional) linear previsions. In general, lower previsions are related to *sets* of linear previsions and therefore, by the results in the previous section, to sets of probability mass functions.

2.6.1 *The lower envelope theorem*

For any conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on an arbitrary domain $\mathscr{C}$, we can consider the corresponding set of dominating linear conditional previsions, as given by

$$\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)} := \{P(\cdot\rfloor\cdot) \in \mathbb{P} \colon P(f\rfloor B) \geq \underline{P}(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}\}. \tag{2.17}$$

The following fundamental result by Williams [112] shows that $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if (a) there is at least one such dominating linear prevision and (b) the lower envelope of all these dominating linear previsions is equal to $\underline{P}(\cdot\rfloor\cdot)$.

Theorem 7 (Lower envelope theorem [112])**.** *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. We then have that $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)} \neq \emptyset$ and*

$$\underline{P}(f\rfloor B) = \min\{P(f\rfloor B) \colon P(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\} \textit{ for all } (f,B) \in \mathscr{C}.$$

Furthermore, in that case, by conjugacy, the corresponding conditional upper prevision is given by

$$\overline{P}(f\rfloor B) = \max\{P(f\rfloor B)\colon P(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\} \text{ for all } (-f,B) \in \mathscr{C}.$$

By combining the lower envelope theorem with the results in Section 2.3.2$_{44}$, we obtain an alternative expression for the natural extension: $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent if and only $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)} \neq \emptyset$ and, in that case, we have that

$$\underline{E}(f\rfloor B) = \min\{P(f\rfloor B)\colon P(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\} \text{ for all } (f,B) \in \mathscr{C}(\Omega). \tag{2.18}$$

Although the set $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ is extremely powerful from a theoretical point of view, it is often too complex to work with in practice. For this reason, it is sometimes convenient to restrict the domain of the conditional linear previsions in $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$. In particular, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$, we may consider the set

$$\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B := \{P(\cdot\rfloor B)\colon P(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\} \subseteq \mathbb{P}_B, \tag{2.19}$$

consisting of linear previsions on $\mathscr{G}(B)$. Alternatively, instead of restricting the domain of the conditional previsions in $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$, we may also regard $\underline{P}(\cdot\rfloor B)$ as an unconditional lower prevision on $\mathscr{C}_B := \{f \in \mathscr{G}(B)\colon (f,B) \in \mathscr{C}\}$—provided that $\mathscr{C}_B \neq \emptyset$—and consider the set of all linear previsions on $\mathscr{G}(B)$ that locally dominate $\underline{P}(\cdot\rfloor B)$:

$$\mathscr{M}_{\underline{P}(\cdot\rfloor B)} := \{P \in \mathbb{P}_B\colon P(f) \geq \underline{P}(f\rfloor B) \text{ for all } f \in \mathscr{C}_B\}.$$

The following result establishes that it does not really matter which road we take. If the domain of $\underline{P}(\cdot\rfloor\cdot)$ is sufficiently large, $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B$ and $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ coincide.

Proposition 8. *Consider an event $B \in \mathscr{P}_{\emptyset}(\Omega)$ and a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ such that $\mathscr{C}_B = \mathscr{G}(B)$. It then holds that $\mathscr{M}_{\underline{P}(\cdot\rfloor B)} = \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B$.*[14]

Proof of Proposition 8. We only prove that $\mathscr{M}_{\underline{P}(\cdot\rfloor B)} \subseteq \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B$. The converse inclusion holds trivially. Consider therefore any $P \in \mathscr{M}_{\underline{P}(\cdot\rfloor B)}$. We show that $P \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B$.

Let $\underline{P}^*(\cdot\rfloor\cdot)$ be the conditional lower prevision on $\mathscr{C}$ that is defined by

$$\underline{P}^*(f\rfloor A) := \begin{cases} \underline{P}(f\rfloor A) & \text{if } A \neq B \\ P(f) & \text{if } A = B \end{cases} \quad \text{for all } (f,A) \in \mathscr{C}$$

[14] Many thanks to Enrique Miranda. I still remember asking him if he knew whether this result was true. He said he did not know. Later the same day, while strolling through town in search for a beer, he handed me a folded sheet of paper. On it, he had written down the central idea of the proof I provide here.

and consider any $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$. We set out to prove that $g \not< 0$. If $g > 0$, this is trivial. If $g \not> 0$, we infer from $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$ that there are $n \in \mathbb{N}$ and, for all $i \in \{1,\dots,n\}$, $\lambda_i \in \mathbb{R}_{>0}$, $(f_i, A_i) \in \mathscr{C}$ and $\varepsilon_i \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} g &\geq \sum_{i=1}^{n} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] \\ &= \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] + \sum_{i\in I'} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] \\ &= \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \sum_{i\in I'} \lambda_i \mathbb{I}_{B}[f_i - P(f_i) + \varepsilon_i], \end{aligned}$$

with $I := \{i \in \{1\dots n\} \colon A_i \neq B\}$ and $I' := \{i \in \{1\dots n\} \colon A_i = B\}$. If $I' = \emptyset$—and therefore $I = \{1,\dots,n\}$—we find that $g \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and therefore, by Proposition 3_{45} and the fact that $\underline{P}(\cdot\rfloor\cdot)$ is coherent, that $g \not< 0$. Hence, we may assume that $I \neq \emptyset$, allowing us to define $f := \sum_{i\in I'} \lambda_i f_i$ and $\varepsilon := \sum_{i\in I'} \lambda_i \varepsilon_i$. By the linearity of P, and the fact that $P \in \mathscr{M}_{\underline{P}(\cdot\rfloor B)}$, we now have that

$$\begin{aligned} g &\geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + P(-f) + \varepsilon] \\ &\geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + \underline{P}(-f\rfloor B) + \varepsilon], \end{aligned}$$

where $\underline{P}(-f\rfloor B)$ is well-defined because $\mathscr{C}_B = \mathscr{G}(B)$ and therefore $(-f, B) \in \mathscr{C}$. Since $\underline{P}(\cdot\rfloor\cdot)$ is coherent, we infer from the lower envelope theorem [Theorem 7_{52}] that there is a linear conditional prevision $P'(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ such that $\underline{P}(-f\rfloor B) = P'(-f\rfloor B) = -P'(f\rfloor B)$ and therefore also

$$\begin{aligned} g &\geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + \underline{P}(-f\rfloor B) + \varepsilon] \\ &\geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - P'(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f - P'(f\rfloor B) + \varepsilon], \end{aligned}$$

implying that $g \in \mathscr{E}_{P'(\cdot\rfloor\cdot)}$ and therefore, by Proposition 3_{45} and the fact that $P'(\cdot\rfloor\cdot)$ is coherent, that $g \not< 0$. Since we have proved that this holds for any $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$, we infer from Proposition 3_{45} that $\mathscr{M}_{\underline{P}^*(\cdot\rfloor\cdot)}$ is non-empty.

Now let $P^*(\cdot\rfloor\cdot)$ be any element of this non-empty set $\mathscr{M}_{\underline{P}^*(\cdot\rfloor\cdot)}$. We infer from $P \in \mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ that P dominates $\underline{P}(\cdot\rfloor B)$ on $\mathscr{C}_B = \mathscr{G}(B)$ and therefore, that $\underline{P}^*(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Furthermore, since $P^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^*(\cdot\rfloor\cdot)}$ implies that $P^*(\cdot\rfloor\cdot)$ dominates $\underline{P}^*(\cdot\rfloor\cdot)$ on $\mathscr{C}$, we find that $P^*(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, or equivalently, that $P^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$. Also, by the linearity of P and $P^*(\cdot\rfloor\cdot)$, we find that

$$P(f) = \underline{P}^*(f\rfloor B) \leq P^*(f\rfloor B) = -P^*(-f\rfloor B) \leq -\underline{P}^*(-f\rfloor B) = -P(-f) = P(f)$$

for all $f \in \mathscr{G}(B)$, implying that $P^*(\cdot\rfloor B) = P$. Since $P^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$, this in turn implies that $P \in \mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B$. □

Focusing on these 'local' sets $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ of linear previsions on $\mathscr{G}(B)$ is not only useful from a practical point of view. They satisfy a fundamental theoretical property as well. As was proved by Walley [106, Section 3.6.1], for any coherent lower prevision $\underline{P}$ on $\mathscr{G}(B)$, $\mathscr{M}_{\underline{P}}$ is closed and convex under the weak*-topology—the topology of pointwise convergence—and is furthermore the only such set that has $\underline{P}$ as its lower envelope. Hence, we find that $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ is the unique closed and convex subset of $\mathbb{P}_B$ that has $\underline{P}(\cdot\rfloor B)$ as its lower envelope.

One practical advantage of the fact that these local sets $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ are closed and convex is that it allows us to characterise them by means of their set of *extreme points* $\operatorname{ext}(\mathscr{M}_{\underline{P}(\cdot\rfloor B)})$—those elements of $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ that cannot be written as a proper convex combination of two other elements. In particular: $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ is the convex hull of $\operatorname{ext}(\mathscr{M}_{\underline{P}(\cdot\rfloor B)})$ [106, Section 3.6.2(b), note 5]. The most important consequence of this result is that $\underline{P}(\cdot\rfloor B)$ is the lower envelope of $\operatorname{ext}(\mathscr{M}_{\underline{P}(\cdot\rfloor B)})$ [106, Section 3.6.2(c)]:

$$\underline{P}(f\rfloor B) = \min\{P(f)\colon P \in \operatorname{ext}(\mathscr{M}_{\underline{P}(\cdot\rfloor B)})\} \quad \text{for all } f \in \mathscr{G}(B). \tag{2.20}$$

This is especially useful if $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ is *finitely generated*, by which we mean that it has a finite number of extreme points. By Equation (2.20), $\underline{P}(f\rfloor B)$ is then simply the minimum of a finite number of previsions.

2.6.2 *Credal sets*

Since there is a one-to-one correspondence between conditional linear previsions and full conditional probability mass functions, it follows that, for any coherent lower prevision $\underline{P}(\cdot\rfloor\cdot)$, the associated set $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ of conditional linear previsions on $\mathscr{C}(\Omega)$ has a unique corresponding set $\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}$ of full conditional probability mass functions on $\mathscr{C}_*(\Omega)$. Similarly, for any $B \in \mathscr{P}_\emptyset(\Omega)$, we let $\mathscr{F}_{\underline{P}(\cdot\rfloor B)}$ be the unique set of probability mass functions on B that corresponds to $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$. If we define

$$\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B := \{p(\cdot\rfloor B)\colon p(\cdot\rfloor\cdot) \in \mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}\}, \tag{2.21}$$

then by the results in the previous section and the one-to-one correspondence between linear previsions and probability mass functions, we know that $\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}\rfloor B = \mathscr{F}_{\underline{P}(\cdot\rfloor B)}$.

These 'local' sets of probability mass functions $\mathscr{F}_{\underline{P}(\cdot\rfloor B)}$ satisfy properties that are similar to those of $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$. Most importantly: $\mathscr{F}_{\underline{P}(\cdot\rfloor B)}$ is a convex and closed subset of $\mathbb{R}^B$, with respect to the natural topology, as induced by the Euclidean metric [63, Section 10.2].[15] Any such closed and convex set of

[15] Other, arguably more intuitive, metrics can be used as well; see References [32,92] for more information.

probability mass functions is called a *credal set* [66]. As is the case for every subset of $\mathbb{R}^B$ that is bounded, closed and convex, $\mathscr{F}_{\underline{P}(\cdot\rfloor B)}$ is the convex hull of its vertices [86, Corollary 18.5.1]—its extreme points. We use $\mathrm{ext}(\mathscr{F}_{\underline{P}(\cdot\rfloor B)})$ to refer to the set that consists of these vertices. If $\mathrm{ext}(\mathscr{F}_{\underline{P}(\cdot\rfloor B)})$ is finite, $\mathscr{F}_{\underline{P}(\cdot\rfloor B)}$ is said to be a *finitely generated* credal set.

2.6.3 *The sensitivity analysis interpretation*

The lower envelope theorem is not merely a theoretical result, it also suggests an alternative interpretation for coherent conditional lower previsions: a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ is simply a convenient representation for the set of linear conditional previsions $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$, or equivalently, the corresponding set of full conditional probability mass functions $\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}$. On this interpretation, one of these full conditional probability mass functions is believed to be the true, ideal model, but because of economic or time-constraints, or due to measurement errors, we are unable to specify it exactly, and can only provide a set of candidates. Walley [106, Section 1.1.5] refers to this belief in the existence of an ideal probability mass function as the dogma of *ideal precision*;[16] the corresponding interpretation of conditional lower previsions is called the *sensitivity analysis interpretation*.

However, care should be taken in adopting this interpretation. As we are about to demonstrate by means of two examples, there are many situations where it is not applicable. See References [106, 110] for extensive additional discussion.

From a frequentist point of view, the existence of a fair betting rate—probability—for an event follows from an hypothesis that the experiment at hand is part of an infinite sequence of—independent or exchangeable—analogous experiments, where analogous is taken to mean that they have the same distribution. This notion of fair betting rate—probability—is problematic because it requires a predefined notion of probability; in that sense, the frequentist argument is circular. In any case, this frequentist hypothesis is only tenable if the limiting frequency of the event actually converges.[17] Nevertheless, it is sometimes applicable. For example, if each of the experiments corresponds to a flip of the same fair coin, it seems reasonable to regard $P(H) = 1/2$ as a fair betting rate for heads. Similarly, if we are told that the coin is not fair, but that its probability for heads lies in between $1/4$ and $3/4$, we are led to consider the set of all linear previsions for which $1/4 \leq P(H) \leq 3/4$, which can be conveniently represented by the—in this case unique—coherent lower prevision for which $\underline{P}(H) = 1/4$ and $\overline{P}(H) = 3/4$. However, the limiting frequency

[16] We will refer to it as 'ideal precision', to avoid the provocative use of the term 'dogma'.

[17] Under this hypothesis, the limiting frequency convergences almost surely because of the law of large numbers.

does not always converge. Consider for example a case where we do not know how the instances of heads and tails are generated—it might be a coin flip, but it might also be by some other generating process—and where we observe that the limiting frequency of heads does not tend to converge, but rather cycles in between $1/4$ and $3/4$. In such a case, for a single experiment in this series, it seems again reasonable to bet on heads at any rate below $1/4$ and to take bets on heads at any rate above $3/4$, leading us to adopt $\underline{P}(\mathrm{H}) = 1/4$ and $\overline{P}(\mathrm{H}) = 3/4$ as our lower and upper prevision for (the indicator of) heads. However, in this case, there seems to be no reason to assume that there is such a thing as a fair (but unknown) betting rate $P(H)$ for betting on heads.[18]

A completely different, and rather extreme situation occurs when we want to model an experiment that is not repeated, and about which we know absolutely nothing, apart from the fact that the outcome is an element of Ω. In that case, for any gamble $f \in \mathscr{G}(\Omega)$, it would be sensible to buy f for any price below $\min f$, and sell it for any price higher than $\max f$, leading us to adopt the so-called *vacuous* lower prevision $\underline{P}_{\mathrm{v}}$ and the corresponding upper prevision $\overline{P}_{\mathrm{v}}$, defined by

$$\underline{P}_{\mathrm{v}}(f) = \min f \text{ and } \overline{P}_{\mathrm{v}}(f) = \max f \text{ for all } f \in \mathscr{G}(\Omega).$$

However, here too, there seems to be no reason to assume the existence of some subject's fair (but unknown) price $P(f)$, in the sense that he should be willing to buy f for any price below $P(f)$, and buy f for any price above.

2.6.4 *Arbitrary sets of linear previsions or mass functions*

If we choose to adopt the sensitivity analysis interpretation, and apply it to a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, we are led to model uncertainty by means of the set of linear conditional previsions $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ or the set of full conditional probability mass functions $\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}$. However, this is merely a special case. As we have seen in Sections 2.6.1$_{52}$ and 2.6.2$_{55}$, $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ and $\mathscr{F}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfy very specific properties.

If we adopt ideal precision as a principle on its own—without the sensitivity analysis interpretation—we are not required to restrict attention to sets that are of this particular form. In principle, uncertainty can be modelled by means of any set $\mathscr{M}$ of conditional linear previsions on $\mathscr{C}(\Omega)$ or any set $\mathscr{F}$ of full conditional probability mass functions on $\mathscr{C}_*(\Omega)$. They need not be conditional either. A set $\mathscr{M}$ of (unconditional) linear previsions or a set $\mathscr{F}$ of probability mass functions can also be used. By the one-to-one correspon-

[18] Such an assumption could be reasonable if there is some time-dependent fair betting rate, taking values in $[1/4, 3/4]$, of which the specific time evolution is unknown. However, we do not consider it reasonable to assume the existence of such a time-dependent fair betting rate based only on an observation that the limiting frequency cycles in between $1/4$ and $3/4$.

dence between these frameworks, these sets can be used interchangeably. We will mainly focus on the framework of (conditional) linear previsions.

We take $\mathscr{F}_{\mathscr{M}}$ to be the set of (full conditional) probability mass functions that corresponds to $\mathscr{M}$ and use $\mathscr{M}_{\mathscr{F}}$ to refer to the set of (conditional) linear previsions that corresponds to $\mathscr{F}$. Furthermore, if $\mathscr{M}$ is a set of conditional linear previsions on $\mathscr{C}(\Omega)$ and $\mathscr{F}$ is a set of full conditional probability mass functions on $\mathscr{C}_*(\Omega)$, then for any event $B \in \mathscr{P}_{\emptyset}(\Omega)$, we let

$$\mathscr{M}\rfloor B := \{P(\cdot\rfloor B)\colon P(\cdot\rfloor\cdot) \in \mathscr{M}\} \tag{2.22}$$

and

$$\mathscr{F}\rfloor B := \{p(\cdot\rfloor B)\colon p(\cdot\rfloor\cdot) \in \mathscr{F}\}$$

Equations $(2.19)_{53}$ and $(2.21)_{55}$ can be regarded as special cases of these definitions. It should also be clear that $\mathscr{M}_{\mathscr{F}}\rfloor B = \mathscr{M}_{\mathscr{F}\rfloor B}$ and $\mathscr{F}_{\mathscr{M}}\rfloor B = \mathscr{F}_{\mathscr{M}\rfloor B}$; this is a direct consequence of the one-to-one correspondence between (conditional) linear previsions and (full conditional) probability mass functions.

A link with lower previsions can still be established, even in this general case. With any set $\mathscr{M}$ of conditional linear previsions on $\mathscr{C}(\Omega)$, we can associate a coherent[19] conditional lower prevision $\underline{P}_{\mathscr{M}}(\cdot\rfloor\cdot)$, defined by

$$\underline{P}_{\mathscr{M}}(f\rfloor B) = \inf\{P(f\rfloor B)\colon P(\cdot\rfloor\cdot) \in \mathscr{M}\} \text{ for all } (f,B) \in \mathscr{C}(\Omega).$$

However, $\underline{P}_{\mathscr{M}}(\cdot\rfloor\cdot)$ is not guaranteed to represent $\mathscr{M}$, in the sense that $\mathscr{M}_{\underline{P}_{\mathscr{M}}(\cdot\rfloor\cdot)}$ might differ from $\mathscr{M}$. In general, we only have that $\mathscr{M} \subseteq \mathscr{M}_{\underline{P}_{\mathscr{M}}(\cdot\rfloor\cdot)}$. Similarly, the lower envelope $\underline{P}_{\mathscr{M}}$ of a set $\mathscr{M}$ of linear previsions on $\mathscr{G}(\Omega)$ may not represent $\mathscr{M}$; the set of unconditional previsions that dominates $\underline{P}_{\mathscr{M}}$ is guaranteed to include $\mathscr{M}$, but the inclusion might be strict; equality is obtained if and only if $\mathscr{M}$ is closed and convex.

As a simple example, consider a situation in which you have a biased coin, but you do not know in which direction it is biased. You only know that it is three times more likely to fall on one of its sides than on the other. In that case, it seems reasonable to model this situation by means of a set $\mathscr{M}$ consisting of two linear previsions P_1 and P_2, defined by $P_1(H) = 1/4$ and $P_2(H) = 3/4$, respectively. The corresponding lower prevision is determined by $\underline{P}_{\mathscr{M}}(H) = 1/4$ and $\overline{P}_{\mathscr{M}}(H) = 3/4$—we obtain the same lower prevision as in Section $2.6.3_{56}$. However, information is lost by using $\underline{P}_{\mathscr{M}}$ rather than $\mathscr{M}$, because $\underline{P}_{\mathscr{M}}(\cdot)$ is dominated not only by P_1 and P_2, but also by any convex combination of these two, including for example the linear prevision that corresponds to a fair coin.

[19] This follows from the fact that coherence is preserved under taking pointwise infima; see Proposition 1_{43}.

2.7 REGULAR EXTENSION VERSUS NATURAL EXTENSION

So far, we have come across two different imprecise-probabilistic methods for conditioning. For sets of desirable gambles, conditioning is fully determined by Equation $(2.2)_{40}$ and, for lower previsions, as explained at the end of Section $2.3.2_{44}$, natural extension can be regarded as a conditioning rule. We have not stressed this yet, but both methods have a surprising property: they are always well-defined, regardless of whether or not the conditioning event has (lower or upper) probability zero. Bayes's rule on the other hand, the most famous probabilistic conditioning rule of all time, is ill-defined whenever the conditioning event B has probability zero. Conditional linear previsions try to remedy this situation by allowing $P(\cdot\rfloor B)$ to be specified separately, but this does not resolve the issue, since the act of conditioning—deriving conditional models from unconditional ones—remains ill-defined: starting from an unconditional linear prevision P on $\mathscr{G}(B)$, with $P(B) = 0$, Bayes's rule places no restrictions on $P(\cdot\rfloor B)$.

Should this lead us to conclude that imprecise-probabilistic approaches are more powerful when it comes to dealing with probability zero? Yes indeed! Does it mean that we should forget about Bayes's rule? Not at all! As we are about to show, Bayes's rule has a prominent place within imprecise-probabilistic conditioning as well. In many cases, it even leads to a unique conditioning rule, which will then coincide with natural extension. In the remaining cases, Bayes's rule also leads to another imprecise-probabilistic conditioning rule, called regular extension. The goal of this section is to introduce this conditioning rule, to compare it with natural extension, and to discuss various related theoretical and computational aspects.

Since regular extension is especially intuitive from a sensitivity analysis point of view, we start by introducing it in terms of sets of linear previsions. Translations to the framework of probability mass functions are trivial and are therefore omitted. The connection with lower previsions and sets of desirable gambles will be established in Sections $2.7.2_{61}$ and $2.7.4_{65}$, respectively. Section $2.7.3_{63}$ discusses computational aspects.

2.7.1 *In terms of sets of linear previsions*

If we adopt ideal precision,[20] and model uncertainty by means of a set $\mathscr{M} \in \mathbb{P}_\Omega$ of—*unconditional*—linear previsions, conditioning on an event $B \in \mathscr{P}_\emptyset(\Omega)$ is commonly performed by conditioning each of the elements of $\mathscr{M}$ separately, through Bayes's rule. For any $P \in \mathscr{M}$ such that $P(B) > 0$, the resulting condi-

[20] See Sections $2.6.3_{56}$ and $2.6.4_{57}$.

tional prevision $P(\cdot\rfloor B)$ is defined by

$$P(f\rfloor B) := \frac{P(\mathbb{I}_B f)}{P(B)} \text{ for all } f \in \mathscr{G}(B). \tag{2.23}$$

If $P(B) > 0$ for all $P \in \mathscr{M}$, this approach leads to a unique conditional model $\mathscr{M}\rfloor B$, which is obtained by applying Equation (2.23) to every element of $\mathscr{M}$:

$$\mathscr{M}\rfloor B := \{P(\cdot\rfloor B) \colon P \in \mathscr{M}\} \subseteq \mathbb{P}_B. \tag{2.24}$$

The difference with Equation $(2.22)_{58}$ is that $\mathscr{M}$ is now a set of *unconditional* linear previsions. However, both equations are obviously closely related. If $\mathscr{M}'$ is a set of conditional linear previsions on $\mathscr{C}(\Omega)$ such that $P(B\rfloor\Omega) > 0$ for all $P(\cdot\rfloor\cdot) \in \mathscr{M}'$ and we let $\mathscr{M} := \mathscr{M}'\rfloor\Omega$ be the corresponding set of unconditional linear previsions on $\mathscr{G}(\Omega)$, then $\mathscr{M}\rfloor B = \mathscr{M}'\rfloor B$.

If $P(B) = 0$ for all $P \in \mathscr{M}$, then Bayes's rule imposes no restrictions on the conditional prevision $P(\cdot\rfloor B)$ and Equation (2.24) can no longer be applied. This leads us to consider the set of all previsions on $\mathscr{G}(B)$ as our conditional model: $\mathscr{M}\rfloor B := \mathbb{P}_B$.

The situation is less clear if $P(B) = 0$ for some $P \in \mathscr{M}$, but $P(B) > 0$ for others. We can then distinguish between two distinct approaches. By analogy with the corresponding notions for lower previsions—which will be discussed shortly—we call them natural and regular extension. Natural extension again considers the set $\mathbb{P}_B$ of all previsions on $\mathscr{G}(B)$, whereas regular extension ignores the previsions in $\mathscr{M}$ for which $P(B) = 0$, and applies Bayes's rule to the others.

By including the aforementioned cases as well, we obtain two different conditioning rules. *Natural extension* leads us to consider the conditional models that are given by

$$\mathscr{M}\rfloor^{\mathrm{n}} B := \begin{cases} \{P(\cdot\rfloor B) \colon P \in \mathscr{M}\} & \text{if } P(B) > 0 \text{ for all } P \in \mathscr{M}; \\ \mathbb{P}_B & \text{otherwise.} \end{cases} \tag{2.25}$$

Regular extension results in the use of the conditional sets that are defined by

$$\mathscr{M}\rfloor^{\mathrm{r}} B := \begin{cases} \{P(\cdot\rfloor B) \colon P \in \mathscr{M} \text{ and } P(B) > 0\} & \text{if } (\exists P \in \mathscr{M})\ P(B) > 0; \\ \mathbb{P}_B & \text{otherwise.} \end{cases} \tag{2.26}$$

The first part of this formula—the case where B has positive probability according to at least one P in $\mathscr{M}$—is called extended Bayesian conditioning in Reference [105]. It has also been called generalised (Bayesian) conditioning [11]. However, care should be taken when using this terminology, because these names are used to refer to other concepts as well; see for example References [65] and [10], respectively.

2.7.2 *In terms of lower previsions*

For lower previsions, the only conditioning rule that we have discussed so far is natural extension: for a given unconditional coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, and any event $B \in \mathscr{P}_{\emptyset}(\Omega)$, the natural extension $\underline{E}(\cdot\rfloor B)$ is the most conservative—most imprecise—conditional lower prevision on $\mathscr{G}(B)$ that is coherent with $\underline{P}$. Even stronger, as we will explain shortly, it is often the only one that is coherent with $\underline{P}$.

By Proposition 5_{49}, we know that a conditional lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{G}(B)$ is coherent with $\underline{P}$ if and only if both of them are separately coherent and if they satisfy $\mathrm{C4}_{48}$. In this particular context, with only these two lower previsions, $\mathrm{C4}_{48}$ reduces to

$$\underline{P}(\mathbb{I}_B[f - \underline{P}(f\rfloor B)]) = 0 \quad \text{for all } f \in \mathscr{G}(B), \tag{2.27}$$

which is referred to as the generalised Bayes rule (GBR) [106, Section 6.4]. One of the reasons why it has this name is because it reduces to Bayes's rule if $\underline{P}$ is a linear prevision; see $\mathrm{P4}_{50}$ and $\mathrm{F4}_{51}$. However, as we are about to show, the GBR has an even more fundamental connection with Bayes's rule.

If $\underline{P}(B) > 0$, the GBR is known to have a unique solution [106, Section 6.4.1]:[21] for any gamble $f \in \mathscr{G}(B)$, there is then a unique value of $\mu \in \mathbb{R}$ such that $\underline{P}(\mathbb{I}_B[f - \mu]) = 0$. Since $\underline{E}(\cdot\rfloor B)$ is coherent with $\underline{P}$, and therefore satisfies the GBR, this unique value coincides with $\underline{E}(f\rfloor B)$. In other words, if $\underline{P}(B) > 0$, $\underline{E}(\cdot\rfloor B)$ is the only coherent lower prevision on $\mathscr{G}(B)$ that satisfies the GBR, and therefore the only one that is jointly coherent with $\underline{P}$.

Let us now consider the set $\mathscr{M}_{\underline{P}}$ of linear previsions that dominate $\underline{P}$. If we adopt the sensitivity analysis interpretation, we can condition this set by means of the methods in the previous section. If $\underline{P}(B) > 0$, then $P(B) > 0$ for all $P \in \mathscr{M}_{\underline{P}}$, and we can simply apply Bayes's rule to each such P to obtain a linear prevision $P(\cdot\rfloor B)$ on $\mathscr{G}(B)$, leading us to adopt the set $\mathscr{M}_{\underline{P}}\rfloor B$ as our conditional model. Now let $\underline{P}(\cdot\rfloor B)$ be the lower envelope of this set. Then $\underline{P}(\cdot\rfloor B)$ is jointly coherent with $\underline{P}$, because for every $P \in \mathscr{M}_{\underline{P}}$, P and $P(\cdot\rfloor B)$ are jointly coherent, and therefore their lower envelopes—$\underline{P}$ and $\underline{P}(\cdot\rfloor B)$, respectively—are jointly coherent as well—see Proposition 1_{43}. Hence, by the results in the previous paragraph, $\underline{P}(\cdot\rfloor B)$ coincides with $\underline{E}(\cdot\rfloor B)$; see Reference [106, Section 6.4.2] as well. Even stronger, as the following result establishes, $\underline{E}(\cdot\rfloor B)$ is not only the lower envelope of $\mathscr{M}_{\underline{P}}\rfloor B$, it even represents it exactly.

Corollary 9. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension. Then for all $B \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\underline{P}(B) > 0$, we have that $\mathscr{M}_{\underline{E}(\cdot\rfloor B)} = \mathscr{M}_{\underline{P}}\rfloor B$.*

[21] This is only true if the domain of $\underline{P}$ is large enough, which is clearly the case here because it is—assumed to be—equal to $\mathscr{G}(\Omega)$.

Proof of Corollary 9$_{\frown}$. We only prove that $\mathscr{M}_{\underline{E}(\cdot\rfloor B)} \subseteq \mathscr{M}_{\underline{P}}\rfloor B$, because the converse inclusion follows trivially from the fact that $\underline{E}(\cdot\rfloor B)$ is the lower envelope of $\mathscr{M}_{\underline{P}}\rfloor B$. So consider any $P^* \in \mathscr{M}_{\underline{E}(\cdot\rfloor B)}$. Then by Proposition 8_{53}, there is some $P(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{E}(\cdot\rfloor\cdot)}$ such that $P^*(\cdot) = P(\cdot\rfloor B)$. If we now let $P(\cdot) := P(\cdot\rfloor\Omega)$, then $P \in \mathscr{M}_{\underline{P}}$, which implies that $P(B) \geq \underline{P}(B) > 0$. Hence, we infer from the coherence of $P(\cdot\rfloor\cdot)$ [$\mathrm{P4}_{50}$] that $P(\cdot\rfloor B)$ is related to $P(\cdot)$ by means of Equation $(2.23)_{60}$. Since $P \in \mathscr{M}_{\underline{P}}$, this implies that $P^* = P(\cdot\rfloor B) \in \mathscr{M}_{\underline{P}}\rfloor B$. □

Hence, in summary: if $\underline{P}(B) > 0$, then by applying Bayes's rule to the linear previsions that dominate $\underline{P}$, we obtain a unique conditional model $\mathscr{M}_{\underline{P}}\rfloor B$ that is fully characterised by its lower envelope. This lower envelope is equal to the natural extension $\underline{E}(\cdot\rfloor B)$, and is the unique solution to Equation $(2.27)_{\frown}$, which is the ultimate reason why this equation is called the generalised Bayes rule.

Unfortunately, these nice results no longer hold if $\underline{P}(B) = 0$. In that case, the GBR might have multiple solutions and coherence alone is not guaranteed to lead to a unique value of $\underline{P}(f\rfloor B)$. The most conservative option is then to resort to the vacuous lower prevision, as defined by $\underline{P}_{\mathrm{v}}(f\rfloor B) := \min f$ for all $f \in \mathscr{G}(B)$; it is the most conservative—most imprecise—coherent lower prevision on $\mathscr{G}(B)$, and it satisfies the GBR whenever $\underline{P}(B) = 0$. Hence, in that case, $\underline{P}_{\mathrm{v}}(\cdot\rfloor B)$ is the most conservative lower prevision on $\mathscr{G}(B)$ that is coherent with $\underline{P}$, and it is therefore equal to the natural extension $\underline{E}(\cdot\rfloor B)$.

By combining this with the results for $\underline{P}(B) > 0$, and also using the fact that $\underline{P}(\mathbb{I}_B[f-\mu])$ is non-increasing in μ because of $(\mathrm{C7})_{49}$, it can be shown that conditioning by means of natural extension results in the use of the following expressions:

$$\underline{E}(f\rfloor B) = \begin{cases} \max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\} & \text{if } \underline{P}(B) > 0 \\ \min f & \text{otherwise} \end{cases} \quad \text{for all } f \in \mathscr{G}(B). \tag{2.28}$$

If we adopt the sensitivity analysis interpretation, this conditioning rule—Equation (2.28)—can be regarded as a special case of the notion of natural extension that was introduced in the previous section: if we apply Equation $(2.25)_{60}$ to the set of dominating linear previsions $\mathscr{M}_{\underline{P}}$, then the resulting conditional model $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{n}}B$ has $\underline{E}(\cdot\rfloor B)$ as its lower envelope, and is even fully characterised by this lower envelope.

Corollary 10. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension. Then $\mathscr{M}_{\underline{E}(\cdot\rfloor B)} = \mathscr{M}_{\underline{P}}\rfloor^{\mathrm{n}}B$ for all $B \in \mathscr{P}_{\emptyset}(\Omega)$.*

Proof of Corollary 10. If $\underline{P}(B) > 0$, this is exactly what is stated in Corollary $9_{\frown}$. If $\underline{P}(B) = 0$, it follows trivially from $\mathrm{P1}_{50}$. □

A similar, but slightly weaker result can be obtained for regular extension as well. If we apply Equation $(2.26)_{60}$ to $\mathscr{M}_{\underline{P}}$, then the lower envelope of the

resulting conditional model $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{r}}B$ is given by

$$\underline{R}(f\rfloor B) := \begin{cases} \max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\} & \text{if } \overline{P}(B) > 0 \\ \min f & \text{otherwise} \end{cases} \quad \text{for all } f \in \mathscr{G}(B). \tag{2.29}$$

and the resulting conditioning rule for lower previsions is also called *regular extension* [106, Appendix J]. It coincides with natural extension whenever $\underline{P}(B) > 0$ or $\overline{P}(B) = 0$, but may differ from it when $0 = \underline{P}(B) < \overline{P}(B)$. In the latter case, in contradistinction with what we found for natural extension, $\underline{R}(\cdot\rfloor B)$ is not guaranteed to fully characterise the conditional model $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{r}}B$ it is the lower envelope of. The set $\mathscr{M}_{\underline{R}(\cdot\rfloor B)}$ of linear previsions that dominate $\underline{R}(\cdot\rfloor B)$ is convex and closed [see Section 2.6.1$_{52}$] but $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{r}}B$ is only guaranteed to be convex [20, 66], and may not be closed [20, Example 1]. However, only very little information is lost; by convexity, and since $\underline{R}(f\rfloor B)$ is the lower envelope of $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{r}}B$, the latter lies in between $\mathscr{M}_{\underline{R}(\cdot\rfloor B)}$ and its relative interior, and therefore approximates it very closely. At the expense of this minimal loss of information, $\mathscr{M}_{\underline{P}}\rfloor^{\mathrm{r}}B$ can be conveniently represented by $\underline{R}(\cdot\rfloor B)$. If $\mathscr{M}_{\underline{P}}$ is finitely generated, this representation is even exact [20, Section 2].

Regular extension can also be introduced without any reference to the sensitivity analysis interpretation or the set $\mathscr{M}_{\underline{P}}$ of dominating linear previsions. Again, as with the natural extension, the resulting lower prevision $\underline{R}(\cdot\rfloor B)$ is coherent with the original model $\underline{P}$ [71, Section 3.3.4]. It only differs from the natural extension if $0 = \underline{P}(B) < \overline{P}(B)$ and is then the largest solution to the GBR, and therefore the *least* conservative—most precise—model that is coherent with $\underline{P}$ [69], whereas natural extension provides the *most* conservative—most imprecise—such model. In order to turn the regular extension into a most conservative model, coherence needs to be combined with additional axioms. Reference [106, Appendix J3, Equation (C16)] provides an abstract condition—if $\overline{P}(B) > 0$ and $\underline{P}(\mathbb{I}_B f) \geq 0$, then $\underline{P}(f\rfloor B) \geq 0$—that does the job, but does not justify why a subject should want to impose this condition as an axiom.

2.7.3 *Computational aspects*

From a computational point of view, calculating $\underline{E}(f\rfloor B)$ or $\underline{R}(f\rfloor B)$ requires two things: evaluating the sign of $\underline{P}(B)$ or $\overline{P}(B)$, respectively, and—in case it is positive—computing the value of $\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$. We consider two distinctly different approaches.

The first approach is to use the extreme points of $\mathscr{M}_{\underline{P}}$. By Equation (2.20)$_{55}$ and conjugacy, we know that $\underline{P}(B)$ is positive if and only if $P(B)$ is positive for all $P \in \operatorname{ext}(\mathscr{M}_{\underline{P}})$, and that $\overline{P}(B)$ is positive whenever there is at least one $P \in \operatorname{ext}(\mathscr{M}_{\underline{P}})$ for which $P(B) > 0$. The following result establishes that $\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$ can be evaluated by applying Bayes's rule to

the element of $\operatorname{ext}(\mathscr{M}_{\underline{P}})$, whenever possible, and then taking the lower envelope of the resulting models.

Proposition 11. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$. Then for any $B \in \mathscr{P}_{\emptyset}(\Omega)$ and any $f \in \mathscr{G}(B)$:*[22]

$$\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\} = \inf\left\{\frac{P(f\mathbb{I}_B)}{P(B)} \colon P \in \operatorname{ext}(\mathscr{M}_{\underline{P}}), P(B) > 0\right\}.$$

Proof of Proposition 11. By Equation $(2.20)_{55}$, we have that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ if and only if $P(\mathbb{I}_B[f-\mu]) \geq 0$ for all $P \in \operatorname{ext}(\mathscr{M}_{\underline{P}})$. Since each of these P is coherent, we also know that $P(\mathbb{I}_B[f-\mu]) = P(\mathbb{I}_B f) - \mu P(B)$, and that $P(\mathbb{I}_B f) = 0$ whenever $P(B) = 0$. Hence, we find that

$$\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0 \Leftrightarrow (\forall P \in \operatorname{ext}(\mathscr{M}_{\underline{P}}) \colon P(B) > 0)\ P(\mathbb{I}_B f) - \mu P(B) \geq 0.$$

This completes the proof, because it implies that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ if and only if μ is lower than or equal to the right-hand side of the equality that we need to prove. □

Theoretically, this approach always works. In practice, it usually only works if the number of extreme points is finite and reasonably small. It also requires that these extreme points are given, or that they can be computed efficiently from $\underline{P}$.

Alternatively, we can work directly with the lower prevision $\underline{P}$, and in particular, with the corresponding real-valued function $\rho_{f,B}$, defined by

$$\rho_{f,B}(\mu) := \underline{P}(\mathbb{I}_B[f-\mu]) \quad \text{for all } \mu \in \mathbb{R}.$$

By Equation $(2.20)_{55}$, we know that $\rho_{f,B}$ is the pointwise minimum of a set of linear, non-increasing functions $P(\mathbb{I}_B f) - \mu P(B)$, with $P \in \operatorname{ext}(\mathscr{M}_{\underline{P}})$. It is therefore (Lipschitz[23]) continuous, concave, and non-increasing, and its first derivative, whenever it exists, lies between $-\overline{P}(B)$ and $-\underline{P}(B)$. The left and right derivatives always exist, and are guaranteed to lie between the same bounds.

Let $\mu_0 < \min f$ and $\mu_1 > \max f$. It then follows from the coherence of $\underline{P}$ that $\rho_{f,B}(\mu_0)$ is positive if and only if $\underline{P}(B)$ is positive too, and similarly for $\rho_{f,B}(\mu_1)$ and $\overline{P}(B)$. Alternatively, the signs of $\underline{P}(B)$ and $\overline{P}(B)$ can be evaluated directly as well. Evaluating $\mu^* := \max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$ is done iteratively. If $\underline{P}(B) > 0$, then $\rho_{f,B}$ is a strictly decreasing function of μ, and μ^* is its only root. By coherence, this root is guaranteed to lie between $\min f$ and $\max f$, and it can therefore be found easily by means of the bisection

[22] This result is essentially well-known; we provide its proof for the sake of completeness.

[23] See Reference [42, Section 6.1].

method, or any other root-finding procedure; see References [42, Section 6.3] and [108, p. 18] for methods that have been specifically designed to exploit the properties of the function $\rho_{f,B}$. If we are able to evaluate $\rho_{f,B}(\mu)$ up to some numerical error ε, and iterate sufficiently often, then by the bounds on the first (left and right) derivative of $\rho_{f,B}$, the error that is made by the bisection method will not exceed $\varepsilon/\underline{P}(B)$. If $\overline{P}(B) > \underline{P}(B) = 0$, then $\rho_{f,B}$ is identically zero in $]-\infty, \mu^*]$ and strictly decreasing in $]\mu^*, +\infty[$, and in this decreasing part, the first derivative, whenever it exists, is bounded above by $-\underline{P}^{>0}(B)$, with $\underline{P}^{>0}(B) := \inf\{P(B) \colon P \in \operatorname{ext}(\mathscr{M}_{\underline{P}}), P(B) > 0\}$, and similarly for the left and right derivative, which always exist. Finding μ^* is now a bit more tricky. Since coherence again implies that μ^* lies in between $\min f$ and $\max f$, we could in principle directly apply the bisection algorithm here as well. However, if during this procedure, numerical errors lead us to mistakenly conclude that $\rho_{f,B}(\mu)$ is negative for some $\mu < \mu^*$, the obtained solution could greatly underestimate the actual value μ^*. The simplest way to fix this is to look for the unique root of $\rho_{f,B} + \delta$, for some sufficiently small $\delta > \varepsilon$. If $\underline{P}^{>0}(B) > 0$—for example, if $\mathscr{M}_{\underline{P}}$ is finitely generated—then the obtained solution will overestimate the actual value of μ^*, but by no more than $(\delta+\varepsilon)/\underline{P}^{>0}(B)$.

2.7.4 *What about sets of desirable gambles?*

After all this elaboration about conditioning with lower previsions, and in particular with natural and regular extension, one could wonder why we even bother to deal with these notions. Why do we not simply work with sets of desirable gambles? The answer is a practical one: from a computational point of view, conditioning by means of natural or regular extension—Equations $(2.28)_{62}$ and $(2.29)_{63}$, respectively—is more tractable. Although Equation $(2.2)_{40}$ provides a conceptually very simple conditioning rule, it is difficult to use in practice. If $\mathscr{D}$ has a complex border structure, it can be very difficult—if not impossible—to check whether a gamble $f \in \mathscr{G}(B)$ belongs to $\mathscr{D}\rfloor B$ or, equivalently, whether $\mathbb{I}_B f \in \mathscr{D}$, especially if $\{\mathbb{I}_B f \colon f \in \mathscr{G}(B)\}$ is a subset of the border of $\mathscr{D}$. The two main reasons are numerical errors and the fact that it is difficult—and sometimes even impossible due to memory limitations—to provide a computer representation for the exact border structure of $\mathscr{D}$.[24]

Nevertheless, even if the actual calculations are performed in terms of lower previsions, sets of desirable gambles remain important, both theoretically and philosophically. One of the key reasons for their importance, is that they provide conditional lower previsions with an interpretation, without any reference to Bayes's rule or the sensitivity analysis interpretation: con-

[24] See References [14, 83] for some ingenious but complex methods that are able to deal with the special case of so-called finitely generated sets of desirable gambles.

ceptually, for every coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, there is a—possibly not given—coherent set of desirable gambles $\mathscr{D}$ such that $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$ coincides with $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.

If $\mathscr{D}$ is given, then in theory, conditioning should be done by applying Equation $(2.2)_{40}$ to $\mathscr{D}$, resulting in a conditional set of desirable gambles $\mathscr{D}\rfloor B$. In practice however, we are usually only interested in the corresponding lower prevision $\underline{P}_{\mathscr{D}\rfloor B}(f) = \underline{P}_{\mathscr{D}}(f\rfloor B)$, for some gamble $f \in \mathscr{G}(B)$. Hence, rather than constructing $\mathscr{D}\rfloor B$, which is often intractable, we will instead try and calculate $\underline{P}_{\mathscr{D}}(f\rfloor B)$ directly. However, even this may be very hard, because—as is the case for $\mathscr{D}\rfloor B$—$\underline{P}_{\mathscr{D}}(f\rfloor B)$ may crucially depend on the exact border structure of $\mathscr{D}$. In order to avoid this dependency on the border structure, the trick is to focus on the unconditional prevision $\underline{P}_{\mathscr{D}}$. Unlike $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$, $\underline{P}_{\mathscr{D}}(\cdot) := \underline{P}_{\mathscr{D}}(\cdot\rfloor\Omega)$ does not depend on the exact border structure of $\mathscr{D}$, and can therefore be evaluated in a more reliable manner. If we now use the techniques in Section $2.7.3_{63}$ to obtain the natural extension $\underline{E}(f\rfloor B)$ of $\underline{P}_{\mathscr{D}}$, then by the results of Section $2.7.2_{61}$, $\underline{E}(f\rfloor B)$ is guaranteed to provide a lower—conservative—bound on $\underline{P}_{\mathscr{D}}(f\rfloor B)$, and when $\underline{P}_{\mathscr{D}}(B) > 0$, this bound will even be exact. For some sets $\mathscr{D}$, it is even possible to prove, on theoretical grounds, that $\underline{P}_{\mathscr{D}}(f\rfloor B)$ is bounded from below by the regular extension $\underline{R}(f\rfloor B)$ of $\underline{P}_{\mathscr{D}}$,[25] thereby providing a lower bound on $\underline{P}_{\mathscr{D}}(f\rfloor B)$ that is guaranteed to be exact whenever $\overline{P}_{\mathscr{D}}(B) > 0$.[26] We will construct such sets in Section 3.4_{79} and Appendix $3.\mathrm{A}_{91}$ and prove that they indeed satisfy this property. In any case, for now, the main message is that even if the underlying model is a set of desirable gambles $\mathscr{D}$, we can still perform all the calculations in terms of lower previsions, using natural or regular extension. This approach is usually more feasible from a computational point of view and is guaranteed to provide lower—conservative—bounds on $\underline{P}_{\mathscr{D}}(f\rfloor B)$, that are often even exact.

In many cases, $\mathscr{D}$ is not given, and is simply an underlying theoretical concept. For example, in the case of conditioning, all we may have to start from is the unconditional lower prevision $\underline{P}(\cdot) := \underline{P}_{\mathscr{D}}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$, with $\mathscr{C} = \{(f,\Omega)\colon f \in \mathscr{G}(\Omega)\}$. In that case, by specifying the set $\mathscr{D}$ ourselves, we are in fact specifying a conditioning rule: for every $B \in \mathscr{P}_{\emptyset}(\Omega)$, the set $\mathscr{D}$ will provide us with a corresponding coherent lower prevision $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ on $\mathscr{G}(B)$. Natural and regular extension correspond to particular choices of $\mathscr{D}$. By definition [see Section $2.3.2_{44}$], natural extension can be obtained by using the set $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, which in this case is equal to $\mathscr{D}_{\underline{P}}$.

[25] This holds trivially if $\underline{P}_{\mathscr{D}}(B) > 0$; if $\overline{P}_{\mathscr{D}}(B) > 0$, this can either be proved directly—as we do in Corollary 21_{86} and Proposition 28_{94}—or, alternatively, by verifying the necessary and sufficient conditions in Reference [74, Appendix A.1] or the sufficient conditions in Reference [14, Theorem 10].

[26] Because $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ is coherent with $\underline{P}_{\mathscr{D}}$, and because, as we have seen in Section $2.7.2_{61}$, $\underline{R}(\cdot\rfloor B)$ is the least conservative lower prevision for which the case.

Proposition 12. *Consider a coherent 'conditional' lower prevision $\underline{P}(\cdot\rfloor\cdot)$ whose domain $\mathscr{C}$ is equal to $\{(f,\Omega)\colon f \in \mathscr{G}(\Omega)\}$, and let $\underline{P}(\cdot) := \underline{P}(\cdot\rfloor\Omega)$ be the corresponding unconditional lower prevision on $\mathscr{G}(\Omega)$. Then $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} = \mathscr{D}_{\underline{P}}$.*

Proof of Proposition 12$_{\frown}$. In this particular case, we find that

$$\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)} = \{f - \underline{P}(f) + \varepsilon\colon f \in \mathscr{G}(\Omega), \varepsilon \in \mathbb{R}_{>0}\} = \{g \in \mathscr{G}(\Omega)\colon \underline{P}(g) > 0\}, \tag{2.30}$$

using C8$_{49}$ to obtain the second equality. Hence, $\mathscr{D}_{\underline{P}}$ is clearly a subset of $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} := \mathscr{E}(\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)})$. In order to establish the converse inclusion, let us consider any $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and prove that $f \in \mathscr{D}_{\underline{P}}$. If $f > 0$, this is trivial. If $f \not> 0$, then by Equation (2.1)$_{39}$, $f \geq \sum_{i=1}^{n} \lambda_i f_i$, with $n \in \mathbb{N}$ and, for all $i \in \{1,\dots,n\}$, $f_i \in \mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$ and $\lambda_i \in \mathbb{R}_{>0}$. Hence, by coherence [C2$_{48}$, C3$_{48}$ and C7$_{49}$] and Equation (2.30), we find that $\underline{P}(f) \geq \underline{P}(\sum_{i=1}^{n} \lambda_i f_i) \geq \sum_{i=1}^{n} \lambda_i \underline{P}(f_i) > 0$ and therefore, that $f \in \mathscr{D}_{\underline{P}}$. □

Regular extension can be obtained by using the set

$$\begin{aligned}\mathscr{D}^{\mathrm{r}}_{\underline{P}} :=& \{f \in \mathscr{G}(\Omega)\colon (\underline{P}(f) \geq 0 \text{ and } \overline{P}(f) > 0) \text{ or } f > 0\}\\ =& \mathscr{D}_{\underline{P}} \cup \{f \in \mathscr{G}(\Omega)\colon \underline{P}(f) = 0 \text{ and } \overline{P}(f) > 0\}.\end{aligned}$$

Proposition 13.[27] *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be the conditional lower prevision on $\mathscr{C}(\Omega)$ that is defined by Equation* (2.29)$_{63}$. *Then $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and $\underline{R}(\cdot\rfloor\cdot)$ are coherent, $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$ and*

$$\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \quad \textit{for all } (f,B) \in \mathscr{C}(\Omega).$$

Proof of Proposition 13. Coherence of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ follows by straightforward verification of D1$_{38}$–D4$_{39}$. D2$_{38}$ holds trivially. D1$_{38}$ and D3$_{39}$ follow directly from the coherence of $\underline{P}$ [C8$_{49}$ and C2$_{48}$, respectively]. In order to prove D4$_{39}$, we consider any $f,g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$, and show that $f+g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. If $f,g \in \mathscr{D}_{\underline{P}}$ this follows from the coherence of $\mathscr{D}_{\underline{P}}$. Otherwise, we may assume without loss of generality that $\underline{P}(f) = 0$ and $\overline{P}(f) > 0$. Since $g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$ clearly implies that $\underline{P}(g) \geq 0$, we infer that $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g) \geq 0$ [C3$_{48}$] and $\overline{P}(f+g) \geq \overline{P}(f) + \underline{P}(g) > 0$ [C5$_{48}$], which in turn implies that $f+g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$.

Next, we prove that $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$. Consider any $f \in \mathscr{G}(\Omega)$ and any $\alpha \in \mathbb{R}$. If $\alpha < \underline{P}(f)$, then $\underline{P}(f-\alpha) > 0$ [C8$_{49}$] and therefore $f - \alpha \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. If $\alpha > \underline{P}(f)$, then $\underline{P}(f-\alpha) < 0$ [C8$_{49}$] and therefore also $f - \alpha \not> 0$ [C1$_{48}$], which implies that $f - \alpha \notin \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Hence, by Equation (2.3)$_{40}$, $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f) = \underline{P}(f)$.

[27] This result was stated without proof in Reference [84, Section 2.6.6]; References [106, Appendix F4] and [14] provide earlier, but less direct statements of the same result.

Since $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ is coherent, $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$ is coherent by construction, and therefore, the only thing that we still need to prove is that $\underline{R}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$. So consider any $(f,B) \in \mathscr{C}(\Omega)$. We will show that $\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Since $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$ is coherent, $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor B)$ is clearly coherent with $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$. First, assume that $\underline{P}(B) > 0$. Then $\underline{P}$ uniquely determines $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$ through the GBR [Equation $(2.27)_{61}$], and therefore $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$ coincides with $\underline{R}(f\rfloor B)$. Next, assume that $\overline{P}(B) > \underline{P}(B) = 0$. Then coherence of $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor B)$ and $\underline{P}$ implies that $\underline{R}(f\rfloor B) \geq \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Consider now any $\mu < \underline{R}(f\rfloor B)$. Since $\underline{P}(\mathbb{I}_B[f-\underline{R}(f\rfloor B)]) = 0$, we know that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ [$\mathrm{C7}_{49}$] and that, with $\lambda := \underline{R}(f\rfloor B) - \mu > 0$:

$$\overline{P}(\mathbb{I}_B[f-\mu]) \geq \underline{P}(\mathbb{I}_B[f-\underline{R}(f\rfloor B)]) + \overline{P}(\mathbb{I}_B\lambda) \geq \lambda\overline{P}(B) > 0$$

[$\mathrm{C5}_{48}$ and $\mathrm{C2}_{48}$]. This implies that $\mathbb{I}_B[f-\mu] \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Since this holds for all $\mu < \underline{R}(f\rfloor B)$, we infer from Equation $(2.10)_{42}$ that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \geq \underline{R}(f\rfloor B)$ and therefore that $\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Finally, assume that $\overline{P}(B) = 0$. Consider any $\mu > \min f$ and choose $\alpha \in \mathbb{R}$ such that $\alpha > \max f$ and $\alpha > 0$. Then $\overline{P}(\mathbb{I}_B[f-\mu]) \leq \overline{P}(\mathbb{I}_B\alpha) = \alpha\overline{P}(B) = 0$ [$\mathrm{C7}_{49}$ and $\mathrm{C2}_{48}$] and $\mathbb{I}_B[f-\mu] \not> 0$, and therefore $\mathbb{I}_B[f-\mu] \notin \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Since this holds for all $\mu > \min f$, we infer from Equation $(2.10)_{42}$ that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \leq \min f$ and therefore, by coherence [$\mathrm{C1}_{48}$], that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) = \min f = \underline{R}(f\rfloor B)$. □

It is important to realise that $\mathscr{D}_{\underline{P}}$ and $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ are not necessarily the only sets of desirable gambles that result in the use of natural or regular extension as a conditioning rule, respectively. Since sets of desirable gambles are more expressive than conditional lower previsions, other sets of desirable gambles may result in the same conditioning rules. However, for natural extension, the set $\mathscr{D}_{\underline{P}}$ is fundamental, because it is the smallest—most conservative, most imprecise—set of gambles whose desirability is implied by $\underline{P}$ and coherence. For regular extension, $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ does not have such a special status; there are smaller sets of desirable gambles that also lead to regular extension. We will construct an example in Appendix $3.\mathrm{A}_{91}$ and will show that it is a—sometimes strict—subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$; see Corollary 32_{96} and Example 1_{96}.

3

UPDATING

"... la théorie des probabilités n'est, au fond, que le bon sens réduit au calcul; elle fait apprécier avec exactitude ce que les esprits justes sentent par une sorte d'instinct, sans qu'ils puissent souvent s'en rendre compte."

Pierre-Simon Laplace

Conditioning is commonly—and successfully—used for a multitude of practical purposes, the most important of which is to solve the following *updating* problem: starting from an initial belief model for an uncertain variable X that takes values in Ω, and given the additional information that some event $O \in \mathscr{P}_{\emptyset}(\Omega)$ has occurred, how then should we update this belief model to reflect this new information? If the belief model is taken to be a probability mass function, then the traditional solution to this problem is to condition the original model by means of Bayes's rule. In fact, in that context, the act of "conditioning on an event O" is often even identified with solving this updating problem. Similarly, if the belief model is an imprecise probability model, the conditioning rules that we have discussed before are commonly used as updating rules.

However, it should not be forgotten that conditioning is just a mathematical concept. Before using this concept to solve a practical problem such as the updating task, one should at least try and justify why it is indeed reasonable to use it for this purpose. The goal of this chapter is to perform this exercise for the two imprecise-probabilistic conditioning rules that we discussed earlier

on: natural and regular extension. We intend to show that, in many cases, it does indeed make sense to use these rules to update a belief model in the light of the occurrence of an event.

Our starting point is an asymmetric version of Walley's updating principle, which we employ to argue for using conditioning as a conservative updating rule for sets of desirable gambles. This will then lead to a rather straightforward justification for updating by means of natural extension. It will also follow from our analysis that there is room for other updating strategies, including less conservative ones, such as updating by means of regular extension. We end this chapter with a justification for using regular extension as an updating rule. We provide two versions of this justification: a simple one, which requires an assumption of ideal precision, and a more involved one, which does not. Appendix $3.A_{91}$ gathers a number of additional technical results that are relevant to the topic, but which do not fit into the main storyline of this chapter. Pointers to these results have been inserted at appropriate places in the main text.

3.1 NARROWING DOWN THE PROBLEM

(Imprecise-probabilistic) belief updating, in a general sense, is highly complicated. Many different, and often ill-defined aspects might come into play: additional background information other than the occurrence of O, complex dynamical aspects, changes of mind, biased information, and so on; see for example References [91, 97, 118]. For our present purposes, we will restrict ourselves to a specific, less general setting, which is nevertheless still applicable in a wide variety of situations.

3.1.1 *Updating by means of a rule*

Our most important restriction is that we only consider situations where the updated models are provided beforehand. In other words: we are looking for an *updating rule*, which—by definition of a rule—is stated in advance. For any event $O \in \mathscr{P}_{\emptyset}(\Omega)$ that is considered, this rule provides an associated updated belief model, that is obtained simply by applying the rule to the original belief model. If O occurs, this updated model is then simply adopted. This situation for example reflects the typical practical two-phased approach to modelling, where the model is first built by experts, based on expert knowledge and/or data, only to be queried afterwards by the user. The user is then usually not an expert in the field, and is therefore provided not only with a model, but also with an updating rule to go with it.

An important consequence of restricting ourselves to this particular case, is that it rules out the possibility of taking into account any additional information other than the occurrence of some event O, because such additional information cannot be anticipated when designing—or choosing—the rule. In

other words, all that is learned is that O has occurred, and nothing else. However, our setting does not rule out the possibility of changing your mind, it only requires you to do so in a way that is specified in advance, and that is based on no information other than that O has occurred.

3.1.2 *What does it mean to learn that an event has occurred?*

Since our updating rule is based only on the initial belief model and the information that O has occurred, we should clearly define what 'to be informed that O has occurred' means. The 'O has occurred' part is easy: it simply means that X has taken a value in O. The 'to be informed that' part is a more tricky. In our set-up, we will only deal with cases where the information that O has occurred is reported to us honestly, correctly and whenever it applies, and where we do not receive any additional information about which particular value X takes in O. We discuss each of these separate requirements below, but basically, they reduce to the following single requirement: *it should be agreed upon beforehand that we will be informed about the occurrence of O—and nothing more—if and only if it actually occurs*.

The requirement that we should not receive any additional information about which particular value X takes in O may seem straightforward, but it is in fact not that trivial. For example, among other things, it implies that the point in time when we are informed that O has occurred should not provide us with any additional information about the value of X, other than that it belongs to O. For example, if X is the outcome of two consecutive coin flips, and O is the event that the outcome of at least one of them is heads, it could be that after the first flip, this event is already known to be true. However, we should not be told so until after the second flip, because otherwise, we would not only learn that the outcome of at least one of the coin flips was heads—that O has occurred—but we would also receive the additional information that the outcome of the first flip was heads.

The requirement of honesty is typically relevant if the information is the answer to some sensitive question, such as whether or not someone is a smoker. The answer may be biased; patients who smoke will often say that they do not. These kind of biases occur frequently, and should be taken into account. However, we do not consider them to be part of the updating problem, but rather of the modelling problem. Whenever such a bias is suspected, the model should be extended in such a way as to differentiate between the answer that is given and the actual truth, and this extended model should try and capture the relationship between both. The event O is then simply the—possibly dishonest—answer that is given by the patient, which is *not* required to coincide with the truthful answer to the question. In this way the event O is trivially guaranteed to be reported honestly, and hence our set-up applies to the extended model.

The requirement of correctness is similar to that of honesty, but refers to unintended errors rather than intentional ones, for example due to measure-

ment errors, miscommunication, and so on. Again, we do not consider this to be part of the updating problem, but rather of the modelling problem. The event O should be taken to refer to a—possibly wrong—actual observation; the observation process itself, as well as the errors it produces, are taken to be part of the model. This guarantees that the observation that O has occurred is trivially correct, thereby making our set-up apply.

Finally, the information that the event O has occurred should be reported whenever it applies. More exactly: it should be agreed upon beforehand that if O occurs, we are guaranteed to receive the information that it indeed does. At first sight, this might seem to follow from the requirements of honesty and correctness, but it does not: honesty and correctness only guarantee that whenever we are informed that O has occurred, it has indeed occurred, but not the other way around. Consider for example a situation where we come to know that the outcome of the roll of a die is even. Then, under the assumptions of correctness and honesty alone, this only allows us to infer that the event O on which we should update is a subset of $O^* := \{2,4,6\}$. For example, it might be that it was decided beforehand—without our knowing—that the information that the outcome is even would only be provided to us if the actual outcome is 2 or 4, and kept from us whenever the outcome is 6. In that case, the information that the outcome is even is correct, as well as honest, but it is too weak, since it also implies that the outcome is not 6. Hence, the event on which we should be updating is $O := \{2,4\}$ rather than O^*.

3.1.3 *Which events should be considered?*

So far, we have restricted attention to a generic single event $O \in \mathscr{P}_{\emptyset}(\Omega)$. However, that is not the end of the story. Can it be any event? Can we consider multiple events? No and yes, respectively.

Since updating is by its very definition concerned with providing a new belief model after getting to know that O has occurred, it clearly only applies to observable events. We do not necessarily have to be able to observe it ourselves, but someone has to be, and needs to be able to communicate it to us. In contrast, conditioning can be applied to any event, regardless of whether it is observable. It is for this reason that we use O—from 'observation'—to refer to a generic event on which we update, rather than B, which we use for events on which we condition.

Many authors do not single out just one event O, but consider it to be part of a structured collection of events that reflects the actual process of gathering information. For example, Walley [106] focuses on partitions of Ω, the elements of which typically correspond to the outcome of some experiment, or the answer to some question. Shafer [89] considers more complex structures, which he calls protocols; they allow for multi-phased set-ups as well, where additional experiments and/or questions can be used to further refine the sample space. These situations are covered by our setting as well; it suffices for

the requirements of Section 3.1.2$_{71}$ to hold for every event in the considered collection. However, our approach is more general. For example, if O is part of some partition, then we do not require that there is some predetermined point in time where we are guaranteed to know which event in this partition has occurred. In any case, it is not necessary to consider a collection of events. Furthermore, if O does belong to some natural collection, then most of our analysis will not depend on it. Therefore, we will usually refrain from mentioning any collection, and will simply consider a single event $O \in \mathscr{P}_{\emptyset}(\Omega)$. Whenever the specific collection to which O belongs does become relevant, it will be mentioned explicitly; see for example Section 3.3.2$_{77}$.

3.2 CONDITIONING AS AN UPDATING STRATEGY

Now that we know exactly what we are dealing with, let us start solving the problem. First of all, by our assumptions, we are guaranteed that whenever we are informed that O has occurred, it has indeed occurred, meaning that X takes a value in O. Hence, after learning that O has occurred, we no longer need to consider the elements in $\Omega \setminus O$. This implies that the updated belief model should capture the uncertainty about which value X takes in the remaining set of possibilities O. Depending on the framework we adopt, this can be a set of desirable gambles on O, a lower prevision on $\mathscr{G}(O)$, or any of the other uncertainty models that we discussed in Chapter 2$_{37}$.

3.2.1 *An asymmetric version of Walley's updating principle*

We first consider the framework of sets of desirable gambles. The initial belief model is then a coherent set of desirable gambles on Ω, which we denote by $\mathscr{D}$, and the updated belief model is a coherent set of desirable gambles on O, which we denote by $\mathscr{D}_O$. By definition of an updated model, $\mathscr{D}_O$ consists of gambles that are desirable *after* O occurs, whereas $\mathscr{D}$ consists of gambles that are desirable *now—before* O occurs. The central question of the updating problem is whether or not, and if yes, how, $\mathscr{D}_O$ should be related to $\mathscr{D}$.

In order to be able to answer this question, we distinguish $\mathscr{D}_O$ from two other sets of gambles on O. The set $\mathscr{D}\rfloor O$ has already been discussed at length; it is fully determined by $\mathscr{D}$, as it consists of those gambles $f \in \mathscr{G}(O)$ for which the gamble $\mathbb{I}_O f$ is desirable *now*—is an element of $\mathscr{D}$. The set $\mathscr{D}^O$ is new. It consists of those gambles on O of which our subject *now* thinks that they *should* become desirable *after* O occurs; following Walley's terminology [106, Section 6.5.1], we call these gambles O-desirable. So how is $\mathscr{D}_O$ related to $\mathscr{D}^O$ and $\mathscr{D}\rfloor O$?

The connection between $\mathscr{D}_O$ and $\mathscr{D}^O$ follows trivially from our assumptions. As explained in Section 3.1.1$_{70}$, we are restricting attention to situations in which the updated model is provided in advance by an expert and then simply adopted by the user once O actually occurs. Within this setting, $\mathscr{D}_O$ and $\mathscr{D}^O$

coincide because the user—by assumption—will consider a gamble $f \in \mathscr{G}(O)$ to be desirable after O occurs if and only if it was stated beforehand by the expert that f should be desirable after O occurs, or equivalently, if the expert considered it to be O-desirable.

The remaining task is to establish a relationship between $\mathscr{D}^O$ and $\mathscr{D}\rfloor O$. The best-known solution is that of Walley [106, Section 6.1.6], who claims that $\mathscr{D}^O$ should be equal to $\mathscr{D}\rfloor O$. We do not agree. We will argue that this requirement is too strong, and that the only thing that can be reasonably imposed is for $\mathscr{D}^O$ to be a superset of $\mathscr{D}\rfloor O$. We start by repeating Walley's argument. The central idea is very elegant: under the assumptions of Section $3.1.2_{71}$,[1] the effect of owning a gamble $f \in \mathscr{G}(O)$ after being informed that O has occurred is indistinguishable from the effect of owning the gamble $\mathbb{I}_O f \in \mathscr{G}(\Omega)$ now; as explained in Section $2.2.2_{40}$, they result in the same payoff if O occurs and have no effect otherwise. So far, we agree. However, Walley does not stop here. He uses this fact to infer—without any actual argumentation—that f should be O-desirable if and only if $\mathbb{I}_O f$ is desirable: $f \in \mathscr{D}^O \Leftrightarrow \mathbb{I}_O f \in \mathscr{D} \Leftrightarrow f \in \mathscr{D}\rfloor O$; he calls this the updating principle.

We think that in order for the updating principle to be compelling, considering a gamble to be desirable should mean being willing to accept it, as Walley seems to assume. If this is not the case, then the fact that two transactions have the same effect does not imply that their desirability should be equivalent. In our framework, where desirability of f means *strict* preference of f over the status quo—a notion that is stronger than just being willing to accept f—such an equivalence would require the status quo to remain identical as well, and this is not the case, because the frame of reference changes: after O has occurred, the status quo is no longer the zero gamble on Ω, but rather the zero gamble on O.

It seems to us that there are two situations where this change of status quo may influence the desirability of a gamble: if—*before* the occurrence of O—the subject believes that O cannot occur or if he has indeterminate beliefs about whether or not O can occur. In those two cases, even if the subject does not prefer $\mathbb{I}_O f$ strictly over the status quo *before* the occurrence of O, it may still make perfect sense for him to think that f should be strictly preferred over the status quo *after* the occurrence of O, because it is then clear that O can occur. It is only if the subject believes that O can occur——again, and obviously, *before* the occurrence of O—that it seems compelling that O-desirability of f should imply the desirability of $\mathbb{I}_O f$. In that case, it can be argued that $\mathscr{D}^O$ should be a subset of $\mathscr{D}\rfloor O$. However, in general, O-desirability of a gamble f does not need to imply the desirability of $\mathbb{I}_O f$.

The converse relation does hold in general: $\mathscr{D}\rfloor O$ should be a subset of

[1] Actually, Walley seems to impose even stronger assumptions, as he requires O to be part of a partition of Ω. We do not consider this to be necessary.

$\mathscr{D}^O$. Indeed, if $\mathbb{I}_O f$ is desirable now, then it must be that the restriction of $\mathbb{I}_O f$ to O—the gamble f—is strictly preferred over the restriction of the status quo to O—the zero gamble on O—because outside of O, $\mathbb{I}_O f$ is identical to the status quo and therefore clearly not strictly preferred to it. Since, under the conditions imposed in Section 3.1.2$_{71}$, these restrictions are exactly the payoffs that become relevant after O has occurred, we conclude that f is O-desirable.

By combining the arguments above, we find that under the conditions described in Section 3.1$_{70}$:

$$\mathscr{D}_O = \mathscr{D}^O \supseteq \mathscr{D}\rfloor O. \tag{3.1}$$

Since we also want $\mathscr{D}_O$ to be coherent, we find that the updated set of desirable gambles $\mathscr{D}_O$ should be a coherent superset of the conditional set of desirable gambles $\mathscr{D}\rfloor O$. Furthermore, as we have seen, if the subject believes that O can occur, $\mathscr{D}^O$ and therefore also $\mathscr{D}_O$ should be equal to $\mathscr{D}\rfloor O$. Of course, this requires a clear definition of what it means for a subject to believe that O can occur. We discuss this further in Section 3.3.2$_{77}$; until then, as we explain in the next section, $\mathscr{D}_O$ will automatically be equal to $\mathscr{D}\rfloor O$.

3.2.2 *Conditioning as a conservative updating strategy*

Without any further assumptions, the only reasonable updating strategy that follows from Equation (3.1) is to use the conditional model $\mathscr{D}\rfloor O$ as our updated model $\mathscr{D}_O$, simply because it is the most conservative—most imprecise—choice of $\mathscr{D}_O$ that is compatible with Equation (3.1), and therefore the only one that is truly implied by it. Furthermore, provided that $\mathscr{D}$ is coherent, $\mathscr{D}\rfloor O$ will be coherent as well. Hence, after all this effort to distinguish between conditioning and updating, it turns out—rather amusingly—that indeed, as is commonly done, conditioning can be regarded as an updating strategy—when the conditions of Section 3.1$_{70}$ are satisfied.

3.2.3 *Justifying natural extension as an updating rule*

The result above easily translates to the framework of lower previsions. If the original belief model is a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, it suffices to apply the result of Section 3.2.2 to the smallest—most conservative—associated set of desirable gambles $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ to find that the most conservative updated lower prevision on $\mathscr{G}(O)$ that is compatible with—the only one that is implied by—Equation (3.1) is given by the natural extension $\underline{E}(\cdot\rfloor O)$ of $\underline{P}(\cdot\rfloor\cdot)$.

If $\underline{P}(\cdot\rfloor\cdot)$ is effectively an unconditional lower prevision $\underline{P}(\cdot) = \underline{P}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$, this natural extension can be calculated using the computational techniques of Section 2.7.3$_{63}$. Otherwise, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is a (possibly strict) superset of $\mathscr{D}_{\underline{E}(\cdot\rfloor\Omega)}$ and therefore, the natural extension $\underline{E}(\cdot\rfloor O)$ of $\underline{P}(\cdot\rfloor\cdot)$ (possibly strictly) dominates the natural extension $\underline{E}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$. If the domain of $\underline{P}(\cdot\rfloor\cdot)$ is large enough, such that $\underline{P}(\cdot\rfloor\Omega)$ is defined on $\mathscr{G}(\Omega)$, $\underline{E}(\cdot\rfloor\Omega)$ can be replaced by $\underline{P}(\cdot\rfloor\Omega)$ because they are then equal. If $\underline{P}(\cdot\rfloor\Omega)$ is known—or if

$\underline{E}(\cdot\rfloor\Omega)$ can be calculated—$\underline{E}^*(\cdot\rfloor O)$ can be obtained by applying the computational techniques of Section 2.7.3$_{63}$. The result is guaranteed to coincide with $\underline{E}(\cdot\rfloor O)$ whenever $\underline{P}(O\rfloor\Omega) > 0$—$\underline{E}(O\rfloor\Omega) > 0$—but will be vacuous and might therefore only be a (safe) lower approximation of $\underline{E}(\cdot\rfloor O)$ if $\underline{P}(O\rfloor\Omega) = 0$—$\underline{E}(O\rfloor\Omega) = 0$.

3.2.4 *Justifying Bayes's rule as an updating rule*

For the special case of linear previsions—and hence also probability mass functions—natural extension coincides with Bayes's rule whenever O has positive probability, and therefore, in that case, we obtain a justification for using Bayes's rule as an updating strategy. Since it is impossible for a coherent lower prevision to dominate a linear one without coinciding with it, we can even conclude that, whenever $P(O) > 0$, Bayes's rule is the *unique* updating strategy that is compatible with Equation (3.1)$_{\frown}$. However, this special case is not our main point of interest here, as there are many other justifications available for updating by means of Bayes's rule; see for example References [88, 89, 91].

3.3 WHAT ABOUT OTHER UPDATING STRATEGIES?

As we have just shown, the use of conditioning as an updating rule is justified by Equation (3.1)$_{\frown}$, and this expression applies whenever the conditions of Section 3.1$_{70}$ are met. However, this is not the end of the story. Other updating strategies can be justified as well, both more and less conservative.

3.3.1 *More conservative strategies*

More conservative updating rules—smaller updated sets of desirable gambles—do not require any actual justification. Since we do not adopt an exhaustive interpretation [see Section 2.2.1$_{39}$] they are fully consistent with the commitments that are implied by the use of conditioning as an updating rule. However, it is rather silly to use these more conservative updating strategies because, as we have just shown, they are unnecessarily weak. The reason why it may nevertheless be reasonable to apply them is a practical one: computing conditional models may be intractable, and in those cases, tractable more conservative updating strategies can serve as a useful safe approximation. For example, as explained in Section 2.7.4$_{65}$, the natural extension $\underline{E}(\cdot\rfloor O)$ of the unconditional prevision $\underline{P}_{\mathscr{D}}$ serves as a tractable outer approximation of $\underline{P}_{\mathscr{D}}(\cdot\rfloor O)$, and $\underline{P}_{\mathscr{D}}(\cdot\rfloor O)$ itself is already more tractable to compute, as well as more conservative, than the actual conditional set of desirable gambles $\mathscr{D}\rfloor O$. In much the same way, as explained in Section 3.2.3$_{\frown}$, the natural extension $\underline{E}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$—or $\underline{P}(\cdot\rfloor\Omega)$—can be used as a tractable lower bound for the natural extension $\underline{E}(\cdot\rfloor O)$ that we are actually after, which is the one that corresponds to $\underline{P}(\cdot\rfloor\cdot)$.

3.3.2 *Less conservative strategies*

Far more attractive are updating strategies that are less conservative, because they result in belief models that are more informative and therefore, ultimately, more powerful. One of the nice aspects of our argumentation in Section 3.2.1$_{73}$—and Equation (3.1)$_{75}$ in particular—is that it is compatible with such strategies: $\mathscr{D}_O$ may be strictly larger than $\mathscr{D}\rfloor O$. However, the fact that these updating strategies are not ruled out by Equation (3.1)$_{75}$ does not suffice to justify them. In order to truly justify adding extra gambles to $\mathscr{D}_O$, we need to come up with a compelling principle that implies their addition. We will introduce such a principle in Section 3.4$_{79}$, and show that it implies the use of regular extension as an updating rule.

It is also important to realise that there are limits to how much larger $\mathscr{D}_O$ can be made. We have already mentioned two of these upper constraints in Section 3.2.1$_{73}$. First of all: $\mathscr{D}_O$ should be kept small enough to keep it coherent. Secondly: if the subject believes that O can occur, then $\mathscr{D}_O$ should be equal to $\mathscr{D}\rfloor O$. In order to make this constraint exact, we need to define what "believing that O can occur" means. In our finitary context, and within the framework of sets of desirable gambles, we consider it reasonable to use the following definition: a subject believes that O can occur if and only if he is willing to bet on its occurrence at some strictly positive (but possibly very small) betting rate—if $\underline{P}_{\mathscr{D}}(O) > 0$; see Section 2.3.5$_{49}$.

A third upper constraint is that, $\mathscr{D}_O$ and $\mathscr{D}$ should avoid partial loss, in the sense that it should not be possible to combine a gamble $f \in \mathscr{D}$ with a gamble $f_O \in \mathscr{D}_O$ such that the combined transaction $f + \mathbb{I}_O f_O$ results in a payoff that is never positive and sometimes negative.[2] This is important because $\mathscr{D}$ and $\mathscr{D}_O$ are both announced before the occurrence of O, and therefore, a subject who has $\mathscr{D}$ and $\mathscr{D}_O$ as its belief models can be forced to accept such a combination of transactions. However, fortunately, as long as $\mathscr{D}_O$ is a coherent superset of $\mathscr{D}\rfloor O$, this will never happen.

Proposition 14. *Let $\mathscr{D}$ be a coherent set of desirable gambles on Ω, let O be an event in $\mathscr{P}_\emptyset(\Omega)$, and let $\mathscr{D}_O$ be a coherent set of desirable gambles on O such that $\mathscr{D}_O \supseteq \mathscr{D}\rfloor O$. Then*

$$(f \in \mathscr{D} \text{ and } f_O \in \mathscr{D}_O) \Rightarrow f + \mathbb{I}_O f_O \not\leq 0.$$

Proof of Proposition 14. Consider any $f \in \mathscr{D}$ and $f_O \in \mathscr{D}_O$ and let f'_O be the restriction of f to O. If there is some $\omega \in \Omega \setminus O$ such that $f(\omega) > 0$, then $(f + \mathbb{I}_O f_O)(\omega) = f(\omega) > 0$ and therefore $f + \mathbb{I}_O f_O \not\leq 0$. Hence, without loss of generality, we may assume that $\mathbb{I}_O f \geq f$. By coherence of $\mathscr{D}$ [D5$_{39}$], this

[2] It suffices to consider just a single gamble from each of the sets involved because, by coherence of the individual sets, any finite combination of gambles of the same set is again a member of that set [D4$_{39}$].

implies that $\mathbb{I}_O f'_O = \mathbb{I}_O f \in \mathscr{D}$ and therefore also that $f'_O \in \mathscr{D}\rfloor O \subseteq \mathscr{D}_O$. Invoking the coherence of $\mathscr{D}_O$, we find that $f'_O + f_O \in \mathscr{D}_O$ [D4$_{39}$] and therefore also that $f'_O + f_O \not\leq 0$ [D6$_{39}$]. Hence, we know that there is some $\omega \in O$ such that $f'_O(\omega) + f_O(\omega) > 0$. Since $f'_O(\omega) = f(\omega)$ and $f_O(\omega) = \mathbb{I}_O(\omega) f_O(\omega)$, this implies that $f(\omega) + \mathbb{I}_O(\omega) f_O(\omega) > 0$ and therefore also that $f + \mathbb{I}_O f_O \not\leq 0$. $\square$

The situation becomes more tricky if a subject announces, besides $\mathscr{D}$, updated models $\mathscr{D}_O$ for multiple events O, for example for every element of some partition $\mathscr{O}$ of Ω, or for the set of events $\mathscr{O}$ that corresponds to a protocol [see Section 3.1.3$_{72}$]. In those cases, even if $\mathscr{D}_O \supseteq \mathscr{D}\rfloor O$ for all $O \in \mathscr{O}$, it is often very easy to combine gambles in $\mathscr{D}$ with gambles from these different updated models $\mathscr{D}_O$, $O \in \mathscr{O}$, in such a way that the combined transaction makes the subject who announced these models suffer a partial loss, or sometimes even a sure loss—a strictly negative payoff regardless of the outcome. Dempster's rule of conditioning is for example known to suffer from this problem [106, Sections 5.13.9–11]. We will not discuss the exact conditions under which a subject can be made to suffer from such a partial or sure loss any further; see Reference [118] for detailed technical discussions of these and other related consistency criteria between initial and updated belief models. For our present purposes, it suffices to realise that updating by means of conditioning will always avoid partial loss, simply because $\mathscr{D}$ is coherent and therefore satisfies D6$_{39}$. As an immediate consequence, we find that whenever there is a coherent set of desirable gambles $\mathscr{D}^*$ such that $\mathscr{D} \subseteq \mathscr{D}^*$ and such that, for all $O \in \mathscr{O}$, $\mathscr{D}_O \subseteq \mathscr{D}^*\rfloor O$, then updating by means of these models $\mathscr{D}_O$, $O \in \mathscr{O}$, is guaranteed to avoid partial loss. Regular extension provides a nice example: by Proposition 13$_{67}$, it can be seen to correspond to the use of updated models that are obtained by conditioning the set of desirable gambles $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ rather than the actual model $\mathscr{D}_{\underline{P}}$; since $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ is a coherent superset of $\mathscr{D}_{\underline{P}}$, we find that updating $\mathscr{D}_{\underline{P}}$ in this way is guaranteed to avoid partial loss, even if multiple updated models are announced at the same time.

3.3.3 *Different settings and interpretations*

It should not be forgotten that our justification for updating by means of natural extension only applies if the conditions that were discussed in Section 3.1$_{70}$ are met, and that it furthermore crucially depends on our subjective interpretation in terms of desirable gambles. If these conditions are not met, or if this interpretation is not adopted, our argumentation is no longer compelling, and other updating strategies could be considered, provided of course that one can find a way to justify them. Reference [106, Section 6.11] and References [91,97,118] provide some ideas on how to deal with situations where the conditions in Section 3.1$_{70}$ are relaxed. Reference [53, Section 6.3.2] compares our approach with updating rules that are not based on interpretations in terms of gambles. Interestingly, many of the alternative updating rules that are provided in the lit-

erature, regardless of the setting they consider or the interpretation they adopt, tend to be at least as informative as natural extension. Maximum likelihood updating (a special case of Dempster's rule of conditioning) [59] and α-cut updating [12] are two examples; see Section 3.4.1 as well. Since we do not adopt an exhaustive interpretation, this implies that, rather surprisingly, and despite the fact that they come from a completely different direction, these other updating rules nevertheless turn out to be compatible with our approach. This being said, we will now refocus on our setting, and our interpretation, and we will use it to explain that it is possible to justify the use of regular extension as an updating rule by combining the results of this section with additional arguments.

3.4 JUSTIFYING UPDATING WITH REGULAR EXTENSION

Regular extension comes across as an intuitive updating rule because of its clear interpretation in terms of sets of probability mass functions or sets of linear previsions. Because of the popularity of Bayes's rule as an updating tool, it seems natural to simply apply it whenever possible, and to ignore the models to which it cannot be applied—those that assign probability zero to the event of interest. The goal of this section is to justify this approach, in two different ways. Our first justification is based on an assumption of ideal precision. It is expressed in terms of sets of linear previsions and closely resembles the intuitive idea sketched above. Our second justification for the use of regular extension as an updating rule starts from less restrictive assumptions; it does not require an assumption of ideal precision and is expressed directly in terms of sets of desirable gambles and/or lower previsions.

3.4.1 *Using an assumption of ideal precision*

As soon as ideal precision is adopted, justifying the use of regular extension as an updating rule for sets of linear previsions is fairly straightforward. This justification is often taken for granted, but for the sake of completeness, let us make the argument explicit.

Let $\mathscr{M}$ be a set of linear previsions. Due to the assumption of ideal precision, each of these linear previsions is considered to be a candidate for the 'correct' linear prevision, but we do not know which one of them it is. However, so the argument goes, after an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ occurs, some of these candidates can be ruled out, in particular those that assign probability zero to O. Indeed, due to our interpretation for linear previsions, adopting a prevision that assigns $P(O) = 0$ implies that you are willing to bet against O at betting rates that are arbitrarily close to 1—at all odds. With hindsight, after observing O, and given that Ω is finite, this seems like an unreasonable commitment, which is the reason why these previsions are no longer judged to be a candidate for

the 'correct' model. Hence, with hindsight, after observing O, the initial set of candidates should have been $\mathscr{M}' := \{P \in \mathscr{M} : P(O) > 0\}$ rather than $\mathscr{M}$. If $\mathscr{M}' \neq \emptyset$, then since every prevision in $\mathscr{M}'$ is a separate candidate model, each of these is to be updated individually. If the conditions of Section 3.1_{70} are met, then as explained in Section $3.2.2_{75}$, the fact that O has positive probability implies that this should be done by means of Bayes's rule, leading us to use $\mathscr{M}' \rfloor O = \mathscr{M} \rfloor^{\mathrm{r}} O$ as our updated set of candidate models. If $\mathscr{M}' = \emptyset$, then none of the candidate models are reasonable, and this procedure cannot be applied. In that case, we are led to consider the set $\mathbb{P}_O = \mathscr{M} \rfloor^{\mathrm{r}} O$ of all possible previsions on O as our updated set of candidates, simply because by the assumption of ideal precision, one of these previsions is guaranteed to be the correct model. Either way, the approach above leads us to adopt $\mathscr{M} \rfloor^{\mathrm{r}} O$ as our updated set of candidate models, and thereby seems to provide a justification for the use of regular extension as an updating rule.

However, there is still a slight issue with this justification, at least in the way we have presented it so far. Indeed, one could argue that after O has occurred, the initial candidate models are no longer relevant, and that therefore, it makes no sense to go back and remove some of them. This problem is solved by the fact that we are adopting the setting of Section $3.1.1_{70}$. Since this setting requires that the updated model is provided in advance, the act of constructing this updated model is necessarily a thought experiment that is conducted in advance, before the occurrence of any event. Within this thought experiment, it is implicitly assumed that O can occur, because otherwise, it makes no sense to provide an updated model for when O actually occurs, and it is this implicit assumption that leads us to adopt $\mathscr{M}'$ instead of $\mathscr{M}$. It is important to realise that $\mathscr{M}'$ exists within this thought experiment only. Our initial belief model, before any event has occurred, is still the set $\mathscr{M}$.

Similar arguments can also be used to try and justify other updating rules. For example, taking $\mathscr{M}'$ to be the set of all previsions that assign maximal probability to O results in maximum likelihood updating, of which Dempster's rule of conditioning can be regarded as a special case [59]. α-cut updating corresponds to the removal of all previsions P in $\mathscr{M}$ for which $P(O) < \alpha \overline{P}_{\mathscr{M}}(O)$, for some $\alpha \in (0,1)$ [12]. However, we consider the case $P(O) = 0$ to be the more fundamental. The removal of extra previsions seems hard to justify on theoretical grounds. For example, the value of α is bound to be arbitrary. Furthermore, the resulting rules are no longer guaranteed to avoid partial or sure loss and will often fail to satisfy the second upper constraint that was discussed in Section $3.3.2_{77}$. Nevertheless, these rules have proved useful in practice, and the argumentation above can be used to motivate their use on theoretical grounds.

These arguments can also be used to justify regular extension as an updating rule for lower previsions. It suffices to adopt the sensitivity analysis interpretation, and to apply the reasoning above to the set of dominating linear previsions. However, this is unnecessarily restrictive, because, as we are about

to show, a similar justification can also be obtained more directly, without any reference to the sensitivity analysis interpretation or ideal precision.

3.4.2 *Dropping the assumption of ideal precision*

One of the crucial points in the previous section was that, since the updated model is (assumed to be) specified in advance, the act of constructing it is necessarily a thought experiment that makes the implicit assumption that the event O can occur. Under this assumption, some of the previsions in $\mathscr{M}$ can be removed, and in this way, we obtain regular extension.

The very same idea can be applied to sets of desirable gambles. Our current set of desirable gambles $\mathscr{D}$ need not be the same as the set of desirable gambles $\mathscr{D}'$ that we would adopt under the extra assumption that O can occur—not to be confused with the set of gambles $\mathscr{D}\rfloor O$ that are desirable contingent on the actual occurrence of O. But what should $\mathscr{D}'$ be? Is it related to $\mathscr{D}$? Can we construct it in an automated way?

3.4.3 *Adding an assessment*

We think that an assumption that O can occur should lead us to add the following assessment:[3]

$$\text{For } \varepsilon \in (0,1) \text{ sufficiently small, } \mathbb{I}_O - \varepsilon \text{ should be desirable.} \tag{3.2}$$

In other words, there should be some positive—but possibly very small—betting rate at which you are willing to bet on O. Although we prefer not to stress this because it might easily be associated with an assumption of ideal precision—which do not want to make—it might be useful to realise that in terms of probabilities, this assessment simply means that O has some positive—but possibly very small—probability ε. We consider our assumption that Ω is finite to be crucial here, because it guarantees that O is 'sufficiently large' with respect to Ω. If Ω were to be infinite, we would not be inclined to adopt Assessment (3.2) for singleton events. Nevertheless, even for finite Ω, one might think that Assessment (3.2) is still not compelling; we leave it to the reader to decide for himself.

We want to stress that we are not assuming that O can occur. Our suggestion here is simply that *if such an assumption is made*, then Assessment (3.2) should be adopted. Our further analysis is based on this principle, and our conclusions therefore only apply to events O for which it is considered reasonable. The main idea is that, while constructing an updated model that is to be used after the occurrence of O, we are conducting a thought experiment in which O

[3] This can also be regarded as a consequence of our definition for "believing that O can occur" in Section 3.3.2$_{77}$.

can obviously occur, thereby allowing us to invoke Assessment $(3.2)_{\cap}$. However, outside of this thought experiment, we do not assume that O can occur.

In any case, if we choose to adopt Assessment $(3.2)_{\cap}$, the first problem we are confronted with is the meaning of 'sufficiently small': how small should ε be? An obvious suggestion is to return to the subject whose beliefs are modelled by means of $\mathscr{D}$—often an expert—and ask him to provide us with an $\varepsilon \in (0,1)$ such that, under the assumption that O can occur, $\mathbb{I}_O - \varepsilon$ would be desirable to him. Alternatively, the choice of ε can be based on someone else's opinion—possibly your own. For now, let us assume that ε is known; we will come back to this shortly.

We are now faced with a classical *belief expansion* problem [56, 66]: we have an initial belief model $\mathscr{D}$—a coherent set of desirable gambles—and want to incorporate the additional assessment that $\mathbb{I}_O - \varepsilon$ is desirable. Similarly to what is done in propositional logic, this can be achieved by considering the deductive closure of the union of these assessments, where in the language of sets of desirable gambles, the deductive closure is obtained by applying the natural extension operator $\mathscr{E}$; see References [38, 77, 114] for more information. Applying this procedure, we obtain the following set of desirable gambles:

$$\mathscr{E}_O^{\varepsilon}(\mathscr{D}) := \mathscr{E}(\mathscr{D} \cup \{\mathbb{I}_O - \varepsilon\}). \tag{3.3}$$

This set is not guaranteed to be coherent; the assessment $\mathbb{I}_O - \varepsilon$ can be inconsistent with $\mathscr{D}$. It is easy to see that $\mathscr{E}_O^{\varepsilon}(\mathscr{D})$ will be coherent if and only if $\varepsilon - \mathbb{I}_O \notin \mathscr{D}$. It is useful to compare this with what happens in propositional logic: a belief base can be expanded with a proposition a if and only if this belief base does not contain the negation of a.

Let us now come back to the problem of choosing ε. In practice, it is often very difficult to do so. The fact that we think that there should be some $\varepsilon \in (0,1)$ for which $\mathbb{I}_O - \varepsilon$ is desirable does not imply that we can actually provide such an ε. This typically occurs if $\mathscr{D}$ was provided by an expert that is no longer available for extra questions. Therefore, instead of fixing ε in some arbitrary way, we propose to restrict attention to the set of desirable gambles

$$\mathscr{E}_O(\mathscr{D}) := \bigcap_{\varepsilon \in (0,1)} \mathscr{E}_O^{\varepsilon}(\mathscr{D}), \tag{3.4}$$

which consists exactly of those gambles whose desirability can *always* be inferred by expanding $\mathscr{D}$ with $\mathbb{I}_O - \varepsilon$, *regardless* of the value of $\varepsilon \in (0,1)$. Assessment $(3.2)_{\cap}$ should clearly lead us to consider—at the very least—the gambles in $\mathscr{E}_O(\mathscr{D})$ as desirable. Other gambles might be desirable as well, but in order to find out which ones, some kind of domain expertise seems necessary. If this kind of expertise is not available, then $\mathscr{E}_O(\mathscr{D})$ seems to be a reasonable, conservative choice of model. Basically, we are then no longer adopting Assessment $(3.2)_{\cap}$, but merely some of its consequences.

3.4.4 *Investigating the consequences of the assessment*

In its current form, our expression for $\mathscr{E}_O(\mathscr{D})$ is rather indirect, making it difficult to get a feeling for which gambles it contains. Therefore, before drawing any conclusions with respect to updating, we start with a theoretical study of the set $\mathscr{E}_O(\mathscr{D})$. We restrict ourselves to results that are directly related to the updating problem we are trying to solve; see Appendix $3.\mathrm{A}_{91}$ for additional properties that are—although they are definitely relevant—not directly related to the present discussion.

The following proposition establishes that in order for $\mathscr{E}_O(\mathscr{D})$ to be coherent, it is sufficient as well as necessary for $\overline{P}_{\mathscr{D}}(O)$ to be strictly positive.

Proposition 15. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_\emptyset(\Omega)$. Then $\mathscr{E}_O(\mathscr{D})$ is incoherent if and only if $\overline{P}_{\mathscr{D}}(O) = 0$. Furthermore, if $\mathscr{E}_O(\mathscr{D})$ is incoherent, then $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$.*

Proof of Proposition 15. First, assume that $\mathscr{E}_O(\mathscr{D})$ is incoherent. Since, by construction, $\mathscr{E}_O(\mathscr{D})$ satisfies $\mathrm{D2}_{38}$–$\mathrm{D4}_{39}$, this implies that $\mathscr{E}_O(\mathscr{D})$ fails to satisfy $\mathrm{D1}_{38}$ and therefore, that $0 \in \mathscr{E}_O(\mathscr{D})$.

Fix $\varepsilon \in (0,1)$. By Equation $(3.4)_\cap$, we know that $0 \in \mathscr{E}_O^\varepsilon(\mathscr{D})$. By Equation $(3.3)_\cap$, and since $\mathscr{D}$ is coherent, this implies that either $\mathbb{I}_O - \varepsilon = 0$ or $g + \lambda(\mathbb{I}_O - \varepsilon) = 0$, with $\lambda \in \mathbb{R}_{>0}$ and $g \in \mathscr{D}$. The first option is impossible because $\varepsilon < 1$. Hence, we have that $\varepsilon - \mathbb{I}_O = {}^1\!/\!_\lambda g$ and therefore, by the coherence of $\mathscr{D}$ and because ${}^1\!/\!_\lambda > 0$ and $g \in \mathscr{D}$, that $\varepsilon - \mathbb{I}_O \in \mathscr{D}$. Since this holds for all $\varepsilon \in (0,1)$, we infer from Equation $(2.14)_{49}$ that $\overline{P}_{\mathscr{D}}(O) \leq 0$ and therefore, by Equation $(2.15)_{49}$, that $\overline{P}_{\mathscr{D}}(O) = 0$.

Next, still assuming that $\mathscr{E}_O(\mathscr{D})$ is incoherent, we set out to prove that $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$. Choose any $f \in \mathscr{G}(\Omega)$. Now choose any $\varepsilon \in (0,1)$. From the first part of this proof, we know that ${}^\varepsilon\!/\!_2 - \mathbb{I}_O \in \mathscr{D}$. Now choose $\alpha \in \mathbb{R}_{>0}$ high enough such that $f + \alpha > 0$. Then, by Equation $(3.3)_\cap$,

$$f = (f + \alpha) + {}^{2\alpha}\!/\!_\varepsilon({}^\varepsilon\!/\!_2 - \mathbb{I}_O) + {}^{2\alpha}\!/\!_\varepsilon(\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O^\varepsilon(\mathscr{D}).$$

Since this holds for all $\varepsilon \in \mathbb{R}_{>0}$, we infer from Equation $(3.4)_\cap$ that $f \in \mathscr{E}_O(\mathscr{D})$. Since this holds for all $f \in \mathscr{G}(\Omega)$, we find that $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$.

Finally, assume that $\overline{P}_{\mathscr{D}}(O) = 0$. Consider any $\varepsilon \in (0,1)$. Then by Equation $(2.14)_{49}$, we know that there is some $0 \leq \alpha < \varepsilon$ such that $\alpha - \mathbb{I}_O \in \mathscr{D}$. Hence, due to Equation $(3.3)_\cap$ and the coherence of $\mathscr{D}$, we obtain that $0 = (\varepsilon - \alpha) + (\alpha - \mathbb{I}_O) + (\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O^\varepsilon(\mathscr{D})$. Since this holds for all $\varepsilon \in (0,1)$, we infer from Equation $(3.4)_\cap$ that $0 \in \mathscr{E}_O(\mathscr{D})$, implying that $\mathscr{E}_O(\mathscr{D})$ is incoherent. □

Therefore, if $\overline{P}_{\mathscr{D}}(O) = 0$, the set $\mathscr{E}_O(\mathscr{D})$ is clearly not very useful. For now, let us assume that $\overline{P}_{\mathscr{D}}(O)$ is strictly positive. In that case, perhaps surprisingly, none of the gambles $\mathbb{I}_O - \varepsilon$ in Assessment $(3.2)_{81}$ is actually added to $\mathscr{D}$.

Proposition 16. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then for all $\varepsilon \in (0,1)$:*

$$\mathbb{I}_O - \varepsilon \in \mathscr{E}_O(\mathscr{D}) \Leftrightarrow \mathbb{I}_O - \varepsilon \in \mathscr{D}.$$

Proof of Proposition 16. Since $\mathscr{D}$ is clearly a subset of $\mathscr{E}_O(\mathscr{D})$, we only need to prove the direct implication. So consider any $\varepsilon \in (0,1)$ and assume that $\mathbb{I}_O - \varepsilon \in \mathscr{E}_O(\mathscr{D})$. We will prove that then $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. If we choose $0 < \varepsilon' < \varepsilon$, then by Equation $(3.4)_{82}$, we know that $\mathbb{I}_O - \varepsilon \in \mathscr{E}_O^{\varepsilon'}(\mathscr{D})$. By Equation $(3.3)_{82}$, and because $\mathscr{D}$ is coherent, we now have that, without loss of generality, either (a) $\mathbb{I}_O - \varepsilon \in \mathscr{D}$ or (b) $\mathbb{I}_O - \varepsilon = g + \lambda(\mathbb{I}_O - \varepsilon')$, with $g \in \mathscr{D} \cup \{0\}$ and $\lambda > 0$. In case of (a), the proof is concluded. In case of (b), we find that, with $\lambda' := 1 - \lambda$,

$$\lambda'(\mathbb{I}_O - \varepsilon) = \mathbb{I}_O - \varepsilon + \lambda(\varepsilon' - \mathbb{I}_O) + \lambda(\varepsilon - \varepsilon') = g + \lambda(\varepsilon - \varepsilon') \in \mathscr{D},$$

where the last inclusion is a consequence of the coherence of $\mathscr{D}$. If $\lambda' = 0$, we have that $0 \in \mathscr{D}$, contradicting the coherence of $\mathscr{D}$. If $\lambda' < 0$, then, again by the coherence of $\mathscr{D}$, we find that $\varepsilon - \mathbb{I}_O \in \mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D})$. By combining this with our assumption, and using the coherence of $\mathscr{E}_O(\mathscr{D})$ [which is a consequence of Proposition $15_{\frown}$ and our assumption that $\overline{P}_{\mathscr{D}}(O) > 0$], we find that $0 = (\varepsilon - \mathbb{I}_O) + (\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O(\mathscr{D})$, contradicting the coherence of $\mathscr{E}_O(\mathscr{D})$. The only remaining possibility is that $\lambda' > 0$. In this case, by the coherence of $\mathscr{D}$, we find that, indeed, $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. □

If $\underline{P}_{\mathscr{D}}(O)$ is strictly positive, we even find that $\mathscr{E}_O(\mathscr{D})$ is equal to $\mathscr{D}$.

Proposition 17. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\underline{P}_{\mathscr{D}}(O) > 0$. Then $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$.*

Proof of Proposition 17. By Equation $(2.13)_{49}$, and since $\underline{P}_{\mathscr{D}}(O) > 0$, we know that there is an $\varepsilon > 0$ such that $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. Since $\mathscr{D}$ is coherent, we also know that $\varepsilon < 1$. From Equation $(3.3)_{82}$, we now infer that $\mathscr{E}_O^{\varepsilon}(\mathscr{D}) = \mathscr{E}(\mathscr{D})$ and therefore, since $\mathscr{D}$ is coherent, that $\mathscr{E}_O^{\varepsilon}(\mathscr{D}) = \mathscr{D}$. Applying Equation $(3.4)_{82}$, we find that $\mathscr{E}_O(\mathscr{D}) \subseteq \mathscr{D}$. Since clearly also $\mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D})$, we conclude that $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$. □

By Equation $(2.15)_{49}$, the remaining option is that $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$. In that case, $\mathscr{E}_O(\mathscr{D})$ is completely characterised by the following theorem, the proof of which can be found in Appendix $3.\mathrm{B}_{97}$.

Theorem 18. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and let $O \in \mathscr{P}_{\emptyset}(\Omega)$ be any event such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then*

$$f \in \mathscr{E}_O(\mathscr{D})$$
$$\Leftrightarrow f \in \mathscr{D} \text{ or } (\forall \varepsilon \in (0,1))(\exists \lambda \in \mathbb{R}_{>0})\ f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathscr{D} \tag{3.5}$$
$$\Leftrightarrow f \in \mathscr{D} \text{ or } (\forall \varepsilon \in (0,1))(\exists \overline{\lambda} \in \mathbb{R}_{>0})(\forall \lambda \in (0,\overline{\lambda}])\ f + \lambda(\varepsilon - \mathbb{I}_O) \in \operatorname{int}(\mathscr{D}). \tag{3.6}$$

Although this result applies whenever $\overline{P}_{\mathscr{D}}(O) > 0$, we are of course mainly interested in the case $\underline{P}_{\mathscr{D}}(O) = 0$. If $\underline{P}_{\mathscr{D}}(O)$ is positive, we already know from Proposition $17_{\frown}$ that $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$. If $\underline{P}_{\mathscr{D}}(O) = 0$, $\mathscr{E}_O(\mathscr{D})$ might be strictly larger than $\mathscr{D}$; however, as the following result shows, only slightly—or should we say, marginally—so.

Corollary 19. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and any $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then $\mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D}) \subseteq \mathrm{cl}(\mathscr{D}) = \mathrm{cl}(\mathscr{E}_O(\mathscr{D}))$ and*

$$\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f) = \underline{P}_{\mathscr{D}}(f) \text{ for all } f \in \mathscr{G}(\Omega).$$

Proof of Corollary 19. We first prove that $\mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D}) \subseteq \mathrm{cl}(\mathscr{D})$. Since $\mathscr{D}$ is clearly a subset of $\mathscr{E}_O(\mathscr{D})$, it suffices to show that $\mathscr{E}_O(\mathscr{D}) \subseteq \mathrm{cl}(\mathscr{D})$. So fix any $f \in \mathscr{E}_O(\mathscr{D})$ and $\delta \in \mathbb{R}_{>0}$. By Equation $(2.5)_{41}$, we need to prove that $f + \delta \in \mathscr{D}$. If $f \in \mathscr{D}$, this follows trivially from the coherence of $\mathscr{D}$. Hence, by Theorem $18_{\frown}$, we can assume without loss of generality that

$$(\forall \varepsilon \in (0,1))(\exists \overline{\lambda} \in \mathbb{R}_{>0})(\forall \lambda \in (0,\overline{\lambda}])\ f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathrm{int}(\mathscr{D}) \subseteq \mathscr{D}. \quad (3.7)$$

Choose any $\varepsilon \in (0,1)$. Equation (3.7) implies that there is some $\overline{\lambda} \in \mathbb{R}_{>0}$ such that, with $\lambda := \min\{\delta, \overline{\lambda}\}$, $f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathscr{D}$. Since $f + \delta \geq f + \lambda(\varepsilon - \mathbb{I}_O)$, we infer from the coherence of $\mathscr{D}$ that $f + \delta \in \mathscr{D}$.

We now know that $\mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D})$ and $\mathscr{E}_O(\mathscr{D}) \subseteq \mathrm{cl}(\mathscr{D})$. By applying the operator cl to both sides of each of these inclusions, we find that

$$\mathrm{cl}(\mathscr{D}) \subseteq \mathrm{cl}(\mathscr{E}_O(\mathscr{D})) \subseteq \mathrm{cl}(\mathrm{cl}(\mathscr{D})) = \mathrm{cl}(\mathscr{D}),$$

where the last equality follows trivially from the fact that, for coherent $\mathscr{D}$, cl coincides with the topological closure operator. Hence, we may conclude that $\mathrm{cl}(\mathscr{D}) = \mathrm{cl}(\mathscr{E}_O(\mathscr{D}))$. By Equation $(2.8)_{41}$, and since $\mathscr{E}_O(\mathscr{D})$ is coherent because of Proposition 15_{83}, this implies that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f) = \underline{P}_{\mathscr{D}}(f)$ for all $f \in \mathscr{G}(\Omega)$. □

Hence, if $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$, the difference between $\mathscr{D}$ and $\mathscr{E}_O(\mathscr{D})$ is situated on their border. Nevertheless, this difference could be important, especially if we start to condition these models.

One particular conditional model that will be especially useful to us is $\mathscr{E}_O(\mathscr{D})\rfloor O$. It has the nice property that the associated set of almost desirable gambles does not depend on the border structure of $\mathscr{E}_O(\mathscr{D})$ or $\mathscr{D}$.

Proposition 20. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then for all $f \in \mathscr{G}(O)$:*

$$f \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O) \Leftrightarrow \mathbb{I}_O f \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})) \Leftrightarrow \mathbb{I}_O f \in \mathrm{cl}(\mathscr{D}).$$

Proof of Proposition 20$_{\frown}$. We prove that $f \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O) \Leftrightarrow \mathbb{I}_O f \in \mathrm{cl}(\mathscr{D})$; this implies the other equivalences because, as we have shown in the proof of Corollary 19$_{\frown}$, $\mathrm{cl}(\mathscr{E}_O(\mathscr{D})) = \mathrm{cl}(\mathscr{D})$.

First assume that $\mathbb{I}_O f \in \mathrm{cl}(\mathscr{D})$. Consider any $\lambda \in \mathbb{R}_{>0}$ and $\varepsilon \in (0,1)$. Then by Equation (2.5)$_{41}$, $\mathbb{I}_O(f+\lambda) + \lambda(\varepsilon - \mathbb{I}_O) = \mathbb{I}_O f + \lambda\varepsilon \in \mathscr{D}$ [because $\delta := \lambda\varepsilon \in \mathbb{R}_{>0}$]. Since this holds for all $\varepsilon \in (0,1)$, we infer from Theorem 18$_{84}$ that $\mathbb{I}_O(f+\lambda) \in \mathscr{E}_O(\mathscr{D})$ or, equivalently, that $f+\lambda \in \mathscr{E}_O(\mathscr{D})\rfloor O$. Since this holds for all $\lambda \in \mathbb{R}_{>0}$, we find that $f \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$.

Conversely, assume that $f \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$. Consider any $\delta \in \mathbb{R}_{>0}$ and choose $\alpha \in (0,1)$ such that $\alpha < \delta$. Then, by Equation (2.5)$_{41}$, $f + \alpha/2 \in \mathscr{E}_O(\mathscr{D})\rfloor O$ and therefore $\mathbb{I}_O(f+\alpha/2) \in \mathscr{E}_O(\mathscr{D})$. If $\mathbb{I}_O(f+\alpha/2) \in \mathscr{D}$, then by coherence of $\mathscr{D}$ also $\mathbb{I}_O(f+\delta) \in \mathscr{D}$. If $\mathbb{I}_O(f+\alpha/2) \notin \mathscr{D}$, then by applying Theorem 18$_{84}$ with $\varepsilon := \alpha/2$, we find some $\overline{\lambda} \in \mathbb{R}_{>0}$ such that

$$\mathbb{I}_O(f+\alpha/2) + \lambda(\alpha/2 - \mathbb{I}_O) \in \mathrm{int}(\mathscr{D}) \subseteq \mathscr{D},$$

with $\lambda := \min\{1, \overline{\lambda}\} \in (0, \overline{\lambda}]$. Since $\lambda \leq 1$ and $0 < \alpha < \delta$, we know that $\delta > \alpha \geq \alpha/2(1+\mathbb{I}_O) \geq \alpha/2(\lambda + \mathbb{I}_O)$, which implies that

$$\begin{aligned}\mathbb{I}_O f + \delta > \mathbb{I}_O f + \frac{\alpha}{2}(\lambda + \mathbb{I}_O) \geq \mathbb{I}_O f + \frac{\alpha}{2}(\lambda + \mathbb{I}_O) - \mathbb{I}_O\lambda \\ = \mathbb{I}_O(f + \frac{\alpha}{2}) + \lambda(\frac{\alpha}{2} - \mathbb{I}_O).\end{aligned}$$

Therefore, because $\mathbb{I}_O(f+\alpha/2) + \lambda(\alpha/2 - \mathbb{I}_O) \in \mathscr{D}$, the coherence of $\mathscr{D}$ implies that, again, $\mathbb{I}_O f + \delta \in \mathscr{D}$. Hence, in all cases, $\mathbb{I}_O f + \delta \in \mathscr{D}$. Since this holds for all $\delta \in \mathbb{R}_{>0}$, we infer from Equation (2.5)$_{41}$ that $\mathbb{I}_O f \in \mathrm{cl}(\mathscr{D})$. □

Therefore, if we are only interested in $\mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$, or equivalently, in $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$, all we need to know is $\mathrm{cl}(\mathscr{D})$, or equivalently, $\underline{P}_{\mathscr{D}}$. The connection between $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ and $\underline{P}_{\mathscr{D}}$ is provided by regular extension.

Corollary 21. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event O in $\mathscr{P}_\emptyset(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > 0$. Let $\underline{R}(\cdot\rfloor O)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, as given by Equation* (2.29)$_{63}$. *Then*

$$\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O) = \underline{R}(f\rfloor O) \quad \textit{for all } f \in \mathscr{G}(O).$$

Proof of Corollary 21. Fix any $f \in \mathscr{G}(O)$ and any $\mu \in \mathbb{R}$. By Proposition 20$_{\frown}$ and Equation (2.6)$_{41}$, we have that

$$\begin{aligned}\underline{P}_{\mathscr{D}}(\mathbb{I}_O[f-\mu]) \geq 0 &\Leftrightarrow \mathbb{I}_O[f-\mu] \in \mathrm{cl}(\mathscr{D}) \\ &\Leftrightarrow f-\mu \in \mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O) \Leftrightarrow \underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f-\mu) \geq 0.\end{aligned}$$

Since $\overline{P}_{\mathscr{D}}(O) > 0$, we know from Proposition 15$_{83}$ that $\mathscr{E}_O(\mathscr{D})$ is coherent. This implies that $\mathscr{E}_O(\mathscr{D})\rfloor O$ is coherent and therefore also that $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}$

is coherent, which in turn implies that $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f-\mu) = \underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f) - \mu$ [C8$_{49}$]. Furthermore, due to Equations (2.3)$_{40}$ and (2.9)$_{42}$, we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f) = \underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)$. Hence, putting it all together, we find that

$$\underline{P}_{\mathscr{D}}(\mathbb{I}_O[f-\mu]) \geq 0 \Leftrightarrow \underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O) \geq \mu.$$

Since this holds for all $\mu \in \mathbb{R}$, we infer from Equation (2.29)$_{63}$ and the fact that $\overline{P}_{\mathscr{D}}(O) > 0$ that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O) \geq \underline{R}(f\rfloor O)$.

Next, since $\mathscr{E}_O(\mathscr{D})$ is coherent, we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ is coherent with $\underline{P}_{\mathscr{E}_O(\mathscr{D})}$ and therefore also, since $\underline{P}_{\mathscr{E}_O(\mathscr{D})} = \underline{P}_{\mathscr{D}}$ because of Corollary 19$_{85}$, that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ is coherent with $\underline{P}_{\mathscr{D}}$. Since $\overline{P}_{\mathscr{D}}(O) > 0$ guarantees that $\underline{R}(\cdot\rfloor O)$ is the largest lower prevision on $\mathscr{G}(O)$ that is coherent with $\underline{P}_{\mathscr{D}}$, this implies that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O) \leq \underline{R}(f\rfloor O)$. □

3.4.5 *Turning the assessment into an updating rule*

Plenty of mathematics so far, but still no updating rule. So let us get back to the beginning: a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$. We are looking for an updated set of desirable gambles $\mathscr{D}_O$ that we intend to adopt after the occurrence of O, but which is specified in advance.

If O cannot occur, it makes no sense to provide an updated model $\mathscr{D}_O$, nor does it matter if we get it wrong. Hence, *while constructing the updated model* $\mathscr{D}_O$, we might as well assume that O can occur. Within this thought experiment, if we are willing to adopt Assessment (3.2)$_{81}$—and we will assume that we are—then as explained in Section 3.4.3$_{81}$, we can combine this assessment with our set of desirable gambles $\mathscr{D}$ to obtain a new set of gambles that is guaranteed to include $\mathscr{E}_O(\mathscr{D})$. Given that our interpretation for sets of desirable gambles is non-exhaustive, using $\mathscr{E}_O(\mathscr{D})$ itself puts us on the safe, conservative side. The rest of the argument now depends on whether or not $\mathscr{E}_O(\mathscr{D})$ is coherent.

If it is, it means that Assessment (3.2)$_{81}$ is compatible with $\mathscr{D}$, and we are led to consider $\mathscr{D}' = \mathscr{E}_O(\mathscr{D})$ as our set of desirable gambles. It is important to realise that $\mathscr{D}'$ is only adopted within the thought experiment of—in advance—constructing an updated model for after the occurrence of O; the belief model of our subject is still $\mathscr{D}$. Within this thought experiment, we can now apply the arguments of Section 3.2$_{73}$ to the set $\mathscr{D}'$, and this results in the use of $\mathscr{D}_O = \mathscr{D}'\rfloor O = \mathscr{E}_O(\mathscr{D})\rfloor O$ as a conservative choice of updated model. If $\underline{P}_{\mathscr{D}}(O) > 0$, Proposition 17$_{84}$ tells us that $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$, which implies that in that case, the updated model $\mathscr{D}_O = \mathscr{D}\rfloor O$ is exactly the same as the one we would have used had we not included Assessment (3.2)$_{81}$.

If $\mathscr{E}_O(\mathscr{D})$ is incoherent, or equivalently, by Proposition 15$_{83}$, if $\overline{P}_{\mathscr{D}}(O) = 0$, it means that Assessment (3.2)$_{81}$ is not compatible with $\mathscr{D}$. In fact, it even implies that $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$. Hence, in that case, it clearly makes no sense to use $\mathscr{E}_O(\mathscr{D})$ as the set $\mathscr{D}'$ that is adopted within the thought experiment of construction $\mathscr{D}_O$. A possible solution to this problem is to drop Assessment (3.2)$_{81}$

and take $\mathscr{D}'$ to be equal to $\mathscr{D}$. By applying the arguments of Section 3.2_{73} to $\mathscr{D}'$, this results in the use of $\mathscr{D}_O = \mathscr{D}'\rfloor O = \mathscr{D}\rfloor O$ as our updated model. However, the same conclusion can also be reached without dropping Assessment $(3.2)_{81}$. The fact that Assessment $(3.2)_{81}$ is incompatible with $\mathscr{D}$ does not imply that this is also the case for subsets of $\mathscr{D}$. Indeed, in particular, the gambles that are of the form $\mathbb{I}_O f$, with $f \in \mathscr{D}\rfloor O$, will always be compatible with Assessment $(3.2)_{81}$.[4] Since these gambles are the only ones that are relevant for Equation $(3.1)_{75}$, the argumentation in Section 3.2_{73} can still be applied, and it leads us to conclude that using $\mathscr{D}_O = \mathscr{D}\rfloor O$ as our updated model is a conservative approach.

By combining the case in which $\mathscr{E}_O(\mathscr{D})$ is coherent with the one in which it is not, we end up with the following updating rule:

$$\mathscr{D}_O^{\mathrm{r}} := \begin{cases} \mathscr{E}_O(\mathscr{D})\rfloor O = \mathscr{D}\rfloor O & \text{if } \underline{P}_{\mathscr{D}}(O) > 0 \\ \mathscr{E}_O(\mathscr{D})\rfloor O & \text{if } \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0 \\ \mathscr{D}\rfloor O & \text{if } \overline{P}_{\mathscr{D}}(O) = 0 \end{cases} \quad \text{for all } O \in \mathscr{P}_{\emptyset}(\Omega). \tag{3.8}$$

It is identical to simply conditioning, except if $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$, in which case it is guaranteed to be at least as informative—since $\mathscr{E}_O(\mathscr{D}) \supseteq \mathscr{D}$. Nevertheless, despite it being more informative, this strategy still avoids partial loss, even if we announce updated models for multiple events O at the same time.[5] Furthermore, it clearly also satisfies the second upper constraint that was discussed in Section $3.3.2_{77}$.

If we let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, then by Theorem 18_{84}, for all $O \in \mathscr{P}_{\emptyset}(\Omega)$, the lower prevision $\underline{P}_{\mathscr{D}_O^{\mathrm{r}}}$ that is associated with the updated set $\mathscr{D}_O^{\mathrm{r}}$ is given by

$$\underline{P}_{\mathscr{D}_O^{\mathrm{r}}}(f) = \begin{cases} \underline{R}(f\rfloor O) = \underline{P}_{\mathscr{D}}(f\rfloor O) & \text{if } \underline{P}_{\mathscr{D}}(O) > 0 \\ \underline{R}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0 \\ \underline{P}_{\mathscr{D}}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) = 0 \end{cases} \tag{3.9}$$

for all $f \in \mathscr{G}(O)$.

[4] Because the coherence of $\mathscr{D}$ implies that $\mathbb{I}_O f \not\leq 0$, which in turn implies that $f \not\leq 0$. Hence, for any $\varepsilon \in (0,1)$ and any $\lambda \in \mathbb{R}_{>0}$, it holds that $f + \lambda(1-\varepsilon) \not\leq 0$ and therefore also that $\mathbb{I}_O f + \lambda(\mathbb{I}_O - \varepsilon) \not\leq 0$.

[5] There is a coherent set of desirable gambles $\mathscr{D}^* = \mathscr{E}^{\mathrm{r}}(\mathscr{D})$ [see Corollary 32_{96} in Appendix $3.\mathrm{A}_{91}$] for which it holds that $\mathscr{D} \subseteq \mathscr{D}^*$ and $\mathscr{E}_O(\mathscr{D}) \subseteq \mathscr{D}^*$ for all $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$. Hence, by Equation (3.8), for all $O \in \mathscr{P}_{\emptyset}(\Omega)$, $\mathscr{D}_O^{\mathrm{r}} \subseteq \mathscr{D}^*\rfloor O$. As explained in Section $3.3.2_{77}$, this implies that updating by means of the updated sets $\mathscr{D}_O^{\mathrm{r}}$ is guaranteed to avoid partial loss, even if multiple updated sets are announced at the same time.

3.4.6 *Translating it into a justification for regular extension*

The updating rule that was derived in the previous section can be translated to the framework of lower previsions in a straightforward manner. For any coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ and any event $O \in \mathscr{P}_{\emptyset}(\Omega)$, we simply apply Equation $(3.8)_{\frown}$ to the smallest associated set of desirable gambles $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and consider the corresponding lower prevision $\underline{P}_{\mathscr{D}_O^{\mathrm{r}}}$, as given by Equation $(3.9)_{\frown}$. In order to reflect this particular choice of $\mathscr{D}$ in our notation, we denote this lower prevision as $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor O)$. The following result establishes that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision on $\mathscr{C}(\Omega)$ and expresses it directly in terms of $\underline{P}(\cdot\rfloor\cdot)$.

Corollary 22. *Consider a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with arbitrary domain $\mathscr{C}$. Let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension and let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{E}(\cdot\rfloor\Omega)$. Then $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ is a coherent lower prevision on $\mathscr{C}(\Omega)$ and, for all $(f,O) \in \mathscr{C}(\Omega)$:*

$$\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(f\rfloor O) = \begin{cases} \underline{R}(f\rfloor O) = \underline{E}(f\rfloor O) & \text{if } \underline{E}(O\rfloor\Omega) > 0 \\ \underline{R}(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) > \underline{E}(O\rfloor\Omega) = 0 \\ \underline{E}(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) = 0 \end{cases}$$

Proof of Corollary 22. The equality follows from Equation $(3.9)_{\frown}$ and the fact that, with $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, by definition of natural extension, $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot) = \underline{E}(\cdot\rfloor\cdot)$ and therefore also $\underline{P}_{\mathscr{D}}(\cdot) = \underline{E}(\cdot\rfloor\Omega)$. Since $\underline{P}(\cdot\rfloor\cdot)$ is coherent, $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent as well, and therefore, Corollary 31_{96} [see Appendix $3.\mathrm{A}_{91}$] implies that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(\cdot\rfloor\cdot)$ is a coherent lower prevision on $\mathscr{C}(\Omega)$. □

Furthermore, if we apply this procedure to an unconditional lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, the resulting conditional lower prevision $\underline{R}_{\underline{P}}(\cdot\rfloor\cdot)$ is equal to the regular extension of $\underline{P}$.

Corollary 23. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be its regular extension. Then*

$$\underline{R}_{\underline{P}}(f\rfloor O) = \underline{R}(f\rfloor O) \quad \text{for all } (f,O) \in \mathscr{C}(\Omega).$$

Proof of Corollary 23. As explained in the proof of Corollary 22, $\underline{R}_{\underline{P}}(\cdot\rfloor\cdot)$ is equal to $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(\cdot\rfloor\cdot)$, with, in this particular case, $\mathscr{D} = \mathscr{E}_{\underline{P}} = \mathscr{D}_{\underline{P}}$. The proof now follows immediately from Corollary 32_{96} [see Appendix $3.\mathrm{A}_{91}$]. □

Given that the only assumptions that were made to obtain $\underline{R}(\cdot\rfloor\cdot)$ are (i) that we are working within the setting described in Section 3.1_{70} and (ii) that within the though experiment of constructing the updated model, the subject whose beliefs are being modelled is willing to adopt Assessment $(3.2)_{81}$, we have finally found our justification for updating by means of regular extension. It follows from our results that, whenever these two conditions are met, regular extension serves as a conservative updating strategy.

3.4.7 *A generalisation of regular extension*

Corollary $23_{\frown}$ implies that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ can be regarded as a generalisation of the notion of regular extension to coherent *conditional* lower previsions. Therefore, from now on, for any conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, we will refer to $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ as the *regular extension* of $\underline{P}(\cdot\rfloor\cdot)$ and, whenever it is clear from the context which conditional lower prevision it is derived from, we drop the index and write $\underline{R}(\cdot\rfloor\cdot)$ instead of $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$. As we have seen in the previous section, this generalised notion of regular extension—similarly to the version for unconditional lower previsions—can be justified as a conservative updating rule.

If we are considering the regular extension $\underline{R}(\cdot\rfloor\cdot)$ of a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, some notational confusion might arise, because in Corollary $22_{\frown}$, we used $\underline{R}(\cdot\rfloor\cdot)$ to refer to the regular extension of $\underline{E}(\cdot\rfloor\Omega)$, with $\underline{E}(\cdot\rfloor\cdot)$ the natural extension of $\underline{P}(\cdot\rfloor\cdot)$. In order to avoid this confusion, we will from now on denote the regular extension of $\underline{E}(\cdot\rfloor\Omega)$ by $\underline{R}^*(\cdot\rfloor\cdot)$. Similarly, as we did in Section $3.2.3_{75}$, we denote the natural extension of $\underline{E}(\cdot\rfloor\Omega)$ by $\underline{E}^*(\cdot\rfloor\cdot)$. Using these conventions, the regular extension $\underline{R}(\cdot\rfloor\cdot)$ of $\underline{P}(\cdot\rfloor\cdot)$ is given, for all $(f,O)\in\mathscr{C}(\Omega)$, by

$$\underline{R}(f\rfloor O)=\begin{cases}\underline{R}^*(f\rfloor O)=\underline{E}^*(f\rfloor O)=\underline{E}(f\rfloor O) & \text{if } \underline{E}(O\rfloor\Omega)>0\\ \underline{R}^*(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega)>\underline{E}(O\rfloor\Omega)=0\\ \underline{E}(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega)=0.\end{cases} \tag{3.10}$$

It coincides with the natural extension $\underline{E}(f\rfloor O)$ whenever $\underline{E}(O\rfloor\Omega)>0$ or $\overline{E}(O\rfloor\Omega)=0$. If $\overline{E}(O\rfloor\Omega)>\underline{E}(O\rfloor\Omega)=0$, regular extension may differ from natural extension and is guaranteed to dominate it.

If $\underline{P}(f\rfloor\Omega)$ is defined for all $f\in\mathscr{G}(\Omega)$, then $\underline{E}(\cdot\rfloor\Omega)$ will be equal to $\underline{P}(\cdot\rfloor\Omega)$ and $\underline{E}^*(\cdot\rfloor\cdot)$ and $\underline{R}^*(\cdot\rfloor\cdot)$ can be taken to be the natural and regular extension of $\underline{P}(\cdot\rfloor\Omega)$, respectively. If $\underline{P}(\cdot\rfloor\Omega)$ is known—or if $\underline{E}(\cdot\rfloor\Omega)$ can be calculated—$\underline{R}^*(\cdot\rfloor O)$ can be obtained by applying the computational techniques of Section $2.7.3_{63}$. The result is guaranteed to coincide with $\underline{R}(\cdot\rfloor O)$ whenever $\overline{P}(O\rfloor\Omega)>0$—$\overline{E}(O\rfloor\Omega)>0$—but will be vacuous and might therefore only be a (safe) lower approximation of $\underline{R}(\cdot\rfloor O)$ if $\overline{P}(O\rfloor\Omega)=0$—$\overline{E}(O\rfloor\Omega)=0$.

The natural extension $\underline{E}(\cdot\rfloor\cdot)$ of a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ can be given a sensitivity analysis interpretation in terms of linear conditional previsions: as we know from Equation $(2.18)_{53}$, it is the lower envelope of the set $\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ consisting of the linear conditional previsions that dominate $\underline{P}(\cdot\rfloor\cdot)$. A similar interpretation can be given to the regular extension $\underline{R}(\cdot\rfloor\cdot)$ of $\underline{P}(\cdot\rfloor\cdot)$ as well. If $\overline{E}(O\rfloor\Omega)=0$ or $\underline{E}(O\rfloor\Omega)>0$, then $\underline{R}(\cdot\rfloor O)$ is equal to $\underline{E}(\cdot\rfloor O)$ and therefore borrows its sensitivity analysis interpretation. If $\overline{E}(O\rfloor\Omega)>0$—which, again, includes the case $\underline{E}(O\rfloor\Omega)>0$—then $\underline{R}(\cdot\rfloor O)$ is equal to the regular ex-

tension $\underline{R}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$, and therefore given by

$$\begin{aligned}\underline{R}(f\rfloor O) &= \inf\{P(f\rfloor O)\colon P\in\mathscr{M}_{\underline{E}(\cdot\rfloor\Omega)} \text{ and } P(O)>0\}\\ &= \inf\{P(f\rfloor O)\colon P(\cdot\rfloor\cdot)\in\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)} \text{ and } P(O\rfloor\Omega)>0\}\end{aligned} \tag{3.11}$$

for all $f\in\mathscr{G}(O)$, where the last equality follows because $\mathscr{M}_{\underline{E}(\cdot\rfloor\cdot)}=\mathscr{M}_{\underline{P}(\cdot\rfloor\cdot)}$ and because, if $P(\cdot\rfloor\cdot)$ is a linear conditional prevision on $\mathscr{C}(\Omega)$ such that $P(O\rfloor\Omega)>0$, then $P(f\rfloor O)$ is fully determined by $P(\cdot\rfloor\Omega)$ through Bayes's rule [P4$_{50}$].

3.A TECHNICAL RESULTS RELATED TO UPDATING

Besides the ones that were already discussed in Section 3.4.4$_{83}$, the operator $\mathscr{E}_O$ that was introduced in Section 3.4.3$_{81}$ has some additional nice properties as well. Since they do not fit nicely into the main discussion of the paper, we gather them in this appendix.

We start by investigating what happens if we apply $\mathscr{E}_O$ multiple, say $n\in\mathbb{N}$, times. Consider a sequence of events $O_i\in\mathscr{P}_\emptyset(\Omega)$, $i\in\{1,\dots,n\}$, and let us apply the corresponding sequence of operators $\mathscr{E}_{O_i}$, $i\in\{1,\dots,n\}$, to a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$, in that order, resulting in a set of gambles

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) := \mathscr{E}_{O_n}(\mathscr{E}_{O_{n-1}}(\cdots\ \mathscr{E}_{O_2}(\mathscr{E}_{O_1}(\mathscr{D}))\ \cdots)).$$

What does this set look like? And does it depend on the order in which the operators are applied? The following results provide an answer to these questions.

Proposition 24. *Consider a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$ and a sequence of $n\in\mathbb{N}$ events $O_i\in\mathscr{P}_\emptyset(\Omega)$, $i\in\{1,\dots,n\}$. Then $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is coherent if and only if $\overline{P}_{\mathscr{D}}(O_i)>0$ for all $i\in\{1,\dots,n\}$. If it is coherent, then* $\mathrm{cl}(\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}))=\mathrm{cl}(\mathscr{D})$. *If it is incoherent, then $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})=\mathscr{G}(\Omega)$.*

Proof of Proposition 24. We provide a proof by induction. For $n=1$, the result follows trivially from Proposition 15$_{83}$ and Corollary 19$_{85}$. Consider now the case $n>1$. Then by the induction hypothesis, we may assume that the result is true for $n-1$.

First, assume that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is coherent. Now assume *ex absurdo* that $\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ is incoherent. The induction hypothesis then allows us to infer that $\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})=\mathscr{G}(\Omega)$, which implies that

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})=\mathscr{E}_{O_n}(\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D}))=\mathscr{E}_{O_n}(\mathscr{G}(\Omega))=\mathscr{G}(\Omega)$$

is incoherent, a contradiction. Hence, we find that $\mathscr{D}' := \mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ must be coherent. By the induction hypothesis, this implies that $\mathrm{cl}(\mathscr{D}')=\mathrm{cl}(\mathscr{D})$ and, for all $i\in\{1,\dots,n-1\}$, that $\overline{P}_{\mathscr{D}}(O_i)>0$. Since $\mathscr{D}'$ and $\mathscr{D}$

are both coherent, we infer from $\mathrm{cl}(\mathscr{D}') = \mathrm{cl}(\mathscr{D})$ and Equation $(2.8)_{41}$ that $\underline{P}_{\mathscr{D}'} = \underline{P}_{\mathscr{D}}$ and therefore, by conjugacy, that $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_{\mathscr{D}}(O_n)$. Since we have that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}')$, we infer from Proposition 15_{83} and the coherence of $\mathscr{D}'$ and $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ that $\overline{P}_{\mathscr{D}'}(O_n) > 0$, which in turn implies that $\mathrm{cl}(\mathscr{E}_{O_n}(\mathscr{D}')) = \mathrm{cl}(\mathscr{D}')$ [Corollary 19_{85}]. In conclusion, we have found that $\mathrm{cl}(\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})) = \mathrm{cl}(\mathscr{E}_{O_n}(\mathscr{D}')) = \mathrm{cl}(\mathscr{D}') = \mathrm{cl}(\mathscr{D})$ and that, for all $i \in \{1,\dots,n\}$, $\overline{P}_{\mathscr{D}}(O_i) > 0$.

Next, assume that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is incoherent. We need to prove that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{G}(\Omega)$ and that $\overline{P}_{\mathscr{D}}(O_i) = 0$ for some $i \in \{1,\dots,n\}$. This clearly the case if $\mathscr{D}' := \mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ is incoherent, because it then follows from the induction hypothesis (a) that $\overline{P}_{\mathscr{D}}(O_i) = 0$ for some $i \in \{1,\dots,n-1\}$ and (b) that $\mathscr{D}' = \mathscr{G}(\Omega)$ and therefore also $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}') = \mathscr{G}(\Omega)$. Hence, without loss of generality, we can assume that $\mathscr{D}'$ is coherent. As shown in the first part of this proof, this implies that $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_{\mathscr{D}}(O_n)$. Since $\mathscr{D}'$ is coherent and $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}')$, we can now combine Proposition 15_{83} with the fact that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is incoherent to infer that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{G}(\Omega)$ and $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_{\mathscr{D}}(O_n) = 0$. □

Proposition 25. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a sequence of $n \in \mathbb{N}$ events $O_i \in \mathscr{P}_{\emptyset}(\Omega)$, $i \in \{1,\dots,n\}$. Then*

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \bigcup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D}).$$

Proof of Proposition 25. Let us define $\mathscr{D}_0 := \mathscr{D}$ and, for all $i \in \{1,\dots,n\}$, $\mathscr{D}_i := \mathscr{E}_{O_i}(\mathscr{D}_{i-1}) = \mathscr{E}_{O_1\dots,O_i}(\mathscr{D})$. Since $\mathscr{D}_0 \subseteq \mathscr{D}_1 \subseteq \cdots \subseteq \mathscr{D}_{n-1} \subseteq \mathscr{D}_n$, we find that, for all $i \in \{1,\dots,n\}$,

$$\mathscr{E}_{O_i}(\mathscr{D}) = \mathscr{E}_{O_i}(\mathscr{D}_0) \subseteq \mathscr{E}_{O_i}(\mathscr{D}_{i-1}) = \mathscr{D}_i \subseteq \mathscr{D}_n = \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}).$$

Hence, we are left to prove that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) \subseteq \cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$. So let us fix any $f \in \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$. We will prove that $f \in \cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$. By Proposition 15_{83}, we may assume without loss of generality that, for all $i \in \{1,\dots,n\}$, $\overline{P}_{\mathscr{D}}(O_i) > 0$, because otherwise, $\cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$ would be equal to $\mathscr{G}(\Omega)$, which would make the proof trivial. Now let i^* be the smallest $i \in \{0,\dots,n\}$ for which it holds that $f \in \mathscr{D}_i$ [since $f \in \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{D}_n$, i^* always exists]. If $i^* = 0$, then $f \in \mathscr{D}$, which makes the proof trivial. Hence, without loss of generality, we may assume that $i^* \geq 1$. This allows us to consider the set $\mathscr{D}_{i^*-1}$, of which we know, by definition of i^*, that $f \notin \mathscr{D}_{i^*-1}$. Since $\overline{P}_{\mathscr{D}}(O_i) > 0$ for all $i \in \{1,\dots,i^*-1\}$, we know that $\mathscr{D}_{i^*-1}$ is coherent and that $\mathrm{cl}(\mathscr{D}_{i^*-1}) = \mathrm{cl}(\mathscr{D})$ [this is trivial if $i^*-1 = 0$ and otherwise follows from Proposition $24_{\frown}$ with $n = i^*-1$]. Since $\mathscr{D}_{i^*-1}$ is coherent and $\overline{P}_{\mathscr{D}}(O_{i^*}) > 0$, we can combine Theorem 18_{84} with the fact that $f \in \mathscr{D}_{i^*} \setminus \mathscr{D}_{i^*-1} = \mathscr{E}_{O_{i^*}}(\mathscr{D}_{i^*-1}) \setminus \mathscr{D}_{i^*-1}$ to infer that

$$(\forall \varepsilon \in (0,1))(\exists \overline{\lambda} \in \mathbb{R}_{>0})(\forall \lambda \in (0,\overline{\lambda}])\ f + \lambda(\varepsilon - \mathbb{I}_{O_{i^*}}) \in \mathrm{int}(\mathscr{D}_{i^*-1}). \quad (3.12)$$

Since $\mathscr{D}$ and $\mathscr{D}_{i^*-1}$ are both coherent, $\mathrm{cl}(\mathscr{D}_{i^*-1}) = \mathrm{cl}(\mathscr{D})$ implies that $\mathrm{int}(\mathscr{D}_{i^*-1}) = \mathrm{int}(\mathscr{D})$ [Equation $(2.8)_{41}$]. Therefore, since $\mathscr{D}$ is coherent and $\overline{P}_{\mathscr{D}}(O_{i^*}) > 0$, we can infer from Equation $(3.12)_{\frown}$ and Theorem 18_{84} that $f \in \mathscr{E}_{O_{i^*}}(\mathscr{D})$ and therefore also, that $f \subseteq \cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$. □

It follows from this last result that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is fully determined by the set of events $\mathscr{O} = \{O_i \colon i \in \{1,\dots,n\}\}$, and we will therefore simply denote it by $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$. The order of the events $O_1,\dots,O_n$ does not matter, neither does the fact that some of the events might appear multiple times. For this reason, from now on, we no longer consider sequences of events, but non-empty subsets $\mathscr{O}$ of $\mathscr{P}_{\emptyset}(\Omega)$.

For any non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$, and any coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$, we define

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D}) := \bigcup_{O \in \mathscr{O}} \mathscr{E}_O(\mathscr{D}).$$

By the results above, and since $\mathscr{P}_{\emptyset}(\Omega)$ and therefore also $\mathscr{O}$ is finite, the following properties are immediate.

Corollary 26. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$. Then $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is coherent if and only if $\overline{P}_{\mathscr{D}}(O) > 0$ for all $O \in \mathscr{O}$. If it is coherent, then $\mathrm{cl}(\mathscr{E}_{\mathscr{O}}(\mathscr{D})) = \mathrm{cl}(\mathscr{D})$. If it is incoherent, then $\mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \mathscr{G}(\Omega)$.*

Proof of Corollary 26. Trivial from Propositions 24_{91} and $25_{\frown}$. □

The necessary and sufficient condition for $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ to be coherent—$\overline{P}_{\mathscr{D}}(O) > 0$ for all $O \in \mathscr{O}$—that is provided in the result above simplifies if $\mathscr{O} = \mathscr{P}_{\emptyset}(\Omega)$. In that case, because $\mathscr{D}$ and therefore also $\underline{P}_{\mathscr{D}}$ is coherent, this condition is satisfied if and only if $\overline{P}_{\mathscr{D}}(\omega) > 0$ for all $\omega \in \Omega$.

It is also possible to characterise $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ differently, in a way that closely resembles our definition for $\mathscr{E}_O(\mathscr{D})$.

Theorem 27. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$. Let $\mathscr{E}^{\varepsilon}_{\mathscr{O}}(\mathscr{D}) := \mathscr{E}(\mathscr{D} \cup \{\mathbb{I}_O - \varepsilon_O \colon O \in \mathscr{O}\})$ for any $\varepsilon \in (0,1)^{\mathscr{O}}$.[6] Then*

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \bigcap_{\varepsilon \in (0,1)^{\mathscr{O}}} \mathscr{E}^{\varepsilon}_{\mathscr{O}}(\mathscr{D}).$$

[6] An element ε of $(0,1)^{\mathscr{O}}$ is a map from $\mathscr{O}$ to $(0,1)$. For any $O \in \mathscr{O}$, the corresponding value is an element of $(0,1)$ and will be denoted by ε_O.

Proof of Theorem 27$_{\frown}$. For the sake of notational convenience, let us denote the right-hand side of the desired equality by $\mathscr{E}^*_{\mathscr{O}}(\mathscr{D})$. We will only prove that $\mathscr{E}^*_{\mathscr{O}}(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. The converse inclusion holds trivially because, for all $O \in \mathscr{O}$, $\mathscr{E}_O(\mathscr{D})$ is clearly a subset of $\mathscr{E}^*_{\mathscr{O}}(\mathscr{D})$.

Fix $f \in \mathscr{E}^*_{\mathscr{O}}(\mathscr{D})$ and assume *ex absurdo* that $f \notin \mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \bigcup_{O\in\mathscr{O}} \mathscr{E}_O(\mathscr{D})$, which already implies that $f \notin \mathscr{D}$. Then for all $O \in \mathscr{O}$, $f \notin \mathscr{E}_O(\mathscr{D})$ and therefore, there is some $\varepsilon_O \in (0,1)$ such that $f \notin \mathscr{E}^{\varepsilon_O}_O(\mathscr{D})$. Let $\varepsilon_{\min} := \min_{O\in\mathscr{O}} \varepsilon_O$ [this minimum is well-defined because Ω and therefore also $\mathscr{P}_\emptyset(\Omega)$ and $\mathscr{O}$ are finite] and let $\varepsilon' \in (0,1)^{\mathscr{O}}$ be defined, for all $O \in \mathscr{O}$, by $\varepsilon'_O := \varepsilon_{\min}|\mathscr{O}|^{-1}$. Since we know that $f \in \mathscr{E}^*_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D} \subseteq \mathscr{E}^{\varepsilon'}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$, we can infer from the definition of $\mathscr{E}^{\varepsilon'}_{\mathscr{O}}(\mathscr{D})$ and the coherence of $\mathscr{D}$ that there are $n \in \mathbb{N}$ such that $n \leq |\mathscr{O}|$, $g \in \mathscr{D} \cup \{0\}$, $(\forall i \in \{1,\dots,n\})\ \lambda_i \in \mathbb{R}_{>0}$ and $O_i \in \mathscr{O}$ such that

$$f = g + \sum_{i=1}^{n} \lambda_i(\mathbb{I}_{O_i} - \varepsilon'_{O_i}) = g + \sum_{i=1}^{n} \lambda_i \mathbb{I}_{O_i} - \frac{\varepsilon_{\min}}{|\mathscr{O}|}\sum_{i=1}^{n}\lambda_i. \tag{3.13}$$

Let $\lambda_{\max} := \max\{\lambda_i\colon 1 \leq i \leq n\}$ and $i_{\max} \in \arg\max\{\lambda_i\colon 1 \leq i \leq n\}$, then

$$\begin{aligned}\sum_{i=1}^{n} \lambda_i \mathbb{I}_{O_i} - \frac{\varepsilon_{\min}}{|\mathscr{O}|}\sum_{i=1}^{n}\lambda_i \geq \lambda_{\max}\mathbb{I}_{O_{i_{\max}}} - \frac{\varepsilon_{\min}}{|\mathscr{O}|} n\lambda_{\max} &\geq \lambda_{\max}\mathbb{I}_{O_{i_{\max}}} - \varepsilon_{\min}\lambda_{\max} \\ &\geq \lambda_{\max}(\mathbb{I}_{O_{i_{\max}}} - \varepsilon_{O_{i_{\max}}})\end{aligned}$$

and therefore, by Equation (3.13), $f \geq g + \lambda_{\max}(\mathbb{I}_{O_{i_{\max}}} - \varepsilon_{O_{i_{\max}}})$, which implies that $f \in \mathscr{E}^{\varepsilon_{O_{i_{\max}}}}_{O_{i_{\max}}}(\mathscr{D})$, a contradiction. Hence, we conclude that $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. Since this holds for all $f \in \mathscr{E}^*_{\mathscr{O}}(\mathscr{D})$, we find that $\mathscr{E}^*_{\mathscr{O}}(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. □

Again, as was the case for $\mathscr{E}_O(\mathscr{D})$, $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is closely connected with regular extension. For every event $O \in \mathscr{O}$, $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O)$ is equal to the regular extension $\underline{R}(\cdot\rfloor O)$ of $\underline{P}_{\mathscr{D}}$.

Proposition 28. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_\emptyset(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, defined by Equation* (2.29)$_{63}$. *Then* $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})} = \underline{P}_{\mathscr{D}}$,

$$\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) = \underline{R}(\cdot\rfloor O) \text{ for all } O \in \mathscr{O}$$

and

$$\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) = \underline{P}_{\mathscr{D}}(\cdot\rfloor O) \text{ for all } O \in \mathscr{P}_\emptyset(\Omega) \text{ such that } \overline{P}_{\mathscr{D}}(O) = 0.$$

Proof of Proposition 28. We know from Corollary 26$_{\frown}$ that $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is coherent, and that $\mathrm{cl}(\mathscr{E}_{\mathscr{O}}(\mathscr{D})) = \mathrm{cl}(\mathscr{D})$. Therefore, by Equation (2.8)$_{41}$, $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})} = \underline{P}_{\mathscr{D}}$.

Fix $O \in \mathscr{O}$. The coherence of $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O)$ is coherent with $\underline{P}_{\mathscr{E}_O(\mathscr{D})}$ and therefore also with $\underline{P}_{\mathscr{D}}$. Since $\overline{P}_{\mathscr{D}}(O) > 0$, this implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) \leq \underline{R}(\cdot\rfloor O)$ [see Section 2.7.2$_{61}$]. By Corollary 21$_{86}$, we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O) = \underline{R}(\cdot\rfloor O)$. Furthermore, $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) \geq \underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ because $\mathscr{E}_O(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. Putting it all together, we find that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) = \underline{R}(\cdot\rfloor O)$.

Consider any $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) = 0$. By Lemma 29, we know that $\mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O = \mathscr{D}\rfloor O$, which implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot\rfloor O) = \underline{P}_{\mathscr{D}}(\cdot\rfloor O)$. □

Lemma 29. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Then*

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O = \mathscr{D}\rfloor O \text{ for all } O \in \mathscr{P}_{\emptyset}(\Omega) \text{ such that } \overline{P}_{\mathscr{D}}(O) = 0.$$

Proof of Lemma 29. Fix $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) = 0$. We only prove that $\mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O \subseteq \mathscr{D}\rfloor O$, because the other inclusion follows trivially from the fact that $\mathscr{D} \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$.

Consider any $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O$ and assume *ex absurdo* that $f \notin \mathscr{D}\rfloor O$. We then have that $\mathbb{I}_O f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$, which implies that there is some $O^* \in \mathscr{O}$ such that $\mathbb{I}_O f \in \mathscr{E}_{O^*}(\mathscr{D}) \setminus \mathscr{D}$. Using Lemma 30, we find that $\overline{P}_{\mathscr{D}}(\mathbb{I}_O f) > 0$. Choose $\lambda \in \mathbb{R}_{>0}$ such that $\lambda \geq \max f$. Then by the coherence of $\underline{P}_{\mathscr{D}}$, we have that $\overline{P}_{\mathscr{D}}(\mathbb{I}_O f) \leq \overline{P}_{\mathscr{D}}(\mathbb{I}_O \lambda) = \lambda \overline{P}_{\mathscr{D}}(O) = 0$. This is a contradiction, allowing us to infer that $f \in \mathscr{D}\rfloor O$. Since this holds for all $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O$, we find that $\mathscr{E}_{\mathscr{O}}(\mathscr{D})\rfloor O \subseteq \mathscr{D}\rfloor O$. □

Lemma 30. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Then for all $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$: $0 = \underline{P}_{\mathscr{D}}(f) < \overline{P}_{\mathscr{D}}(f)$.*

Proof of Lemma 30. Consider any $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$, which implies that there is some $O \in \mathscr{O}$ such that $f \in \mathscr{E}_O(\mathscr{D}) \setminus \mathscr{D}$. Choose $\varepsilon \in (0,1)$ such that $\varepsilon < \overline{P}_{\mathscr{D}}(O)$. Since $\overline{P}_{\mathscr{D}}(O) > 0$, we know from Theorem 18$_{84}$ that there is some $\lambda \in \mathbb{R}_{>0}$ such that $f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathscr{D}$, which implies that $\underline{P}_{\mathscr{D}}(f + \lambda[\varepsilon - \mathbb{I}_O]) \geq 0$. Using the coherence of $\underline{P}_{\mathscr{D}}$ [C2$_{48}$, C5$_{48}$ and C8$_{49}$], this allows us to infer that

$$\begin{aligned}\overline{P}_{\mathscr{D}}(f) &= \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon] + f + \lambda[\varepsilon - \mathbb{I}_O]) \\ &\geq \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon]) + \underline{P}_{\mathscr{D}}(f + \lambda[\varepsilon - \mathbb{I}_O]) \geq \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon]) \\ &= \lambda(\overline{P}_{\mathscr{D}}(O) - \varepsilon) > 0.\end{aligned}$$

Since $f \in \mathscr{E}_O(\mathscr{D})$, we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f) \geq 0$ and therefore, by Corollary 19$_{85}$, that $\underline{P}_{\mathscr{D}}(f) \geq 0$. Finally, the fact that $f \notin \mathscr{D}$ implies that, for all $\alpha \geq 0$, $f - \alpha \notin \mathscr{D}$, which in turn implies that $\underline{P}_{\mathscr{D}}(f) \leq 0$. □

If we let $\mathscr{O}$ be equal to $\mathscr{O}_1(\mathscr{D}) := \{O \in \mathscr{P}_{\emptyset}(\Omega) \colon \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0\}$ or $\mathscr{O}_2(\mathscr{D}) := \{O \in \mathscr{P}_{\emptyset}(\Omega) \colon \overline{P}_{\mathscr{D}}(O) > 0\}$, or any set in between these two, we

obtain an important special instance of the operator $\mathscr{E}_{\mathscr{O}}$, which we denote by $\mathscr{E}^{\mathrm{r}}(\mathscr{D})$. For any coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$, it is defined by

$$\mathscr{E}^{\mathrm{r}}(\mathscr{D}) := \bigcup_{\substack{O \in \mathscr{P}_{\emptyset}(\Omega) \\ \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0}} \mathscr{E}_O(\mathscr{D}) = \bigcup_{\substack{O \in \mathscr{P}_{\emptyset}(\Omega) \\ \overline{P}_{\mathscr{D}}(O) > 0}} \mathscr{E}_O(\mathscr{D}).$$

The last equality is a consequence of Proposition 17_{84}, which is also the reason why $\mathscr{E}^{\mathrm{r}}(\mathscr{D}) = \mathscr{E}_{\mathscr{O}}(\mathscr{D})$ for all $\mathscr{O}_1(\mathscr{D}) \subseteq \mathscr{O} \subseteq \mathscr{O}_2(\mathscr{D})$.

Corollary 31. *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{P}_{\mathscr{D}}$. Then $\mathscr{E}^{\mathrm{r}}(\mathscr{D})$ is coherent, $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})} = \underline{P}_{\mathscr{D}}$ and*

$$\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(f\rfloor O) = \begin{cases} \underline{R}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) > 0 \\ \underline{P}_{\mathscr{D}}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) = 0 \end{cases} \quad \text{for all } (f,O) \in \mathscr{C}(\Omega).$$

Proof of Corollary 31. Since $\mathscr{E}^{\mathrm{r}}(\mathscr{D}) = \mathscr{E}_{\mathscr{O}_2(\mathscr{D})}(\mathscr{D})$, this follows immediately from Corollary 26_{93} and Proposition 28_{94}. □

If $\mathscr{D}$ is the set of desirable gambles $\mathscr{D}_{\underline{P}}$ that corresponds to a coherent lower prevision $\underline{P}$, then $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ is a subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and—as is the case for $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$—its associated conditional lower prevision $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})}(\cdot\rfloor\cdot)$ is equal to the regular extension of $\underline{P}$.

Corollary 32. *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be its regular extension. Then $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ is a coherent subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$, $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})} = \underline{P}$ and*

$$\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})}(f\rfloor O) = \underline{R}(f\rfloor O) \quad \text{for all } (f,O) \in \mathscr{C}(\Omega).$$

Proof of Corollary 32. Since $\mathscr{D}_{\underline{P}}$ is coherent and $\underline{P}_{\mathscr{D}_{\underline{P}}} = \underline{P}$, we infer from Corollary 31 that it suffices to prove that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) \subseteq \mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and that, for all $(f,O) \in \mathscr{C}(\Omega)$ such that $\overline{P}(O) = 0$, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f\rfloor O) = \underline{R}(f\rfloor O)$. The inclusion follows from Lemma $30_{\frown}$ and the fact that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{E}_{\mathscr{O}_2(\mathscr{D}_{\underline{P}})}(\mathscr{D}_{\underline{P}})$. So consider any $(f,O) \in \mathscr{C}(\Omega)$ such that $\overline{P}(O) = 0$. Let $\underline{E}(\cdot\rfloor\cdot)$ be the natural extension of $\underline{P}$. Then as explained in Section $2.7.4_{65}$, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f\rfloor O) = \underline{E}(f\rfloor O)$. Hence, since natural and regular extension coincide if the conditioning event has upper probability zero, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f\rfloor O) = \underline{R}(f\rfloor O)$. □

The following example illustrates that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ can be a strict subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$.

Example 1. *Let $\Omega = \{a,b\}$ and consider the lower prevision $\underline{P}$ that is defined by*

$$\underline{P}(f) := \min\{f(a), \frac{f(a)+f(b)}{2}\} \text{ for all } f \in \mathscr{G}(\Omega).$$

Then

$$\begin{aligned}\mathscr{D}_{\underline{P}} &= \{f \in \mathscr{G}(\Omega)\colon f(a) > 0 \text{ and } f(a)+f(b) > 0\} \cup \mathscr{G}(\Omega)_{>0} \\ &= \{f \in \mathscr{G}(\Omega)\colon f(a) \geq 0 \text{ and } f(a)+f(b) > 0\} \end{aligned} \tag{3.14}$$

and

$$\begin{aligned}\mathscr{D}^{r}_{\underline{P}} &= \mathscr{D}_{\underline{P}} \cup \{f \in \mathscr{G}(\Omega)\colon -f(b) = f(a) > 0\} \\ &= \{f \in \mathscr{G}(\Omega)\colon f(a) \geq 0 \text{ and } f(a)+f(b) \geq 0\} \setminus \{0\}.\end{aligned}$$

Since $\underline{P}(\Omega) = \overline{P}(\Omega) = 1$, $1/2 = \underline{P}(\{a\}) < \overline{P}(\{a\}) = 1$, $0 = \underline{P}(\{b\}) < \overline{P}(\{b\}) = 1/2$ *and* $\underline{P}_{\mathscr{D}_{\underline{P}}} = \underline{P}$, *we know that* $\mathscr{O}_1(\mathscr{D}_{\underline{P}})$ *consists of the singleton* $\{b\}$, *which implies that* $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}})$. *Assume* ex absurdo *that* $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}} \neq \emptyset$. *Consider any* $f \in \mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}}$. *Then by Theorem* 18_{84}, *and since* $\overline{P}_{\mathscr{D}_{\underline{P}}}(\{b\}) > 0$, *there is some* $\overline{\lambda} \in \mathbb{R}_{>0}$ *such that, for all* $\lambda \in (0, \overline{\lambda}]$, $f + \lambda(1/2 - \mathbb{I}_{\{b\}}) \in \mathscr{D}_{\underline{P}}$, *which implies, by Equation* (3.14), *that* $f(a) + \lambda 1/2 \geq 0$ *and* $f(a)+f(b) > 0$. *Since this holds for all* $\lambda \in (0, \overline{\lambda}]$, *we infer that* $f(a) \geq 0$ *and* $f(a)+f(b) > 0$, *or equivalently, that* $f \in \mathscr{D}_{\underline{P}}$. *This is a contradiction, allowing us to infer that* $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}} = \emptyset$ *and therefore, by Theorem* 18_{84}, *and since* $\overline{P}_{\mathscr{D}_{\underline{P}}}(\{b\}) > 0$, *that* $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) = \mathscr{D}_{\underline{P}}$ *and therefore also* $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{D}_{\underline{P}}$. *Hence, in conclusion:* $\mathscr{D}_{\underline{P}} = \mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) \subset \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. ◊

3.B PROOF OF THEOREM 18

Proof of Theorem 18_{84}. First, assume that the right-hand side of Equation $(3.5)_{84}$ or Equation $(3.6)_{84}$ holds. Then clearly, for all $\varepsilon \in (0,1)$, by Equation $(3.3)_{82}$, $f \in \mathscr{E}^{\varepsilon}_{O}(\mathscr{D})$. Hence, by Equation $(3.4)_{82}$, $f \in \mathscr{E}_{O}(\mathscr{D})$.

Next, assume that $f \in \mathscr{E}_{O}(\mathscr{D})$. We will prove that this implies the right-hand side of Equations $(3.5)_{84}$ and $(3.6)_{84}$. Clearly, without loss of generality, we may assume that $f \in \mathscr{E}_{O}(\mathscr{D}) \setminus \mathscr{D}$. Now fix $\varepsilon \in (0,1)$. By Equation $(3.3)_{82}$, and because $\mathscr{D}$ is coherent, we find that there are $g \in \mathscr{D} \cup \{0\}$ and $\overline{\lambda}_{\varepsilon} \in \mathbb{R}_{>0}$ such that $f = g + \overline{\lambda}_{\varepsilon}(\mathbb{I}_{O} - \varepsilon)$. If $g = 0$, then by the coherence of $\mathscr{E}_{O}(\mathscr{D})$ [because of Proposition 15_{83} and because $\overline{P}_{\mathscr{D}}(O) > 0$], $\mathbb{I}_{O} - \varepsilon \in \mathscr{E}_{O}(\mathscr{D})$. Due to Proposition 16_{84} and the coherence of $\mathscr{D}$, this implies that $f = \overline{\lambda}_{\varepsilon}(\mathbb{I}_{O} - \varepsilon) \in \mathscr{D}$, a contradiction. Hence, $g \in \mathscr{D}$ and therefore $f + \overline{\lambda}_{\varepsilon}(\varepsilon - \mathbb{I}_{O}) \in \mathscr{D}$. By repeating this argument, we obtain a set of coefficients $\overline{\lambda}_{\varepsilon} \in \mathbb{R}_{>0}$, one for every $\varepsilon \in (0,1)$, that satisfies the following condition:

$$(\forall \varepsilon \in (0,1))\ f + \overline{\lambda}_{\varepsilon}(\varepsilon - \mathbb{I}_{O}) \in \mathscr{D}. \tag{3.15}$$

This already proves Equation $(3.5)_{84}$ [simply denote $\overline{\lambda}_{\varepsilon}$ as λ].

Now fix $\varepsilon \in (0,1)$ and $\lambda \in (0, \overline{\lambda}_{\varepsilon^*}]$, with $\varepsilon^* := \varepsilon/2$. We will show that $f + \lambda(\varepsilon - \mathbb{I}_{O}) \in \operatorname{int}(\mathscr{D})$ and thereby finish the proof for Equation $(3.6)_{84}$ [simply denote $\overline{\lambda}_{\varepsilon^*}$ as $\overline{\lambda}$]. We consider two possibilities. The first possibility is

that, for all $\varepsilon' \in (0,\varepsilon^*]$, $\overline{\lambda}_{\varepsilon'}\varepsilon' \geq \lambda\varepsilon^*$. Then for all $\varepsilon'' \in (0,1)$, we can choose $\alpha > \max\{0, \max f\}$ and $\varepsilon' \in (0,\varepsilon^*]$ small enough such that $\varepsilon'(1+\alpha/\lambda\varepsilon^*) \leq \varepsilon''$ and therefore also

$$\begin{aligned}\overline{\lambda}_{\varepsilon'}(\varepsilon''-\mathbb{I}_O) \geq \overline{\lambda}_{\varepsilon'}(\varepsilon'(1+\frac{\alpha}{\lambda\varepsilon^*})-\mathbb{I}_O) \geq &\overline{\lambda}_{\varepsilon'}(\varepsilon'(1+\frac{\alpha}{\overline{\lambda}_{\varepsilon'}\varepsilon'})-\mathbb{I}_O)\\ =&\alpha+\overline{\lambda}_{\varepsilon'}(\varepsilon'-\mathbb{I}_O) > f+\overline{\lambda}_{\varepsilon'}(\varepsilon'-\mathbb{I}_O).\end{aligned}$$

By combining this with Equation $(3.15)_{\frown}$ and the coherence of $\mathscr{D}$, we find that $\varepsilon''-\mathbb{I}_O \in \mathscr{D}$. Since this holds for all $\varepsilon'' \in (0,1)$, we infer from Equation $(2.14)_{49}$ that $\overline{P}_{\mathscr{D}}(O) \leq 0$, which contradicts our assumption. Hence, we only have to consider the second, remaining possibility, namely that there is some $\varepsilon' \in (0,\varepsilon^*]$ for which $\overline{\lambda}_{\varepsilon'}\varepsilon' < \lambda\varepsilon^*$. Since also $\lambda\varepsilon^* \leq \overline{\lambda}_{\varepsilon^*}\varepsilon^*$, we can use this particular ε' to define

$$\delta_1 := \frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\lambda\varepsilon^*}{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'} \geq 0,\ \delta_2 := 1-\delta_1 = \frac{\lambda\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'}{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'} > 0$$

and

$$\begin{aligned}\beta := \delta_1\overline{\lambda}_{\varepsilon'}+\delta_2\overline{\lambda}_{\varepsilon} =&\frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\lambda\varepsilon^*}{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'}\overline{\lambda}_{\varepsilon'}+\frac{\lambda\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'}{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'}\overline{\lambda}_{\varepsilon^*}\\ =&\frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'}-\lambda\varepsilon^*\overline{\lambda}_{\varepsilon'}+\lambda\varepsilon^*\overline{\lambda}_{\varepsilon^*}-\overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*}}{\overline{\lambda}_{\varepsilon^*}\varepsilon^*-\overline{\lambda}_{\varepsilon'}\varepsilon'}.\end{aligned}$$

Since

$$\begin{aligned}\beta \geq \lambda &\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'}-\lambda\varepsilon^*\overline{\lambda}_{\varepsilon'}+\lambda\varepsilon^*\overline{\lambda}_{\varepsilon^*}-\overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*} \geq \overline{\lambda}_{\varepsilon^*}\varepsilon^*\lambda-\overline{\lambda}_{\varepsilon'}\varepsilon'\lambda\\ &\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'}-\lambda\varepsilon^*\overline{\lambda}_{\varepsilon'}-\overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*}+\overline{\lambda}_{\varepsilon'}\varepsilon'\lambda \geq 0\\ &\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^*-\lambda\varepsilon^*-\varepsilon'\overline{\lambda}_{\varepsilon^*}+\varepsilon'\lambda \geq 0 \Leftrightarrow (\overline{\lambda}_{\varepsilon^*}-\lambda)(\varepsilon^*-\varepsilon') \geq 0,\end{aligned}$$

we find that $\beta \geq \lambda$. Since we also know that

$$\begin{aligned}\delta_1\overline{\lambda}_{\varepsilon'}\varepsilon'+\delta_2\overline{\lambda}_{\varepsilon^*}\varepsilon^* &= \delta_1\overline{\lambda}_{\varepsilon'}\varepsilon'+(1-\delta_1)\overline{\lambda}_{\varepsilon^*}\varepsilon^*\\ &= \overline{\lambda}_{\varepsilon^*}\varepsilon^*+\delta_1(\overline{\lambda}_{\varepsilon'}\varepsilon'-\overline{\lambda}_{\varepsilon^*}\varepsilon^*) = \lambda\varepsilon^*,\end{aligned}$$

we find that

$$\begin{aligned}f+\lambda(\varepsilon^*-\mathbb{I}_O) =&f+\lambda\varepsilon^*-\lambda\mathbb{I}_O\\ \geq&f+\lambda\varepsilon^*-\beta\mathbb{I}_O\\ =&(\delta_1+\delta_2)f+\delta_1\overline{\lambda}_{\varepsilon'}\varepsilon'+\delta_2\overline{\lambda}_{\varepsilon^*}\varepsilon^*-(\delta_1\overline{\lambda}_{\varepsilon'}+\delta_2\overline{\lambda}_{\varepsilon})\mathbb{I}_O\\ =&\delta_1[f+\overline{\lambda}_{\varepsilon'}(\varepsilon'-\mathbb{I}_O)]+\delta_2[f+\overline{\lambda}_{\varepsilon^*}(\varepsilon^*-\mathbb{I}_O)].\end{aligned}$$

This implies that $f + \lambda(\varepsilon^* - \mathbb{I}_O) \in \mathscr{D}$ because of Equation $(3.15)_{97}$ and the coherence of $\mathscr{D}$. Hence, if we define $\delta := \lambda\varepsilon^* > 0$, then since $\lambda\varepsilon^* = \lambda\varepsilon - \delta$ [because $\varepsilon^* = \varepsilon/2$], we find that $f + \lambda(\varepsilon - \mathbb{I}_O) - \delta \in \mathscr{D}$ and therefore also, by Equation $(2.7)_{41}$, that $f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathrm{int}(\mathscr{D})$. $\square$

4

MULTIVARIATE MODELS

"We neither fear complexity nor embrace it for its own sake, but rather face it with the faith that simplicity and understanding are within reach."

Frederick R. Adler

The frameworks introduced in Chapter 2_{37} were all concerned with modelling the uncertainty that is related to a single variable X. In contrast, practical modelling tasks are usually concerned with multiple variables at once. Consider for example a situation where we want to model the weather. In that case, instead of modelling 'the weather' directly, we would usually separate it into multiple more manageable variables such as rainfall, wind speed, barometric pressure, and so on. Not only do we want to model the uncertainty that is associated with each of these variables individually, we also wish the capture the connection between them and how they influence each other.

Fortunately, from a theoretical point of view, these multivariate set-ups are identical to the univariate case that we considered before, and all the machinery we have introduced so far keeps on working. In fact, as we will see, the multivariate case even is a special case of the univariate one, and it comes with its own additional toolbox, including concepts such as marginalisation, structural assessments, and so on. The goal of this chapter is to introduce the multivariate framework, to link it with the univariate one, and to discuss the additional tools that become available. With respect to structural assessments, we will mainly focus on epistemic irrelevance, an asymmetric notion of independence, because it is one of the cornerstones of the theory of credal networks

under epistemic irrelevance that will be the topic of the following chapters. As we will see, and in contrast with the precise-probabilistic case, independence becomes a highly non-trivial concept.

The underlying ideas behind many of the things that we do in this chapter are not new. Our contribution mainly consists in presenting, justifying, unifying, combining and extending existing ideas in the literature [19, 45, 76] into a full-fledged and consistent account.

4.1 REDUCING EVERYTHING TO THE UNIVARIATE CASE

Consider a finite number of individual variables X_s, indexed by the elements s of some finite set G. For every $s \in G$, the variable X_s assumes values in some non-empty finite state space $\mathscr{X}_s$. In the aforementioned weather example, we could take G to be equal to $\{r, w, b\}$, using r, w and b as respective shorthand notations for rainfall, wind speed and barometric pressure, respectively. A simple choice of associated state spaces could be $\mathscr{X}_r = \{\text{yes}, \text{no}\}$, $\mathscr{X}_w = \{\text{light}, \text{strong}, \text{storm}\}$ and $\mathscr{X}_b = \{\text{high}, \text{low}\}$.

For any subset S of G, we now let X_S be the tuple whose components are the variables X_s, with $s \in S$. This new joint variable X_S assumes values x_S in the finite Cartesian product set $\mathscr{X}_S := \times_{s \in S} \mathscr{X}_s$. For any such tuple-valued element x_S of $\mathscr{X}_S$, and any $s \in S$, the s-component of x_S is an element of $\mathscr{X}_s$, and we will denote it by x_s. If $S = \emptyset$, $\mathscr{X}_\emptyset$ is taken to be a singleton—the empty map. $X_\emptyset$ can then only assume this single value, so there is no uncertainty about it. For $S = \{s\}$, X_S can be identified with X_s. The case $S = G$ is the most important one, because it provides us with a single tuple-valued variable X_G that represents all the individual variables we are interested in.

Modelling the uncertainty that is associated with the collection of variables X_s, $s \in G$, is now extremely easy. It suffices to apply the frameworks introduced in Chapter 2_{37} to the single tuple-valued variable X_G, simply by choosing $X := X_G$ and $\Omega := \mathscr{X}_G$. Depending on the chosen framework, one can use a coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$, a coherent conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, a set $\mathscr{M}_G$ of linear conditional previsions $P_G(\cdot\rfloor\cdot)$ on $\mathscr{G}(\mathscr{X}_G)$, a set $\mathscr{F}_G$ of full conditional probability mass functions $p_G(\cdot\rfloor\cdot)$ on $\mathscr{C}_*(\mathscr{X}_G)$, or any of their unconditional versions. Whenever we need to distinguish them from other—local—models that will be introduced further on, we will refer to these uncertainty models for X_G as *global* or *joint* models. The theory [see Chapter 2_{37}] remains exactly the same and our results for conditioning and updating [see Chapter 3_{69}] are directly applicable.

4.2 MARGINALISATION, CONDITIONING AND UPDATING

One of the major advantages of the multivariate set-up is that it corresponds to the use of a state space $\Omega = \times_{s \in G} \mathscr{X}_s$ that is highly structured. This structure

allows for the introduction of some new concepts—such as marginalisation—as well as special cases of existing ones—such as conditioning and updating. This is all well known and relatively easy in terms of probabilities, but for some of the other frameworks that we consider, clear definitions are often missing. We present an overview of existing tools, connect ideas from different frameworks to one another and add new material whenever necessary. The result is a complete and well-founded theory of multivariate imprecise probabilities that is, to the best of our knowledge, novel in its completeness, generality and consistency.

4.2.1 *Marginalisation*

In practice, we may not always wish to consider all the variables that are represented by X_G. Instead, we may want to focus on a specific subset of variables X_s only, indexed by the elements of some subset S of G. As an important special case, we might want to focus on just a single variable X_s—choosing $S = \{s\}$. Of course, we could simply construct a separate model for the uncertainty that is associated with X_S. However, in doing so, we would be ignoring the fact that this information is already contained within the uncertainty model for X_G. In the precise-probabilistic case, as we all know, the model for X_S is related to that for X_G by means of marginalisation. As we are about to show, this concept generalises easily to the imprecise-probabilistic case.

Regardless of the framework we adopt, the uncertainty model for X_S will be expressed in terms of gambles on $\mathscr{X}_S$, either by assessing their desirability directly—sets of desirable gambles—or by providing supremum buying and infimum selling prices for them—lower and upper previsions. Any such gamble $f \in \mathscr{G}(\mathscr{X}_S)$ is a map on $\mathscr{X}_S$ that associates a real number $f(x_S)$ with every $x_S \in \mathscr{X}_S$. For any $S \subseteq U \subseteq G$, we now let f_U be the *cylindrical extension* of f to $\mathscr{X}_U$, defined by $f_U(x_U) := f(x_S)$ for all $x_U \in \mathscr{X}_U$, letting x_S be the projection of x_U on $\mathscr{X}_S$. Although f_U is formally a function of x_U—the values of the variables X_s that are indexed by an element of U—it only depends on x_S—the values of the variables that are indexed by an element of S—and remains constant if $x_{U\setminus S}$ is varied within $\mathscr{X}_{U\setminus S}$. Hence, conceptually, f_U is indistinguishable from f, which allows us to identify them with one another and simply denote them both by f. Using this convention, for any $\mathscr{K} \subseteq \mathscr{G}(\mathscr{X}_G)$, we can write $\mathscr{K} \cap \mathscr{G}(\mathscr{X}_S)$ to refer to the gambles in $\mathscr{G}(\mathscr{X}_S)$ whose cylindrical extension belongs to $\mathscr{K}$, or equivalently, those gambles in $\mathscr{K}$ that only depend on the variables X_s, $s \in S$.

In terms of sets of desirable gambles, marginalisation is now based on the intuitive idea that a gamble $f \in \mathscr{G}(\mathscr{X}_S)$ should be desirable if and only if its cylindrical extension is desirable. This leads to the following definition for the marginal model for X_S [45]:

$$\operatorname{marg}_S(\mathscr{D}_G) := \left\{ f \in \mathscr{G}(\mathscr{X}_S) \colon f \in \mathscr{D}_G \right\} = \mathscr{D}_G \cap \mathscr{G}(\mathscr{X}_S). \tag{4.1}$$

Loosely speaking, the marginal model $\mathrm{marg}_S(\mathscr{D}_G)$ consists of the gambles in the global model $\mathscr{D}_G$ that only depend on the variable X_S. It is a matter of straightforward verification to see that marginalisation preserves coherence: if $\mathscr{D}_G$ is coherent, $\mathrm{marg}_S(\mathscr{D}_G)$ will also be coherent. See Reference [45] for additional properties of this marginalisation operator.

Similarly, in terms of lower previsions, we adopt the intuitive idea that the lower prevision of a gamble $f \in \mathscr{G}(\mathscr{X}_S)$ should be identical to that of its cylindrical extension. Starting from a lower prevision $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$, this leads us to consider the following expression for the corresponding marginal model $\underline{P}_S$:

$$\underline{P}_S(f) := \underline{P}_G(f) \text{ for all } f \in \mathscr{G}(\mathscr{X}_S). \tag{4.2}$$

Coherence is again preserved: if $\underline{P}_G$ is coherent, $\underline{P}_S$ will be coherent as well. For coherent lower previsions, Definition (4.2) can also be regarded as a consequence of Definition $(4.1)_{\frown}$: if $\underline{P}_G$ is related to a coherent set of desirable gambles $\mathscr{D}_G$ by means of Equation $(2.3)_{40}$, then $\underline{P}_S$ and $\mathrm{marg}_S(\mathscr{D}_G)$ will be related in the same way.

Since linear previsions are a special type of lower previsions, they can be marginalised in the same way. In this special case, if we let p_G be the probability mass function for X_G that corresponds to a linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$, then

$$\begin{aligned} P_S(f) := P_G(f) &= \sum_{x_G \in \mathscr{X}_G} f(x_G) p_G(x_G) \\ &= \sum_{x_S \in \mathscr{X}_S} \sum_{x_{G\setminus S} \in \mathscr{X}_{G\setminus S}} f(x_S, x_{G\setminus S}) p_G(x_S, x_{G\setminus S}) \\ &= \sum_{x_S \in \mathscr{X}_S} f(x_S) \sum_{x_{G\setminus S} \in \mathscr{X}_{G\setminus S}} p_G(x_S, x_{G\setminus S}) = \sum_{x_S \in \mathscr{X}_S} f(x_S) p_S(x_S), \end{aligned}$$

letting p_S be the marginalised probability mass function for the variable X_S, as derived from p_G by summing out the variable $X_{G\setminus S}$. In other words: in the case of linear previsions, our notion of marginalisation is equivalent to the usual precise-probabilistic one.

A set $\mathscr{M}_G$ of linear previsions on $\mathscr{G}(\mathscr{X}_G)$ or a set $\mathscr{F}_G$ of probability mass functions on $\mathscr{X}_G$ can be marginalised by marginalising each of its elements separately, leading to the following expressions for the corresponding marginal models:

$$\mathrm{marg}_S(\mathscr{M}_G) := \{P_S \colon P_G \in \mathscr{M}_G\}$$

and

$$\mathrm{marg}_S(\mathscr{F}_G) := \{p_S \colon p_G \in \mathscr{F}_G\}.$$

When applied to such sets, the operator $\mathrm{marg}_S(\cdot)$ preserves closedness and convexity: if $\mathscr{M}_G$ is closed and/or convex, so too is $\mathrm{marg}_S(\mathscr{M}_G)$, and similarly for $\mathscr{F}_G$ and $\mathrm{marg}_S(\mathscr{F}_G)$.[1]

Although we prefer to interpret—and motivate the use of—Equation $(4.2)_{\frown}$ directly in terms of supremum buying prices for gambles, $\underline{P}_S$ can also be given an alternative interpretation, which is especially intuitive if the sensitivity analysis interpretation is adopted. If $\underline{P}_G$ is coherent, it follows from the lower envelope theorem [Theorem 7_{52}] that

$$\underline{P}_S(f) = \min\{P_S(f) \colon P_S \in \mathrm{marg}_S(\mathscr{M}_{\underline{P}_G})\} \text{ for all } f \in \mathscr{G}(\mathscr{X}_S). \tag{4.3}$$

Since $\underline{P}_G$ and therefore also $\underline{P}_S$ are both coherent, we know from Section $2.6.1_{52}$ that $\mathscr{M}_{\underline{P}_G}$ and $\mathscr{M}_{\underline{P}_S}$ are the unique closed and convex sets of linear previsions of which $\underline{P}_G$ and $\underline{P}_S$ are the respective lower envelopes. By combining this with Equation 4.3 and the fact that $\mathrm{marg}_S(\cdot)$ preserves convexity and closedness, we find that $\mathscr{M}_{\underline{P}_S} = \mathrm{marg}_S(\mathscr{M}_{\underline{P}_G})$. Consequently, by the one-to-one correspondence between linear previsions and probability mass functions, we also have that $\mathscr{F}_{\underline{P}_S} = \mathrm{marg}_S(\mathscr{F}_{\underline{P}_G})$.

4.2.2 *A special type of conditional models*

Given that the multivariate set-up is just a special case of the univariate one, we could in theory consider conditional models with respect to any chosen conditioning event $B \in \mathscr{P}_\emptyset(\mathscr{X}_G)$. However, in the multivariate case, some conditioning events are more fundamental than others, as they lead to conditional models that are particularly intuitive.

The most fundamental type of events are those that correspond to fixing the value of some subset of the variables X_s, $s \in G$, or in other words, fixing the value x_I of X_I, for some subset I of G. With any such $x_I \in \mathscr{X}_I$, we can associate a corresponding event $\{x_I\} \times \mathscr{X}_{G\setminus I} := \{z_G \in \mathscr{X}_G \colon z_I = x_I\} \in \mathscr{P}_\emptyset(\mathscr{X}_G)$ that consists of the tuples z_G in $\mathscr{X}_G$ for which, for all $s \in I$, the s-component of z_G is equal to that of x_I. The indicator $\mathbb{I}_{\{x_I\}\times\mathscr{X}_{G\setminus I}}$ of $\{x_I\} \times \mathscr{X}_{G\setminus I}$ clearly only depends on the variable X_I. In fact, it is the cylindrical extension of the indicator $\mathbb{I}_{\{x_I\}} \in \mathscr{G}(\mathscr{X}_I)$, which allows us to identify these two indicators and simply denote them both by $\mathbb{I}_{\{x_I\}}$. In the same spirit, the events $\{x_I\} \times \mathscr{X}_{G\setminus I}$ and $\{x_I\}$ can also be identified with each other, and we will use $\{x_I\}$ to refer to them both or, whenever it is clear that we are referring to a set, simply denote them both as x_I.

[1] Convexity is trivially preserved. For sets of linear previsions, closedness is preserved because, with respect to the topology of pointwise convergence, restricting the domain of a linear prevision P_G—which is what marginalisation essentially does—is a continuous operation. For sets of probability mass functions, closedness is preserved because, with respect to the topology that is induced by the Euclidean metric, summing components together—which is what marginalisation essentially does—is a continuous operation.

If we start from a set of desirable gambles $\mathscr{D}_G$ on $\mathscr{G}(\mathscr{X}_G)$ and condition it on such an event $\{x_I\}$, then according to Equation $(2.2)_{40}$, the resulting conditional model[2,3]

$$\mathscr{D}_G\rfloor x_I = \big\{f \in \mathscr{G}(\{x_I\} \times \mathscr{X}_{G\setminus I})\colon \mathbb{I}_{\{x_I\}} f \in \mathscr{D}_G\big\}.$$

is a subset of $\mathscr{G}(\{x_I\} \times \mathscr{X}_{G\setminus I})$. However, in practice, the gambles in $\mathscr{D}_G\rfloor x_I$ only depend on the value of $X_{G\setminus I}$, because the value of X_I is equal to x_I and therefore fixed. It is therefore far more intuitive to identify these gambles with gambles on $\mathscr{X}_{G\setminus I}$. By the obvious one-to-one correspondence between $\{x_I\} \times \mathscr{X}_{G\setminus I}$ and $\mathscr{X}_{G\setminus I}$, such an identification is trivial. The resulting conditional model is[4]

$$\operatorname{marg}_{G\setminus I}(\mathscr{D}_G\rfloor x_I) := \big\{f \in \mathscr{G}(\mathscr{X}_{G\setminus I})\colon \mathbb{I}_{\{x_I\}} f \in \mathscr{D}_G\big\}. \tag{4.4}$$

It represents our subject's beliefs about the values of the variables that are represented by $X_{G\setminus I}$, contingent on the fact that X_I assumes the value x_I. Coherence is again trivially preserved. Also, since $\mathbb{I}_{\{x_\emptyset\}} = 1$, the degenerate case $I = \emptyset$ yields no problems. As is to be expected from conditioning on the value of a deterministic variable such as $X_\emptyset$, it amounts to not conditioning at all.

Similar definitions can be given for conditional lower previsions, for (sets of) conditional linear previsions, and for (sets of) full conditional probability mass functions, by using a similar identification. However, since we are about to introduce a generalised version of Equation (4.4), we will—in order to avoid having to repeat ourselves—defer the translations to these other frameworks to the next section, where we will introduce them directly for the general version that is considered there, rather than the specific case that was considered here.

4.2.3 *Marginalised conditional models*

The concepts that were introduced in Sections $4.2.1_{102}$ and $4.2.2_{\frown}$ can be combined with each other to obtain a single generalised operator that includes them both as special cases. For any two disjoint subsets S and I of G, the idea is to

[2]This is a nice example of an instance in where it is safe—as well as far more elegant—to use x_I as a shorthand notation for the set $\{x_I\} = \{x_I\} \times \mathscr{X}_{G\setminus I}$, because it is clear from the context that we are referring to a set.

[3]In Reference [45], the authors used $\mathscr{D}_G\rfloor x_I$ to refer to a different set of desirable gambles: the set $\operatorname{marg}_{G\setminus I}(\mathscr{D}_G\rfloor x_I)$, as given by Equation (4.4). We prefer to distinguish between these two sets because it allows us to establish a clear connection with the general—non-multivariate—notion of conditioning that was discussed in Chapter 2_{37}.

[4]For those among you who wonder why the marginalisation operator is part of our notation: we will generalise this operator in the next section, in such a way that, in this particular case, it maps gambles on $\{x_I\} \times \mathscr{X}_{G\setminus I}$ to their counterpart on $\mathscr{X}_{G\setminus I}$. However, it would be overkill to formally introduce this generalised version of the marginalisation operator already here, because in this case, it corresponds to a rather degenerate notion of marginalisation.

construct a model that represents our subject's beliefs about the value of the variable X_S contingent on the occurrence of an event $B_S \times B_I \in \mathscr{P}_\emptyset(\mathscr{X}_{S\cup I})$, for some $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$—contingent on the fact that X_S assumes a value in B_S and X_I assumes a value in B_I. This model should reflect our subject's beliefs about which value X_S will take in the remaining set of possibilities B_S. Let us start by explaining how this works in terms of sets of desirable gambles.

The first step consists in generalising the concept of conditioning on an event $\{x_I\}$ to conditioning on an event $B_S \times B_I$, with $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$. As we did in the special case where $B_S = \mathscr{X}_S$ and $B_I = \{x_I\}$ [see Section 4.2.2$_{104}$] we can identify the event $B_S \times B_I$ with the event $B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)} \in \mathscr{P}_\emptyset(\mathscr{X}_G)$, and similarly for their indicators. Using this convention, we also have that $\mathbb{I}_{B_S\times B_I} = \mathbb{I}_{B_S}\mathbb{I}_{B_I}$. Conditioning on the event $B_S \times B_I$ is now again a matter of applying Equation (2.2)$_{40}$, which results in a conditional model

$$\mathscr{D}_G \rfloor B_S \times B_I = \left\{ f \in \mathscr{G}(B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)}) \colon \mathbb{I}_{B_S}\mathbb{I}_{B_I} f \in \mathscr{D}_G \right\} \tag{4.5}$$

that consists of gambles on $B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)}$. However, unlike what happened in the special case where $B_S = \mathscr{X}_S$ and $B_I = \{x_I\}$, there is no immediate one-to-one correspondence with gambles on $B_S \times \mathscr{X}_{G\setminus(S\cup I)}$, and therefore definitely not with gambles on B_S. Consequently, we do not trivially obtain a set of desirable gambles on B_S.

The second step therefore consists in generalising the concept of marginalisation. The central idea of marginalisation was to restrict attention to those gambles in $\mathscr{D}_G$ that, although they are formally defined on $\mathscr{X}_G$, depend on the value of X_S only. The very same trick can also be applied to a set $\mathscr{D}'_G$ of gambles on $B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)}$. Formally, we identify a gamble $f \in \mathscr{G}(B_S)$ with its *cylindrical extension*[5] $f_G \in \mathscr{G}(B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)})$, defined by

$$f_G(x_G) := f(x_S) \text{ for all } x_G \in B_S \times B_I \times \mathscr{X}_{G\setminus(S\cup I)},$$

and denote them both by f. Using this identification, we can then define

$$\operatorname{marg}_S(\mathscr{D}'_G) := \mathscr{D}'_G \cap \mathscr{G}(B_S) = \left\{ f \in \mathscr{G}(B_S) \colon f \in \mathscr{D}'_G \right\}.$$

Loosely speaking, the elements of $\operatorname{marg}_S(\mathscr{D}'_G)$ are the gambles in $\mathscr{D}'_G$ that depend on the value of X_S only, meaning that they remain constant if the value of $X_{G\setminus S}$ varies within its remaining set of options $B_I \times \mathscr{X}_{G\setminus(S\cup I)}$. It is easy to check that this generalised notion of marginalisation preserves coherence: if $\mathscr{D}'_G$ is a coherent set of desirable gambles on $B_S \times B_I \times \mathscr{X}_{G\setminus I}$, then $\operatorname{marg}_S(\mathscr{D}'_G)$ is a coherent set of desirable gambles on B_S.

[5]This is a more general notion of cylindrical extension than the one that was considered in Section 4.2.1$_{102}$, which restricted attention to the special case where $B_S = \mathscr{X}_S$ and $B_I = \mathscr{X}_I$.

By sequentially applying the two steps that were discussed above, we obtain the following marginalised conditional model, which represents our subject's beliefs about the value of the variable X_S, contingent on the fact that X_S assumes a value in B_S and X_I assumes a value in B_I:

$$\begin{aligned}\mathrm{marg}_S(\mathscr{D}_G\rfloor B_S\times B_I) &:= (\mathscr{D}_G\rfloor B_S\times B_I)\cap\mathscr{G}(B_S)\\ &=\big\{f\in\mathscr{G}(B_S)\colon \mathbb{I}_{B_S}\mathbb{I}_{B_I}f\in\mathscr{D}_G\big\}.\end{aligned}\tag{4.6}$$

The operator $\mathrm{marg}_S(\cdot\rfloor B_S\times B_I)$ is guaranteed to preserve coherence, simply because this is the case for each of the two individual operators $\cdot\rfloor B_S\times B_I$ and $\mathrm{marg}_S(\cdot)$.

If $B_S=\mathscr{X}_S$, we use $\mathscr{D}_G\rfloor B_I$ as a shorthand notation for $\mathscr{D}_G\rfloor\mathscr{X}_S\times B_I$,[6] and similarly for $\mathscr{D}_G\rfloor B_S:=\mathscr{D}_G\rfloor B_S\times\mathscr{X}_I$. If $B_S=\mathscr{X}_S$ and $B_I=\mathscr{X}_I$, we find that $\mathscr{D}_G=\mathscr{D}_G\rfloor\mathscr{X}_S\times\mathscr{X}_I$. In this way, we see that Equations $(4.1)_{102}$ and $(4.4)_{105}$ correspond to particular cases of $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_S\times B_I)$. The former corresponds to letting $B_S=\mathscr{X}_S$ and $B_I=\mathscr{X}_I$ and the latter can be obtained by choosing $S=G\setminus I$, $B_S=\mathscr{X}_S$ and $B_I=\{x_I\}$.

A similar story can be told in terms of lower previsions. If we start from a conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, then $\underline{P}_G(\cdot\rfloor B_S\times B_I)$ is a lower prevision on $\mathscr{G}(B_S\times B_I\times\mathscr{X}_{G\setminus(S\cup I)})$. If all we are interested in is the conditional lower prevision of gambles that only depend on the value that X_S takes in its remaining set of options B_S, we can simply restrict the domain of $\underline{P}_G(\cdot\rfloor B_S\times B_I)$ to (cylindrical extensions of) gambles in $\mathscr{G}(B_S)$. The resulting operator is denoted by $\underline{P}_S(\cdot\rfloor B_S\times B_I)$ and is trivially defined as

$$\underline{P}_S(f\rfloor B_S\times B_I):=\underline{P}_G(f\rfloor B_S\times B_I)\text{ for all }f\in\mathscr{G}(B_S).\tag{4.7}$$

Coherence is again preserved and—as was the case for marginalisation—this definition can also be regarded as an immediate consequence of its counterpart for desirable gambles: if $\underline{P}_G(\cdot\rfloor\cdot)$ is obtained from a coherent set of desirable gamble $\mathscr{D}_G$—by means of Equation $(2.10)_{42}$—then $\underline{P}_S(\cdot\rfloor B_S\times B_I)$ is the coherent lower prevision that corresponds to $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_S\times B_I)$—as given by Equation $(2.3)_{40}$.

Linear previsions correspond to a special case. Furthermore, in that case, if we let $p_G(\cdot\rfloor\cdot)$ be the unique full conditional probability mass function that is associated with a linear conditional prevision $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, then for all $f\in\mathscr{G}(B_S)$:

[6] In previous work [31, Section 3.3], $\mathscr{D}_G\rfloor B_I$ was used to refer to a different set of desirable gambles: the set $\mathrm{marg}_{G\setminus I}(\mathscr{D}_G\rfloor B_I)$, as given by Equation (4.6). We prefer the current notation because it allows for a clear connection with the general approach to conditioning that was discussed in Chapter 2_{37}; see Footnote 3_{105} as well.

$$
\begin{aligned}
P_S(f\rfloor B_S\times B_I) &:= P_G(f\rfloor B_S\times B_I) && (4.8)\\
&= \sum_{x_G\in B_S\times B_I\times\mathscr{X}_{G\setminus(S\cup I)}} f(x_G)p_G(x_G\rfloor B_S\times B_I)\\
&= \sum_{x_S\in B_S}\ \sum_{x_{G\setminus S}\in B_I\times\mathscr{X}_{G\setminus(I\cup S)}} f(x_S,x_{G\setminus S})p_G(x_S,x_{G\setminus S}\rfloor B_S\times B_I)\\
&= \sum_{x_S\in B_S} f(x_S)\sum_{x_{G\setminus S}\in B_I\times\mathscr{X}_{G\setminus(I\cup S)}} p_G(x_S,x_{G\setminus S}\rfloor B_S\times B_I)\\
&= \sum_{x_S\in B_S} f(x_S)p_S(x_S\rfloor B_S\times B_I), && (4.9)
\end{aligned}
$$

where $p_S(x_S\rfloor B_S\times B_I)$ is the probability that X_S has the value x_S, conditional on $B_S\times B_I$, as obtained in the conventional way, by summing out the remaining possible values for $X_{G\setminus S}$. Hence, we see that—as was the case for marginalisation—the usual precise-probabilistic concept is again obtained as a special case.

These combinations of conditioning and marginalisation can be introduced for sets of conditional linear previsions and sets of full conditional probability mass functions as well, simply by applying the above procedures elementwise. For a set $\mathscr{M}_G$ of conditional linear previsions $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, we define

$$
\begin{aligned}
\mathrm{marg}_S(\mathscr{M}_G\rfloor B_S\times B_I) &:= \{P_S(\cdot\rfloor B_S\times B_I)\colon P_G(\cdot\rfloor B_S\times B_I)\in\mathscr{M}_G\rfloor B_S\times B_I\}\\
&= \{P_S(\cdot\rfloor B_S\times B_I)\colon P_G(\cdot\rfloor\cdot)\in\mathscr{M}_G\},
\end{aligned}
$$

and for a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\mathscr{X}_G)$, we let

$$
\begin{aligned}
\mathrm{marg}_S(\mathscr{F}_G\rfloor B_S\times B_I) &:= \{p_S(\cdot\rfloor B_S\times B_I)\colon p_G(\cdot\rfloor B_S\times B_I)\in\mathscr{F}_G\rfloor B_S\times B_I\}\\
&= \{p_S(\cdot\rfloor B_S\times B_I)\colon p_G(\cdot\rfloor\cdot)\in\mathscr{F}_G\}.
\end{aligned}
$$

The operator $\mathrm{marg}_S(\cdot)$ will again preserve closedness and convexity:[7] if $\mathscr{M}_G\rfloor B_S\times B_I$ is closed and/or convex, then $\mathrm{marg}_S(\mathscr{M}_G\rfloor B_S\times B_I)$ will be closed and/or convex as well, and similarly for $\mathscr{F}_G\rfloor B_S\times B_I$ and $\mathrm{marg}_S(\mathscr{F}_G\rfloor B_S\times B_I)$.

Finally, Equation $(4.7)_{\frown}$ can also be given a sensitivity analysis interpretation. If $\underline{P}_G(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision on $\mathscr{C}(\mathscr{X}_G)$, it follows from the lower envelope theorem [Theorem 7_{52}] that, for all $f\in\mathscr{G}(B_S)$:

$$
\begin{aligned}
&\underline{P}_S(f\rfloor B_S\times B_I)\\
&\quad= \min\{P_S(f\rfloor B_S\times B_I)\colon P_G(\cdot\rfloor\cdot)\in\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\}\\
&\quad= \min\{P_S(f\rfloor B_S\times B_I)\colon P_S(\cdot\rfloor B_S\times B_I)\in\mathrm{marg}_S(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor B_S\times B_I)\}. \quad (4.10)
\end{aligned}
$$

[7] This follows from an argument that is analogous to the one in Footnote 1_{104}.

Since $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ is a coherent lower prevision, we know from Section 2.6.1$_{52}$ that $\mathscr{M}_{\underline{P}_S(\cdot\rfloor B_S\times B_I)}$ is the unique closed and convex set of linear previsions that has $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ as its lower envelope. However, since $\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor B_S \times B_I = \mathscr{M}_{\underline{P}_G(\cdot\rfloor B_S\times B_I)}$ [see Proposition 8$_{53}$] and since $\mathscr{M}_{\underline{P}_G(\cdot\rfloor B_S\times B_I)}$ is a closed and convex set of linear previsions [see Section 2.6.1$_{52}$] $\mathrm{marg}_S(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor B_S \times B_I)$ is also a closed and convex set of linear previsions that, by Equation (4.10)$_{\cap}$, has $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ as its lower envelope. Hence, we may conclude that

$$\mathscr{M}_{\underline{P}_S(\cdot\rfloor B_S\times B_I)} = \mathrm{marg}_S(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor B_S \times B_I) \tag{4.11}$$

Due to the one-to-one correspondence between conditional linear previsions and full conditional probability mass functions, this also implies that $\mathscr{F}_{\underline{P}_S(\cdot\rfloor B_S\times B_I)} = \mathrm{marg}_S(\mathscr{F}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor B_S \times B_I)$.

4.2.4 *Keeping all the conditional information together*

The models that we have introduced in the previous section are fairly general, as they consider two conditioning events: B_S and B_I. In many practical situations, the event B_S will be trivial, in the sense that it is equal to $\mathscr{X}_S$. In that case

$$\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I) = \{f \in \mathscr{G}(\mathscr{X}_S)\colon \mathbb{I}_{B_I} f \in \mathscr{D}_G\}$$

represents our subject's beliefs about X_S, contingent on partial information about X_I—the fact that X_I takes a value in B_I. It is important to realise that $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)$ truly reflects all of our subject's beliefs about X_S, including conditional beliefs: for any $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$, we can obtain the corresponding conditional model $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)\rfloor B_S$ by applying Equation (2.2)$_{40}$. We find that

$$\begin{aligned}\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)\rfloor B_S &= \{f \in \mathscr{G}(B_S)\colon \mathbb{I}_{B_S} f \in \mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)\}\\ &= \{f \in \mathscr{G}(B_S)\colon \mathbb{I}_{B_I}[\mathbb{I}_{B_S} f] \in \mathscr{D}_G\}\\ &= \mathrm{marg}_S(\mathscr{D}_G\rfloor B_S \times B_I).\end{aligned} \tag{4.12}$$

In other words: $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)$ can be regarded as a convenient representation for all the models $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_S \times B_I)$, with $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$.

This is no longer true if we move from the framework of sets of desirable gambles to that of coherent lower previsions. Consider a coherent conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ and an event $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$. Then for any $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$, we can derive $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ from $\underline{P}_G(\cdot\rfloor\cdot)$ by applying Equation (4.7)$_{107}$. However, it may be impossible to derive $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ from $\underline{P}_S(\cdot\rfloor B_I)$—unless $\underline{P}_S(\mathbb{I}_{B_S}\rfloor B_I) > 0$. This means that, in contradistinction with what we have found for $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)$, $\underline{P}_S(\cdot\rfloor B_I)$ does not truly represent all of our subject's beliefs about X_S contingent on B_I, but only his 'unconditional' beliefs. In order to represent all of our subject's beliefs about X_S contingent on B_I

by a single operator, we introduce the conditional lower prevision $\underline{P}_S(\cdot\rfloor\cdot\times B_I)$ on $\mathscr{C}(\mathscr{X}_S)$, defined by

$$\underline{P}_S(f\rfloor B_S\times B_I) := \underline{P}_G(f\rfloor B_S\times B_I) \text{ for all } (f,B_S)\in\mathscr{C}(\mathscr{X}_S). \tag{4.13}$$

For a fixed event $B_S\in\mathscr{P}_\emptyset(\mathscr{X}_S)$, $\underline{P}_S(\cdot\rfloor B_S\times B_I)$ is a lower prevision on $\mathscr{G}(B_S)$ that is equal to the identically denoted operator that was defined in the previous section [see Equation (4.7)$_{107}$]. For $B_I=\mathscr{X}_I$, we let $\underline{P}_S(\cdot\rfloor\cdot) := \underline{P}_S(\cdot\rfloor\cdot\times\mathscr{X}_I)$. It is a matter of straightforward verification that if $\underline{P}_G(\cdot\rfloor\cdot)$ is derived from a coherent set of desirable gambles $\mathscr{D}_G$, then $\underline{P}_S(\cdot\rfloor\cdot\times B_I)$ is the coherent conditional lower prevision on $\mathscr{C}(\mathscr{X}_S)$ that corresponds to $\mathrm{marg}_S(\mathscr{D}_G\rfloor B_I)$. As an immediate consequence, we find that Equation (4.13) preserves coherence: if $\underline{P}_G(\cdot\rfloor\cdot)$ is coherent, then $\underline{P}_S(\cdot\rfloor\cdot\times B_I)$ is also coherent.

A similar concept can be introduced for a set $\mathscr{M}_G$ of conditional linear previsions on $\mathscr{C}(\Omega)$. By applying Equation (4.13) to every element of $\mathscr{M}_G$, we obtain a set

$$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I) := \{P_S(\cdot\rfloor\cdot\times B_I)\colon P_G(\cdot\rfloor\cdot)\in\mathscr{M}_G\}$$

of conditional linear previsions on $\mathscr{C}(\mathscr{X}_S)$. For all $B_S\in\mathscr{P}_\emptyset(\mathscr{X}_S)$, we have that

$$\begin{aligned}\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I)\rfloor B_S =&\{P_S(\cdot\rfloor B_S\times B_I)\colon P_S(\cdot\rfloor\cdot\times B_I)\in\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I)\}\\ =&\{P_S(\cdot\rfloor B_S\times B_I)\colon P_G(\cdot\rfloor\cdot)\in\mathscr{M}_G\}\\ =&\,\mathrm{marg}_S(\mathscr{M}_G\rfloor B_S\times B_I).\end{aligned} \tag{4.14}$$

We add the superscript c to indicate that the set $\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I)$ is more than just a convenient representation of the sets $\mathrm{marg}_S(\mathscr{M}_G\rfloor B_S\times B_I)$, for $B_S\in\mathscr{P}_\emptyset(\mathscr{X}_S)$. The set $\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I)$ also tells us, for a given combination of elements of $\mathrm{marg}_S(\mathscr{M}_G\rfloor B_S\times B_I)$, one for every $B_S\in\mathscr{P}_\emptyset(\mathscr{X}_S)$, whether or not there is some conditional linear prevision $P(\cdot\rfloor\cdot)\in\mathscr{M}_G$ from which they can all be derived by means of Equation (4.8)$_{108}$. This is a major difference with $\underline{P}_S(\cdot\rfloor\cdot\times B_I)$, which only serves as a convenient representation for the lower previsions $\underline{P}_S(\cdot\rfloor B_S\times B_I)$, for $B_S\in\mathscr{P}_\emptyset(\mathscr{X}_S)$. Nevertheless, rather surprisingly, we can still establish the following connection.

Theorem 33. *Consider a coherent lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\Omega)$. Let S and I be two disjoint subsets of G. Then for all $B_I\in\mathscr{P}_\emptyset(\mathscr{X}_I)$, it holds that* $\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor|B_I)=\mathscr{M}_{\underline{P}_S(\cdot\rfloor\cdot\times B_I)}$.

Proof of Theorem 33. First consider any $P^*_S(\cdot\rfloor\cdot)\in\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}\rfloor|B_I)$. This means that there is some $P_G(\cdot\rfloor\cdot)\in\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}$ such that $P^*_S(\cdot\rfloor\cdot)=P_S(\cdot\rfloor\cdot\times B_I)$. Hence, for all $(f,B_S)\in\mathscr{C}(\mathscr{X}_S)$, we find that

$$P^*_S(f\rfloor B_S)=P_S(f\rfloor B_S\times B_I)\geq\underline{P}_S(f\rfloor B_S\times B_I),$$

which implies that $P^*_S(\cdot\rfloor\cdot)\in\mathscr{M}_{\underline{P}_S(\cdot\rfloor\cdot\times B_I)}$.

Next, consider any $P_S^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}_S(\cdot\rfloor\cdot\times B_I)}$. Let $\underline{P}_G^*(\cdot\rfloor\cdot)$ be the conditional lower prevision on $\mathscr{C}(\mathscr{X}_G)$ that is defined by

$$\underline{P}_G^*(f\rfloor B) := \begin{cases} P_S^*(f\rfloor B_S) & \text{if } f \in \mathscr{G}(\mathscr{X}_S) \text{ and } (\exists B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S))\ B = B_S \times B_I; \\ \underline{P}_G(f\rfloor B) & \text{otherwise.} \end{cases}$$

Assume *ex absurdo* that $\mathscr{E}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$ is incoherent. As explained in Section 2.2.1$_{39}$, this implies that $\mathscr{E}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$ does not avoid null gain [D1$_{38}$]: $0 \in \mathscr{E}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$. Since $\mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}$ and $\mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$ are coherent, it now follows from the definition of $\mathscr{E}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$ that there are $f_S \in \mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}$ and $f_G \in \mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$ such that $\mathbb{I}_{B_I} f_S + f_G = 0$.

By definition of $\mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}$, we know that there are $n \in \mathbb{N}$, $(f_i, B_i) \in \mathscr{C}(\mathscr{X}_S)$, $\varepsilon_i \in \mathbb{R}_{>0}$, $\lambda_i \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} f_S &= \sum_{i=1}^n \lambda_i \mathbb{I}_{B_i}[f_i - P_S^*(f_i\rfloor B_i) + \varepsilon_i] \\ &= \sum_{i=1}^n \lambda_i \mathbb{I}_{B_i}[f_i - P_S^*(f_i\rfloor B_i) + \frac{\varepsilon_i}{2}] + \sum_{i=1}^n \lambda_i \mathbb{I}_{B_i} \frac{\varepsilon_i}{2} \geq \mathbb{I}_{B_S}[h_S + \varepsilon], \end{aligned} \tag{4.15}$$

where we let $B_S := \cup_{i=1}^n B_i$, $\varepsilon := 1/2 \min_{i=1}^n \lambda_i \varepsilon_i$ and

$$h_S := \sum_{i=1}^n \lambda_i \mathbb{I}_{B_i} \left[f_i - P_S^*(f_i\rfloor B_i) + \frac{\varepsilon_i}{2} \right] \in \mathscr{G}(B_S).$$

Observe that $\mathbb{I}_{B_S} h_S \in \mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}$. Since $P_S^*(\cdot\rfloor\cdot)$ is coherent and therefore coincides with its natural extension, we have that $P_S^*(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}}$. Hence, since $\mathbb{I}_{B_S} h_S \in \mathscr{E}_{P_S^*(\cdot\rfloor\cdot)}$, we know that $P_S^*(h_S\rfloor B_S) \geq 0$.

Let f_S' be the restriction of f_S to B_S. It then follows from Equation (4.15) that $f_S = \mathbb{I}_{B_S} f_S = \mathbb{I}_{B_S} f_S'$ and that $f_S' \geq h_S + \varepsilon$. The linearity of $P_S^*(\cdot\rfloor B_S)$ now implies that $P_S^*(-f_S'\rfloor B_S) \leq P_S^*(-h_S - \varepsilon\rfloor B_S) = -P_S^*(h_S\rfloor B_S) - \varepsilon < 0$. Since $P_S^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}_S(\cdot\rfloor\cdot\times B_I)}$ and therefore $P_S^*(\cdot\rfloor\cdot) \geq \underline{P}_S(\cdot\rfloor\cdot\times B_I)$, this in turn implies that $\underline{P}_S(-f_S'\rfloor B_S \times B_I) < 0$. Hence, we find that $\underline{P}_G(-f_S'\rfloor B_S \times B_I) < 0$. Since the coherence of $\underline{P}_G(\cdot\rfloor\cdot)$ implies that it coincides with its natural extension, we also know that $\underline{P}_G(\cdot\rfloor\cdot)$ is the conditional lower prevision that corresponds to $\mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$. By combining this with the fact that $\underline{P}_G(-f_S'\rfloor B_S \times B_I) < 0$, we find that $\mathbb{I}_{B_I}\mathbb{I}_{B_S}[-f_S'] \notin \mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$. Since $f_G = -\mathbb{I}_{B_I} f_S = -\mathbb{I}_{B_I}\mathbb{I}_{B_S} f_S' = \mathbb{I}_{B_I}\mathbb{I}_{B_S}[-f_S']$ and $f_G \in \mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$, this is a contradiction. Hence, we may conclude that $\mathscr{E}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$ is coherent.

Due to Proposition 3$_{45}$, this allows us to infer that there is some conditional linear prevision $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ such that $P_G(\cdot\rfloor\cdot) \geq \underline{P}_G^*(\cdot\rfloor\cdot)$. Since $P_S^*(\cdot\rfloor\cdot) \geq \underline{P}_S(\cdot\rfloor\cdot\times B_I)$, this implies that $P_G(\cdot\rfloor\cdot) \geq \underline{P}_G(\cdot\rfloor\cdot)$ or, equivalently, that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}$. Furthermore, $P_G(\cdot\rfloor\cdot) \geq \underline{P}_G^*(\cdot\rfloor\cdot)$ also implies that $P_S(\cdot\rfloor\cdot\times B_I) \geq P_S^*(\cdot\rfloor\cdot)$. Since $P_S(\cdot\rfloor\cdot\times B_I)$ and $P_S^*(\cdot\rfloor\cdot)$ are both conditional linear previsions on $\mathscr{C}(\mathscr{X}_S)$, this implies that $P_S(\cdot\rfloor\cdot\times B_I) = P_S^*(\cdot\rfloor\cdot)$, which allows us to conclude that $P_S^*(\cdot\rfloor\cdot) \in \mathrm{marg}_S^{\mathrm{c}}(\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}|B_I)$. □

Finally, for a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\mathscr{X}_G)$, we define

$$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_G\rfloor|B_I) := \{p_S(\cdot\rfloor\cdot\times B_I)\colon p_G(\cdot\rfloor\cdot)\in\mathscr{F}_G\},$$

where, for every $p_G(\cdot\rfloor\cdot)\in\mathscr{F}_G$, $p_S(\cdot\rfloor\cdot\times B_I)$ is a full conditional probability mass function on $\mathscr{C}_*(\mathscr{X}_S)$, defined for all $(x_S,B_S)\in\mathscr{C}_*(\mathscr{X}_S)$ by

$$p_S(x_S\rfloor B_S\times B_I) := \sum_{x_{G\setminus S}\in B_I\times\mathscr{X}_{G\setminus(I\cup S)}} p_G(x_S,x_{G\setminus S}\rfloor B_S\times B_I).$$

As a direct consequence of Equation $(4.9)_{108}$ and the one-to-one correspondence between conditional linear previsions and full conditional probability mass functions, we find that $\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_{\mathscr{F}_G}\rfloor|B_I)=\mathscr{M}_{\mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_G\rfloor|B_I)}$. Similarly, for any set $\mathscr{M}_G$ of conditional linear previsions on $\mathscr{C}(\mathscr{X}_G)$, we have that $\mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_{\mathscr{M}_G}\rfloor|B_I)=\mathscr{F}_{\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|B_I)}$.

4.2.5 *Conditioning and updating*

In the previous two sections, all the required conditional models were available, simply because we started from a *conditional* lower prevision, a set of *conditional* linear previsions, and so on. All we did was consider special cases of these conditional models, and show how they can be marginalised. However, in practice, these conditional models are not always given. If they are not, they have to be derived from unconditional ones or from other conditional ones, using the methods discussed in the previous chapters.[8]

The concepts that we introduced in Section $4.2.3_{105}$ remain applicable in these situations as well, in the way you would expect them to. For example, if $\mathscr{M}_G$ is a set of unconditional linear previsions on $\mathscr{G}(\mathscr{X}_G)$, we let $\mathrm{marg}_S(\mathscr{M}_G\rfloor^{\mathrm{r}}B_S\times B_I)$ be the set of linear previsions on $\mathscr{G}(\mathscr{X}_S)$ obtained by applying Equation $(4.7)_{107}$ to every element of $\mathscr{M}_G\rfloor^{\mathrm{r}}B_S\times B_I$, where $B_S\times B_I$—as before—serves as a shorthand notation for $B_S\times B_I\times\mathscr{X}_{G\setminus(S\cup I)}$. Similarly, for the natural extension $\underline{E}(\cdot\rfloor\cdot)$ and regular extension $\underline{R}(\cdot\rfloor\cdot)$ of a (conditional) lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}\subseteq\mathscr{C}(\mathscr{X}_G)$, the lower previsions $\underline{E}_S(\cdot\rfloor B_S\times B_I)$ and $\underline{R}_S(\cdot\rfloor B_S\times B_I)$ are defined by

$$\underline{E}_S(f\rfloor B_S\times B_I) := \underline{E}_G(f\rfloor B_S\times B_I) \text{ for all } f\in\mathscr{G}(\mathscr{X}_S)$$

and

$$\underline{R}_S(f\rfloor B_S\times B_I) := \underline{R}_G(f\rfloor B_S\times B_I) \text{ for all } f\in\mathscr{G}(\mathscr{X}_S),$$

respectively. By not fixing the set B_S in $\underline{E}_S(\cdot\rfloor B_S\times B_I)$, we obtain—as in Section $4.2.4_{109}$—a conditional lower prevision $\underline{E}_S(\cdot\rfloor\cdot\times B_I)$ on $\mathscr{C}(\mathscr{X}_S)$, and similarly for $\underline{R}_S(\cdot\rfloor\cdot\times B_I)$.

[8] In particular: Sections $2.3.2_{44}$, 2.7_{59} and $3.4.7_{90}$.

Our justifications for updating by means of conditioning, as discussed in Chapter 3_{69}, remain applicable as well. Once these updated models have been constructed—for example, by means of conditioning—we can use the techniques discussed in Section $4.2.3_{105}$ to marginalise them. In this way, we find that the lower previsions $\underline{E}_S(\cdot\rfloor B_S \times B_I)$ and $\underline{R}_S(\cdot\rfloor B_S \times B_I)$ can be interpreted as updated belief models for the variable X_S, and similarly for $\mathrm{marg}_S(\mathscr{M}_G\rfloor^{\mathrm{r}} B_S \times B_I)$, $\mathrm{marg}_S(\mathscr{M}_G\rfloor^{\mathrm{n}} B_S \times B_I)$, and so on. These models are applicable in situations where we have some—possibly partial—observation about the variables that are represented by X_I—the event B_I—and X_S—the event B_S—and we want to use these observations to update our subject's beliefs about X_S. In many practical situations, there will not be an observation for X_S. In those cases, $\underline{E}_S(\cdot\rfloor \cdot \times B_I)$, $\underline{R}_S(\cdot\rfloor \cdot \times B_I)$, and so on, are the appropriate updated belief models for X_S.

With regard to updating and conditioning, the multivariate set-up is often easier than the univariate one because it allows us to specify events more intuitively. Events of the form $B_I = \{x_I\}$, with $x_I \in \mathscr{X}_I$, constitute an important example. Observing such an event simply means observing the value x_s of X_s for all $s \in I$. These events are typical in the context of updating: we learn the value of a number of variables and then wish to use this information to update our beliefs about other variables. For these kinds of events, it will often also be easier to fulfil the requirement that we should be informed about the occurrence of an event—and nothing more—if and only if it actually occurs, as needed in order for our justifications for updating by means of natural or regular extension to apply [see Section $3.1.2_{71}$].

4.2.6 *Conditional lower previsions as gambles*

A conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}(\Omega)$ takes two arguments, the first of which is a gamble f and the second of which is an event B. Because of how we have defined a conditional lower prevision, the event needs to be fixed first, because it determines the set $\mathscr{G}(B)$ from which f is to be taken—the domain of the unconditional lower prevision $\underline{P}(\cdot\rfloor B)$.

In a multivariate context, it is rather intuitive to reverse this process, especially if we are considering lower previsions of the form $\underline{P}_S(f\rfloor x_I)$. Indeed, for a fixed gamble $f \in \mathscr{G}(\mathscr{X}_S)$, it feels rather natural to regard $\underline{P}_S(f\rfloor x_I)$ as a function of x_I. More formally, we let $\underline{P}_S(f\rfloor X_I)$ be a gamble on $\mathscr{G}(\mathscr{X}_I)$, defined by

$$\underline{P}_S(f\rfloor X_I)(x_I) := \underline{P}_S(f\rfloor x_I) \text{ for all } x_I \in \mathscr{X}_I.$$

The operator $\underline{P}_S(\cdot\rfloor X_I)$ is map from $\mathscr{G}(\mathscr{X}_S)$ to $\mathscr{G}(\mathscr{X}_I)$. With every gamble $f \in \mathscr{G}(\mathscr{X}_S)$, it associates a corresponding gamble $\underline{P}_S(f\rfloor X_I) \in \mathscr{G}(\mathscr{X}_I)$. The domain of the operator $\underline{P}_S(\cdot\rfloor X_I)$ can also be extended to include gambles $f \in \mathscr{G}(\mathscr{X}_W)$, for some $S \subseteq W \subseteq G$, by letting $\underline{P}_S(f\rfloor X_I)$ be a gamble on

$\mathscr{X}_{(W\setminus S)\cup I}$, defined by

$$\underline{P}_S(f\rfloor X_I)(x_{(W\setminus S)\cup I}) := \underline{P}_S(f(X_S, x_{W\setminus S})\rfloor x_I) \text{ for all } x_{(W\setminus S)\cup I} \in \mathscr{X}_{(W\setminus S)\cup I}.$$

In this way, $\underline{P}_S(\cdot\rfloor X_I)$ can also be regarded as a map from $\mathscr{G}(\mathscr{X}_W)$ to $\mathscr{G}(\mathscr{X}_{(W\setminus S)\cup I})$. For any $V \supseteq W \cup I$, the concept of cylindrical extension [see Section 4.2.1$_{102}$] even allows us to regard $\underline{P}_S(f\rfloor X_I)$ as a gamble on $\mathscr{X}_{V\setminus S}$.

For conditional linear previsions, these conventions allow us to state the well-known law of iterated expectation or, in our terminology, the *law of iterated prevision*. Consider pairwise disjoint subsets S_1, S_2 and I of G and let $S := S_1 \cup S_2$. Then for any conditional linear prevision $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, we have that

$$P_S(f\rfloor X_I) = P_{S_2}(P_{S_1}(f\rfloor X_{S_2\cup I})\rfloor X_I) \text{ for all } f \in \mathscr{G}(\mathscr{X}_G). \tag{4.16}$$

Proof of Equation (4.16). Let $p_G(\cdot\rfloor\cdot)$ be the full conditional probability mass function that corresponds to $P_G(\cdot\rfloor\cdot)$. Fix any gamble $f \in \mathscr{G}(\mathscr{X}_G)$ and regard $g := P_{S_1}(f\rfloor X_{S_2\cup I})$ as a gamble on $\mathscr{X}_{G\setminus S_1}$, as defined for all $x_{G\setminus S_1} \in \mathscr{X}_{G\setminus S_1}$ by

$$g(x_{G\setminus S_1}) := P_{S_1}(f\rfloor X_{S_2\cup I})(x_{G\setminus S_1}) := P_{S_1}(f(X_{S_1}, x_{G\setminus S_1})\rfloor x_{S_2\cup I}).$$

For any $x_{G\setminus S} \in \mathscr{X}_{G\setminus S}$, we then find that

$$\begin{aligned}
P_S(f\rfloor X_I)(x_{G\setminus S}) &= P_S(f(X_S, x_{G\setminus S})\rfloor x_I) \\
&= \sum_{x_S\in\mathscr{X}_S} f(x_S, x_{G\setminus S}) p_S(x_S\rfloor x_I) \\
&= \sum_{x_{S_2}\in\mathscr{X}_{S_2}} \sum_{x_{S_1}\in\mathscr{X}_{S_1}} f(x_{S_1}, x_{S_2}, x_{G\setminus S}) p_{S_1}(x_{S_1}\rfloor x_{S_2\cup I}) p_{S_2}(x_{S_2}\rfloor x_I) \\
&= \sum_{x_{S_2}\in\mathscr{X}_{S_2}} P_{S_1}(f(X_{S_1}, x_{S_2}, x_{G\setminus S})\rfloor x_{S_2\cup I}) p_{S_2}(x_{S_2}\rfloor x_I) \\
&= \sum_{x_{S_2}\in\mathscr{X}_{S_2}} g(x_{S_2}, x_{G\setminus S}) p_{S_2}(x_{S_2}\rfloor x_I) = P_{S_2}(g(X_{S_2}, x_{G\setminus S})\rfloor x_I) \\
&= P_{S_2}(g\rfloor X_I)(x_{G\setminus S}),
\end{aligned}$$

using Bayes's rule for the third equality. □

This proof shows that Equation (4.16) is an alternative formulation of the law of total probability.

4.3 EPISTEMIC IRRELEVANCE

At this point, we have all the tools necessary to introduce one of the most important concepts in this thesis: epistemic irrelevance. We will distinguish between three variants: epistemic value-irrelevance, epistemic subset-irrelevance and epistemic h-irrelevance. We focus on their conditional versions, as the unconditional ones can be recovered easily as a special case.

4.3.1 *Epistemic value-irrelevance*

Consider three pairwise disjoint subsets C, I, and S of G. When a subject judges X_I to be *epistemically value-irrelevant to X_S conditional on X_C*, as denoted by $\mathrm{VIR}(I,S \mid C)$, he assumes that if he knew the value of X_C, then knowing in addition which value X_I assumes in $\mathscr{X}_I$ would not affect his beliefs about X_S. Formally, depending on the framework that is used, he assumes that, for all $x_C \in \mathscr{X}_C$ and $x_I \in \mathscr{X}_I$:

$$\mathrm{marg}_S(\mathscr{D}_G \rfloor x_{C\cup I}) = \mathrm{marg}_S(\mathscr{D}_G \rfloor x_C).$$

or

$$\underline{P}_S(\cdot \rfloor x_{C\cup I}) = \underline{P}_S(\cdot \rfloor x_C).$$

or

$$\mathrm{marg}_S(\mathscr{M}_G \rfloor x_{C\cup I}) = \mathrm{marg}_S(\mathscr{M}_G \rfloor x_C).$$

or

$$\mathrm{marg}_S(\mathscr{F}_G \rfloor x_{C\cup I}) = \mathrm{marg}_S(\mathscr{F}_G \rfloor x_C).$$

The conditional models at both sides of these expressions should not be interpreted as updated ones. We take epistemic value-irrelevance to be a statement that is concerned with current beliefs about X_S, contingent on the fact that X_C (and X_I) take specific values.

For the unconditional versions of value-irrelevance, it suffices to let $C = \emptyset$. This makes sure the variable X_C has only one possible value $x_\emptyset$—and is therefore deterministic—so conditioning on that variable amounts to not conditioning at all.

4.3.2 *Epistemic subset-irrelevance*

Alternatively, a subject can make the stronger statement that he judges X_I to be *epistemically subset-irrelevant to X_S conditional on X_C*, as denoted by $\mathrm{SIR}(I,S \mid C)$. In that case, he assumes that if he knew the value of X_C, then receiving the additional information that X_I is an element of any non-empty subset B_I of $\mathscr{X}_I$ would not affect his beliefs about X_S. In other words, he assumes that, for all $x_C \in \mathscr{X}_C$ and all $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$:

$$\mathrm{marg}_S(\mathscr{D}_G \rfloor \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{D}_G \rfloor x_C) \tag{4.17}$$

or

$$\underline{P}_S(\cdot \rfloor \{x_C\} \times B_I) = \underline{P}_S(\cdot \rfloor x_C) \tag{4.18}$$

or

$$\mathrm{marg}_S(\mathscr{M}_G \rfloor \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{M}_G \rfloor x_C) \tag{4.19}$$

or

$$\mathrm{marg}_S(\mathscr{F}_G \rfloor \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{F}_G \rfloor x_C). \tag{4.20}$$

Again, the conditional models at both sides of these equations should be interpreted as contingent models, not as updated ones. The unconditional version corresponds to $C = \emptyset$.

Making a subset-irrelevance statement $\mathrm{SIR}(I, S \mid C)$ clearly implies the corresponding value-irrelevance statement $\mathrm{VIR}(I, S \mid C)$. Even stronger, it implies that $\mathrm{VIR}(I', S \mid C)$ for all $I' \subseteq I$. As the following counterexample illustrates, the converse relation does not hold in general.

Example 2. *Consider two variables X_{s_1} and X_{s_2} that take values in their respective state spaces $\mathscr{X}_{s_1} := \{a, b, c\}$ and $\mathscr{X}_{s_2} := \{0, 1\}$. Hence, $G = \{s_1, s_2\}$. Now let $I := \{s_1\}$, $S := \{s_2\}$, $C := \emptyset$ and $B_I := \{a, c\} \subseteq \mathscr{X}_I$, let $g \in \mathscr{G}(\mathscr{X}_S)$ be defined by $g(1) := 1$ and $g(0) := -1$, and define*

$$\mathscr{D}_G := \mathscr{E}(\{\mathbb{I}_{B_I} g\}) = \{f \in \mathscr{G}(\mathscr{X}_G) \colon (\exists \lambda \in \mathbb{R}_{>0})\ f \geq \lambda \mathbb{I}_{B_I} g \text{ or } f > 0\}.$$

According to this set of desirable gambles, X_I is epistemically value-irrelevant to X_S because, for all $x_I \in \mathscr{X}_I$, $\mathrm{marg}_S(\mathscr{D}_G \rfloor x_I) = \mathscr{G}(\mathscr{X}_S)_{>0} = \mathrm{marg}_S(\mathscr{D}_G)$. However, X_I is not subset-irrelevant to X_S, because

$$\mathrm{marg}_S(\mathscr{D}_G \rfloor B_I) = \mathscr{E}(\{g\}) = \{f \in \mathscr{G}(\mathscr{X}_O) \colon (\exists \lambda \in \mathbb{R}_{>0})\ f \geq \lambda g \text{ or } f > 0\}$$

is strictly larger than $\mathscr{G}(\mathscr{X}_S)_{>0}$ and therefore not equal to $\mathrm{marg}_S(\mathscr{D}_G)$. $\Diamond$

We consider epistemic subset-irrelevance to be the more natural of the two concepts, as it requires *all* information about the value of X_I to be irrelevant, including partial information, which is what—in our opinion—epistemic irrelevance should mean.[9] For example, in Example 2, although X_{s_1} is epistemically value-irrelevant to X_{s_2}, knowing that $X_{s_1} \neq b$ does affect our subjects beliefs about X_{s_2}; this would be impossible if X_{s_1} were epistemically subset-irrelevant to X_{s_2}.

Epistemic value-irrelevance and epistemic subset-irrelevance seem to be the prevailing notions of epistemic irrelevance in the literature on imprecise probabilities. However, the prefixes 'value' and 'subset' are not added.[10] Both notions are simply referred to as epistemic irrelevance and the distinction between them is usually not made [15, 18, 22, 45, 46, 106].[11] In an attempt to

[9] If the state space is infinite—a case we do not consider here—then the advantages of epistemic-subset irrelevance (and the notion of epistemic h-irrelevance that we consider further on) over epistemic-value irrelevance are even more prominent because epistemic value-irrelevance tends to become a very weak concept; see Example 4 (and Remark 1) in Reference [75].

[10] With the exception of References [28, 31], where we recently introduced the notion of epistemic subset-irrelevance for sets of desirable gambles; however, in these papers, we have also used the term epistemic irrelevance to refer to what we now prefer to call epistemic value-irrelevance.

[11] The notion of epistemic irrelevance that is mentioned in References [15, 22, 45, 46, 106] coincides with wat we have chosen to call epistemic value-irrelevance. On the other hand, Reference [18] uses the term epistemic irrelevance to refer to what we have called epistemic subset-irrelevance.

clarify the difference, we have chosen to let it be reflected in our terminology. However, the possible definitions of epistemic irrelevance have not been exhausted yet. The—in our opinion—most important variant is a very recent notion of irrelevance, called (epistemic) h-irrelevance by Cozman [19].[12]

4.3.3 *Epistemic h-irrelevance*

For frameworks other than sets of desirable gambles, epistemic value- and subset-irrelevance share a common weakness, which is that they require information about the value of X_I—be it the actual value or partial information—to be irrelevant only to unconditional beliefs about X_S, and not to beliefs about X_S that are conditional on some event $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$. For example, in the framework of lower previsions, an assessment that X_I is epistemically subset-irrelevant to X_S only requires that $\underline{P}_S(\cdot\rfloor B_I)$ should be equal to $\underline{P}_S(\cdot)$. As we know from the discussion in Section 4.2.4$_{109}$, this may not imply the equality of $\underline{P}_S(\cdot\rfloor B_S \times B_I)$ and $\underline{P}_S(\cdot\rfloor B_S)$. Nevertheless, it seems to us that any property that claims to reflect epistemic irrelevance should impose this equality; there seems to be no reason to restrict attention to unconditional beliefs about X_S only.

In order to strengthen the concept of epistemic (subset-)irrelevance in such a way that it displays this desired behaviour, Cozman introduced the concept of h-irrelevance [19, Definition 2], which he named after a similar precise-probabilistic notion introduced by Hammond [61]. Within the framework of lower previsions, Cozman calls X_I h-irrelevant to X_S conditional on X_C if, for all $x_C \in \mathscr{X}_C$, $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$ and $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$:

$$\underline{P}_S(\cdot\rfloor B_S \times \{x_C\} \times B_I) = \underline{P}_S(\cdot\rfloor B_S \times \{x_C\}). \tag{4.21}$$

Inspired by this definition, and using the notation introduced in Section 4.2.4$_{109}$, we say that a subject judges X_I to be *epistemically h-irrelevant to X_S conditional on X_C*, as denoted by $\mathrm{HIR}(I,S \mid C)$, if he assumes that, for all $x_C \in \mathscr{X}_C$ and $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$:

$$\mathrm{marg}_S(\mathscr{D}_G\rfloor\{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{D}_G\rfloor x_C) \tag{4.22}$$

or

$$\underline{P}_S(\cdot\rfloor \cdot \times \{x_C\} \times B_I) = \underline{P}_S(\cdot\rfloor \cdot \times x_C) \tag{4.23}$$

or

$$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|\{x_C\} \times B_I) = \mathrm{marg}^{\mathrm{c}}_S(\mathscr{M}_G\rfloor|x_C) \tag{4.24}$$

or

$$\mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_G\rfloor|\{x_C\} \times B_I) = \mathrm{marg}^{\mathrm{c}}_S(\mathscr{F}_G\rfloor|x_C), \tag{4.25}$$

[12] Besides epistemic value-, subset- and h-irrelevance, other—in our opinion less fundamental—notions of (epistemic) irrelevance have been defined as well; some of them are briefly mentioned in Section 4.4$_{120}$.

depending on the framework that is adopted. We add the prefix 'epistemic' because we believe that it better reflects the meaning of such an assessment. The irrelevance that is imposed here is with respect to a belief model—subjective knowledge—and need not be related to (a lack of) causation or some kind of 'physical' independence, concepts with which the word irrelevance—without the prefix 'epistemic'—might otherwise easily be associated.

Our definition for conditional lower previsions [Equation $(4.23)_{\frown}$] is clearly equivalent to that of Cozman [Equation $(4.21)_{\frown}$]. We prefer our notation because a similar equivalence does not hold for some of the other frameworks that we consider. For example, for the framework of sets of conditional linear previsions, mimicking Cozman's definition would lead us to impose that, for all $x_C \in \mathscr{X}_C$, $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$ and $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$:

$$\mathrm{marg}_S(\mathscr{M}_G \rfloor B_S \times \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{M}_G \rfloor B_S \times \{x_C\}). \tag{4.26}$$

However, as explained in Section $4.2.4_{109}$, this is not equivalent to our definition of epistemic h-irrelevance, which is stronger, in the sense that it implies Equation (4.26). A similar statement can be made for the framework of sets of full conditional probability mass functions.

The framework of sets of desirable gambles is special. For that framework, as the reader might already have noticed, our definitions for subset-irrelevance and h-irrelevance [Equations $(4.22)_{\frown}$ and $(4.17)_{115}$] are identical. This makes perfect sense because it follows from the discussion in Section $4.2.4_{109}$—in particular, Equation $(4.12)_{109}$—that, for a coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$, epistemic subset-irrelevance [Equation $(4.17)_{115}$] implies that

$$\mathrm{marg}_S(\mathscr{D}_G \rfloor B_S \times \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{D}_G \rfloor B_S \times \{x_C\}) \tag{4.27}$$

for all $x_C \in \mathscr{X}_C$, $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$ and $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$. Hence, for sets of desirable gambles, epistemic subset-irrelevance already displays the behaviour that is expected of a notion of epistemic h-irrelevance. It is therefore not necessary—nor possible—to strengthen the notion of epistemic subset-irrelevance any further; it is already strong enough. The notion of epistemic h-irrelevance is therefore redundant in the framework of sets of desirable gambles. The reason why we have chosen to define it anyway, is because it allows us to make general statements about epistemic h-irrelevance that apply to each of the four frameworks that we consider.

We consider epistemic h-irrelevance to be more fundamental than epistemic value- and subset-irrelevance. Not only does it require *all* information about the value of X_I—including partial information—to be irrelevant conditional on the value of X_C, it also requires it to be irrelevant to *all* beliefs about X_S—conditional *and* unconditional beliefs. We think that these are exactly the properties that a notion of epistemic irrelevance should have. *For*

this reason, from now on, we will drop the prefix 'h' and refer to epistemic h-irrelevance simply as epistemic irrelevance, denoting it by $\mathrm{IR}(I,S \mid C)$ *instead of* $\mathrm{HIR}(I,S \mid C)$. This is unconventional; as we have mentioned in the previous section, most authors use the term epistemic irrelevance to refer to either epistemic value-irrelevance of epistemic subset-irrelevance. However, we are convinced that it is appropriate, and that over time, epistemic h-irrelevance will become the prevailing notion of epistemic irrelevance. We think that epistemic value- and subset-irrelevance should be regarded as simplifications, which should only be adopted if the added strength of epistemic h-irrelevance is not important for the problem at hand.

Whenever the difference between the three different versions of epistemic irrelevance is not relevant, we will also use epistemic irrelevance as a generic term that refers to all three of them.

4.3.4 *Does it matter which framework we use?*

Suppose now that we have a coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$, a coherent conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, a set $\mathscr{M}_G$ of conditional linear previsions on $\mathscr{C}(\mathscr{X}_G)$ and a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\mathscr{X}_G)$ that are all related to each other, in the sense that $\underline{P}_G(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}_G}(\cdot\rfloor\cdot)$, $\mathscr{M}_G = \mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}$ and $\mathscr{F}_G = \mathscr{F}_{\mathscr{M}_G}$. Does it matter in which framework we express an assessment of epistemic irrelevance? Yes, sometimes, but not that much.

The versions for coherent conditional lower previsions, sets of conditional linear previsions and sets of full conditional probability mass functions are equivalent: $\underline{P}_G(\cdot\rfloor\cdot)$ satisfies Equation $(4.23)_{117}$ if and only if $\mathscr{M}_G$ satisfies Equation $(4.24)_{117}$ if and only if $\mathscr{F}_G$ satisfies Equation $(4.25)_{117}$. The first equivalence follows from Equations $(4.10)_{108}$ and $(4.14)_{110}$ and Theorem 33_{110}. The second one follows from the one-to-one correspondence between conditional linear previsions and full conditional probability mass functions; see the end of Section 4.2.4 as well. Similar equivalences also hold for value-irrelevance and subset-irrelevance; Equation $(4.14)_{110}$ is then no longer necessary, and the role of Theorem 33_{110} is then taken up by Equation $(4.11)_{109}$.

The version for coherent sets of desirable gambles [Equation $(4.17)_{115}$] clearly implies the one for conditional lower previsions; this follows from Equation $(4.27)_{\frown}$ and the text that follows Equation $(4.7)_{107}$. Consequently, it also implies the other two versions. However, the converse relation does not necessarily hold; it may be the case that $\underline{P}_G(\cdot\rfloor\cdot)$ satisfies Equation $(4.23)_{117}$ whereas $\mathscr{D}_G$ does not satisfy Equation $(4.22)_{117}$. Similar statements apply to epistemic value- and subset-irrelevance too.

It follows from these considerations that if we prove an irrelevance statement in terms of sets of desirable gambles, we immediately obtain a corresponding result in terms of the other frameworks.

4.4 VARIOUS OTHER NOTIONS OF INDEPENDENCE

Epistemic irrelevance is not the only imprecise-probabilistic notion of independence that has been discussed in the literature. In fact, it is not even among the more popular ones. The reason why we focus on it here, is (a) because it will—obviously—be one of the essential tools further on in our study of credal networks under epistemic irrelevance, and (b) because we believe that it deserves more attention than it is currently receiving. In order to try and motivate (b), we now give a brief overview of other imprecise-probabilistic notions of independence, compare them to epistemic irrelevance, and discuss some of the advantages of the latter. We focus on independence for variables; in particular, independence of X_I and X_S. We could easily add a variable X_C and discuss independence of X_I and X_S conditional on X_C; our reason for not doing so is because it would only distract from the main message.[13] Other, more comprehensive overviews of imprecise-probabilistic notions of independence can be found in References [15, 19, 37].

In order to understand why there are multiple notions of imprecise-probabilistic independence, it is instrumental to start by taking a closer look at precise-probabilistic independence or, as we prefer to call it, stochastic independence. Consider a linear prevision on $\mathscr{G}(\mathscr{X}_G)$ and its corresponding probability measure. Then X_S and X_I are said to be stochastically independent if any of the following three equivalent requirements holds. For all $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$:

$$P(B_S \times B_I) = P(B_S)P(B_I) \tag{4.28}$$

or

$$P_S(B_S \rfloor B_I) = P_S(B_S) \text{ whenever } P(B_I) > 0 \tag{4.29}$$

or

$$P_I(B_I \rfloor B_S) = P_I(B_I) \text{ whenever } P(B_S) > 0. \tag{4.30}$$

Equation (4.28) is the most well-known of these three equivalent formulations. Its popularity is mainly due to the mathematical convenience of this so-called factorisation property. However, it provides little intuition about what independence means. In that respect, Equations (4.29) and (4.30) are more important: they clearly illustrate that independence means mutual irrelevance.

The way we have presented it so far, stochastic independence is a relatively weak concept, especially in the presence of zero probabilities. It seems that the only purpose of the inequalities in Equations (4.29) and (4.30) is to circumvent the issue of conditioning on events with probability zero. Since conditional linear previsions regard conditional models as primitive instead of as models that need to be obtained through conditioning, they do not require such an

[13] And because the mathematical details of notions of independence other than epistemic irrelevance will not be used in the following chapters anyway.

escape clause, and these inequalities can therefore simply be dropped. For Equation $(4.29)_{\frown}$, this results in the following requirement:

$$P_S(B_S \rfloor B_I) = P_S(B_S) \text{ for all } B_S \in \mathscr{P}_{\emptyset}(\mathscr{X}_S) \text{ and } B_I \in \mathscr{P}_{\emptyset}(\mathscr{X}_I).$$

This is equivalent to imposing that X_I should be epistemically subset-irrelevant to X_S. We could take this even further, and impose epistemic (h-)irrelevance instead. In any case, the main point here is that imposing epistemic irrelevance on a conditional linear prevision implies stochastic independence, but not the other way around. *We consider this to be a first important property of epistemic irrelevance: when applied to a precise model, it is a generalisation of the usual notion of stochastic independence.* For an overview of other precise-probabilistic generalisations of stochastic independence, see Reference [21].

When moving from precise- to imprecise-probabilistic notions of independence, two different routes can be taken. The first route is to stick with the notion of stochastic independence—or any of its generalisations—and to impose it on each—or some—of the elements of a set of precise models. The two best-known examples are complete and strong independence.

Consider a set $\mathscr{M}_G$ of linear previsions on $\mathscr{G}(\mathscr{X}_G)$. Then X_I and X_S are said to be *completely independent* if X_I and X_S are stochastically independent according to every $P_G \in \mathscr{M}_G$ [18]. For sets of conditional linear previsions, elementwise epistemic (value-, subset- or h-)irrelevance can be defined in the same way. These notions of independence are especially intuitive if the ideal of precision is adopted—if each of the elements of $\mathscr{M}_G$ is regarded as a candidate for some 'true' precise probabilistic model. Indeed, in that case, it often makes sense for a structural judgement—such as independence—to be imposed on each of the individual elements of $\mathscr{M}_G$ rather than on the set $\mathscr{M}_G$ itself. However, without an assumption of ideal precision, complete independence—and any generalised version of it—is merely a mathematical property that seems to have no intuitive meaning.[14] Furthermore, these notions of independence can only be applied to sets of precise models. They cannot be imposed on a coherent lower prevision $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$ because complete independence is incompatible with the convexity of the set $\mathscr{M}_{\underline{P}_G}$ [18, Section 3.1], and definitely not on a coherent set of desirable gambles. *We consider this to be a second advantage of epistemic irrelevance: it can easily be imposed in each of the four frameworks that we consider.*

If $\mathscr{M}_G$ is a closed and convex set of linear previsions on $\mathscr{G}(\mathscr{X}_G)$, then X_I and X_S are said to be *strongly independent* if X_I and X_S are stochastically in-

[14] In all fairness, there are some very recent results that indicate that complete independence could be axiomatised in terms of choice functions, a framework for modelling uncertainty that falls beyond the scope of the present discussion [18, Section 4]. We believe this to be promising. In cases where these kinds of axioms can be defended, they could lead to a justification for complete independence that does not require an assumption of ideal precision.

dependent according to every $P \in \mathrm{ext}(\mathscr{M}_G)$, or equivalently, if X_G is a convex hull of linear previsions for which X_I and X_S are stochastically independent. More generally, for any set $\mathscr{M}_G$ of conditional linear previsions on $\mathscr{C}(\mathscr{X}_G)$, we could require $\mathscr{M}_G$ to be of the form $\mathscr{M}_{\underline{P}_G(\cdot\rfloor\cdot)}$, with $\underline{P}_G(\cdot\rfloor\cdot)$ the lower envelope of some set of conditional linear previsions for which X_I and X_S satisfy some generalised notion of stochastic independence, such as epistemic value-irrelevance, epistemic subset-irrelevance, and so on. Strong independence seems to have been introduced out of mathematical convenience, and is an attempt to reconcile complete independence with convexity. However, it seems to have little intuitive meaning, and is therefore hard to justify as a concept of independence.[15] Nevertheless, rather surprisingly, it is by far the most popular imprecise-probabilistic notion of independence.

A second route that can be taken to move from precise- to imprecise-probabilistic independence, is to forget about the individual elements of a set of precise models, and to develop properties that can be expressed directly in terms of the imprecise model. *The advantage of taking this route—as epistemic irrelevance does—is that it does not require an assumption of ideal precision.* Furthermore, since they do not need to refer to individual precise models, the resulting notions of independence can also be expressed in the framework of lower previsions and sets of desirable gambles. For this second route, three different approaches can be distinguished, each of which tries to generalise stochastic independence in a different way.

The first approach is to develop a generalised version of the factorisation property [Equation $(4.28)_{120}$]. Reference [46, Section 3.1] provides an overview of some of the ways in which this can be done. With the exception of Kuznetsov independence [20], most of these properties are not referred to as notions of independence, and rightly so. We believe that the use of Equation $(4.28)_{120}$ is justified only because of its equivalence with Equations $(4.29)_{120}$ and $(4.30)_{120}$. Without a similar equivalence, generalisations of factorisation are merely mathematical properties; we see no reason why they should be referred to as notions of independence.

The second approach is to develop a generalised version of Equation $(4.29)_{120}$. This is exactly what we have done in Section 4.3_{114}; the three versions of *epistemic irrelevance* that we have discussed can all be regarded as generalisations of Equation $(4.29)_{120}$.[16] The philosophical advantage of such an approach is that *it has a very clear and intuitive meaning: it simply means that, according to the subject whose beliefs we are modelling, informa-*

[15] See References [18, 26] for (partial) justifications of special cases, using an assessment of infinite exchangeability.

[16] At first sight, it might seem like this is not the case for epistemic value-irrelevance, since it does not consider conditioning on events B_S, but only on values x_S. However, the difference only appears in the presence of (lower) probability zero, and this case is excluded by Equation $(4.29)_{120}$.

tion about the value of X_I will not alter his belief model for X_S—his 'epistemic' uncertainty about X_S.

As illustrated by the three different versions that we have introduced, this intuitive property can be formalised in different ways. We believe epistemic h-irrelevance to be the most fundamental of those three. Some authors have adopted yet other versions than the ones we consider, by requiring the conditional models that appear in a statement of irrelevance to be derived by means of regular extension—in the sense of Section 2.7.1$_{59}$—instead of considering them to be given as part of a 'full conditional' model such as a conditional lower prevision, a set of conditional linear previsions, and so on; Reference [19, Section 2.1] provides an overview. It seems that the sole purpose for doing so is to avoid the issue of conditioning on events with lower probability zero. However, given that this issue can easily be dealt with by considering conditional models as primitive notions—as 'full conditional' models do—we see no reason why a conditioning rule—such as regular extension—should be part of the definition of epistemic irrelevance. Epistemic irrelevance relates different conditional models; where these models come from and how they are obtained should—in our opinion—not be part of the definition.

Although an assessment of epistemic irrelevance does not require an assumption of ideal precision, it is nevertheless fully compatible with it. If such an assumption is made, epistemic subset-irrelevance assesses that the *set* $\mathrm{marg}_S(\mathscr{M}_G\rfloor B_I)$ of candidate belief models for X_S is not affected by information about the value of X_I—the event B_I. However, for an individual conditional linear prevision $P_G(\cdot\rfloor\cdot)$ in $\mathscr{M}_G$, epistemic subset-irrelevance allows for $P_S(\cdot\rfloor B_I)$ to be different from $P_S(\cdot\rfloor B_I')$, for $B_I \neq B_I'$, and thereby allows for dependencies between X_I and X_S. These dependencies are however limited by the fact that epistemic subset-irrelevance requires that $P_S(\cdot\rfloor B_I)$ and $P_S(\cdot\rfloor B_I')$ should both be elements of $\mathrm{marg}_S(\mathscr{M}_G)$. If $\mathrm{marg}_S(\mathscr{M}_G)$ is small, the difference between $P_S(\cdot\rfloor B_I)$ and $P_S(\cdot\rfloor B_I')$ is bounded, and this limits the amount of dependence between X_I and X_S. In that sense, *under an assumption of ideal precision, from the point of view of the individual candidate models, epistemic (subset-)irrelevance can be regarded as a notion of almost-independence or almost-irrelevance*. If $\mathrm{marg}_S(\mathscr{M}_G)$ is a singleton, it reduces to actual independence. Since assessments of independence are usually regarded as 'approximations of reality' anyway, we consider such a notion of almost-independence to be particularly realistic in practice.

The third approach is to impose mutual irrelevance, which is simply a combination of two irrelevance statements: X_I is epistemically irrelevant to X_S—using any of the definitions that were discussed before—and X_S is epistemically irrelevant to X_I. This notion of independence is called *epistemic independence* [106]. Depending on the notion of irrelevance on which it is based, one can distinguish between epistemic value-independence, epistemic subset-independence, etcetera.

Epistemic independence, strong independence and complete independence

can all be regarded as strengthened versions of epistemic irrelevance, in the sense that—when suitably defined—they *all imply epistemic irrelevance*. Making an assessment of epistemic irrelevance is therefore easier to justify. As we have seen in this section, replacing epistemic irrelevance by one of the other notions of independence requires additional assessments or assumptions. Sections 5.3_{132} and 5.5_{145} discuss some additional advantages of epistemic irrelevance in the context of graphical models.

Notions of independence are sometimes also compared with respect to the graphoid properties that they satisfy [18, 21, 22]. For example, epistemic h-irrelevance satisfies more graphoid properties than epistemic value- and subset-irrelevance, and this could be regarded as an argument in favour of epistemic h-irrelevance. However, we believe that graphoid properties are overrated as axioms. In an imprecise-probabilistic context, some graphoid properties—such as contraction and intersection—should not be regarded as axioms that a notion of independence should satisfy, but rather as mathematically convenient properties of models, which may or may not be satisfied. We discuss this point further in Section $6.5.3_{168}$.

4.5 GENERALISING THE NOTION OF NATURAL EXTENSION

We have managed to fill a chapter with all kinds of tools for—and properties of—multivariate models. However, we have not yet explained how to construct one. So, in order to end this chapter: how can we construct a multivariate model? The general answer is very simple: in the same way as a univariate one, by means of natural extension.

The starting point is to gather assessments, either from a domain expert or from data. In the framework of sets of desirable gambles, this would be some set $\mathscr{A}_G$ of gambles that are considered to be desirable. For lower previsions, this would be a partial specification of a conditional lower prevision, for some subdomain $\mathscr{C} \subseteq \mathscr{C}(\mathscr{X}_G)$, or equivalently, assessments of supremum contingent buying prices for some—usually limited—set of combinations of gambles and events. For sets of conditional linear previsions or sets of full conditional probability mass functions, an assessment consists of linear constraints on the individual elements of these sets. We will not discuss the practicalities of gathering such assessments any further; References [106, Chapter 4], [4, 8, 93, 107, 108] and [81, Section 3.2] provide plenty of information on the topic. For our present purposes, it suffices to say that these assessments can either be given directly by the expert, or obtained indirectly from data, in which case the role of the expert lies in selecting an appropriate model for deriving assessments from data.

A problem that arises when gathering such assessments in a multivariate set-up is that the size of the state space $\Omega = \mathscr{X}_G = \times_{i \in G} \mathscr{X}_i$ is exponential in the number of individual variables. For this reason, making assessments

about the global tuple-valued variable X_G directly is rather impractical. Fortunately, the multivariate set-up itself already solves this problem, because it provides us with tools that allow us to focus on individual variables X_s or small sets of variables, as represented by X_S, with $S \subseteq G$. Using the example of Section 4.1_{101}, we could for example make an assessment such as: given that there is no rainfall—$X_r = \text{no}$—the barometric pressure is at least as likely to be high than it is to be low—$p_b(X_b = \text{high} \rfloor X_r = \text{no}) \geq p_b(X_b = \text{low} \rfloor X_r = \text{no})$. If we let $S = \{b\}$, $I = \{r\}$, and $B_I = \{\text{no}\}$, such an assessment is concerned with the local model $\text{marg}_S(\mathscr{F}_G \rfloor B_S)$. In the same way, local assessments in terms of other frameworks can also be considered. Making assessments about these local models is easier and more intuitive than trying to deal directly with the tuple-valued variable X_G. Nevertheless, since these local models are related to—are derived from—a global model, local assessments can also be interpreted as global assessments. For example, stating that a gamble $f \in \mathscr{G}(\mathscr{X}_S)$ belongs to $\text{marg}_S(\mathscr{D}_G \rfloor B_S)$ is equivalent to stating that the gamble $\mathbb{I}_{B_S} f \in \mathscr{G}(\mathscr{X}_G)$ is globally desirable—belongs to $\mathscr{D}_G$. This connection between local and global models is crucial for elicitation. On the one hand, we can gather intuitive local assessments that are concerned with only few variables. On the other hand, we can translate them into statements about the global variable X_G. In this way, many local assessments that are concerned with different local variables can be combined into a single global set of assessments.

Within a multivariate set-up, we do not need to restrict ourselves to direct—local or global—assessments, such as the ones discussed above. Other very important types of assessments are the structural ones, such as assessments of irrelevance or independence. On their own, structural assessments have little impact. For example, the vacuous conditional lower prevision $\underline{P}_{G,\text{v}}(\cdot \rfloor \cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, defined by

$$\underline{P}_{G,\text{v}}(f \rfloor B) := \min f \text{ for all } (f, B) \in \mathscr{C}(\mathscr{X}_G),$$

satisfies every possible epistemic irrelevance statement that one could come up with. However, combining these structural assessments with direct ones can lead to a substantial increase in information. For example, if X_I is assessed to be epistemically irrelevant to X_S, it implies that any direct assessment about $\underline{P}_S(\cdot)$ immediately leads to an analogous assessment about $\underline{P}_S(\cdot \rfloor B_I)$, simply because the assessment of epistemic irrelevance requires that $\underline{P}_S(\cdot)$ and $\underline{P}_S(\cdot \rfloor B_I)$ should be equal.

Once we have a collection of assessments—which may include both direct and structural ones—the next step is to use it to construct a global belief model. Depending on the framework that is adopted, this might be a coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$, a coherent conditional lower prevision $\underline{P}_G(\cdot \rfloor \cdot)$ on $\mathscr{C}(\mathscr{X}_G)$, a set $\mathscr{M}_G$ of conditional linear previsions on $\mathscr{C}(\mathscr{X}_G)$ or a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\mathscr{X}_G)$. In order to be able to associate a unique such global model with a collection of assessments, this collection needs to satisfy two crucial properties. First of all, there must be at least

one global model that is compatible with each of the assessments. Secondly, out of all global models for which this is the case, there should be a unique most conservative—most imprecise—one. If that is the case, this unique most conservative model is the only model that can reasonably be inferred from the assessments.

Satisfying the first property is a matter of careful elicitation. If it fails, some of the assessments need to be reconsidered. In order for the second property to be satisfied, it suffices that each assessment allows for *conservative reasoning*. The exact definition depends on the framework. For sets of desirable gambles, an assessment allows for conservative reasoning if it is preserved under taking intersections. Indeed, suppose that this is the case, and consider the set $\mathscr{D}_{G,\gamma}$, $\gamma \in \Gamma$, of all coherent sets of desirable gambles that satisfy our assessments. Then $\bigcap_{\gamma\in\Gamma} \mathscr{D}_{G,\gamma}$ will also be coherent, will also satisfy the assessments, and will furthermore be the smallest—most conservative—set of desirable gambles on $\mathscr{G}(\mathscr{X}_G)$ for which this is the case. Direct assessments clearly allow for conservative reasoning: if $f \in \mathscr{D}_{G,\gamma}$ for all $\gamma \in \Gamma$, then $f \in \bigcap_{\gamma\in\Gamma} \mathscr{D}_{G,\gamma}$. If there is a global model—a coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$—that is compatible with a direct assessment $\mathscr{A}_G$—such that $\mathscr{A}_G \subseteq \mathscr{D}_G$—then the unique most conservative such model is the natural extension $\mathscr{E}(\mathscr{A}_G)$ of $\mathscr{A}_G$, as introduced in Section 2.2.1$_{39}$. However, we can go further than that, because many structural assessments also allow for conservative reasoning. For our present purposes, epistemic irrelevance is the most important example: it is not hard to see that if $\mathscr{D}_{G,\gamma}$ satisfies a given epistemic irrelevance statement (using any of the three versions we consider) for every $\gamma \in \Gamma$, then $\bigcap_{\gamma\in\Gamma} \mathscr{D}_{G,\gamma}$ also satisfies this statement. When a type of structural assessments allows for conservative reasoning, we can combine such structural assessments with direct ones, and as soon as there is a single global model that satisfies the resulting collection of assessments, we are guaranteed that there is a unique most conservative model that satisfies them. *This model can be regarded as a generalised notion of the natural extension of Section 2.2.1*$_{39}$.

This approach also works for the other three frameworks that we consider, but the definition of 'allowing for conservative reasoning' changes. For coherent lower previsions, an assessment allows for conservative reasoning if it is preserved under taking lower envelopes, in the sense that if an assessment is satisfied by all coherent lower previsions $\underline{P}_{G,\gamma}(\cdot\rfloor\cdot)$ that are indexed by the elements γ of some set Γ, then the pointwise infimum of these coherent lower previsions should also satisfy the assessment. For sets of conditional linear previsions and sets of full conditional probability mass functions, an assessment allows for conservative reasoning if it is preserved under taking unions. Again, as was the case for sets of desirable gambles, it is not hard to see that assessments of epistemic irrelevance allow for conservative reasoning. Hence, also in these other frameworks, we can consider a generalised notion of natural extension that can deal with assessments of epistemic irrelevance.

Although the concept of conservative reasoning and the resulting gener-

alised notion of natural extension guarantee the existence of a unique most conservative model that is compatible with a collection of assessments (provided that there is at least one compatible model), they do not provide practical guidelines for constructing this most conservative model, because the index set Γ is usually infinite. The actual construction of the most conservative model that is compatible with a given collection of assessments is therefore a problem that needs to be solved on a case-by-case basis and that—depending on the case—may have a practical solution method or not. In the following chapters, we will focus on one specific case, where local direct assessments are combined with structural assessments of epistemic irrelevance that are generated automatically from a graphical model.

However, before we start focusing on this specific case, we should mention that epistemic irrelevance is not the only structural assessment for which this generalised notion of natural extension can be considered. There are many other examples of structural assessments that allow for conservative reasoning: symmetry, exchangeability, epistemic independence, complete independence, and so on. For more information, the interested reader can take a look at, among others, References [40, 45–47, 72, 99] and [106, Chapter 9]. We believe that the concept of conservative reasoning is important, and that it could be successfully applied in various contexts. The specific case that we are about to introduce and study in detail serves as a promising example.

5

CREDAL NETWORKS UNDER EPISTEMIC IRRELEVANCE

"...all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind..."

George E. P. Box

Although multivariate models provide an intuitive framework for modelling a multitude of real-life situations, they are far from easy to construct. Theoretically speaking, as explained in Section 4.5_{124}, we can consider any combination of direct assessments and structural ones, and consider the most conservative model that is compatible with them. However, in practice, coming up with such a combination of assessments may be hard, especially since they should satisfy two opposing criteria. On the one hand, it should be possible to construct a model that satisfies all these assessments. On the other hand, the most conservative such model should be informative enough to yield useful inferences.

In order to simplify this task, one can use graphical models as a tool for helping domain experts in providing such assessments, and ultimately, to construct multivariate models. In the precise-probabilistic case, Bayesian networks have proved to be a successful example. In this chapter, we will extend this approach to the imprecise-probabilistic setting. The resulting generalised notion of a Bayesian network is called a credal network, after the credal sets

that were initially used by Cozman to define it [16]. However, as we will see, it can be developed in terms of other imprecise-probabilistic frameworks as well, the most important of which are sets of desirable gambles and lower previsions. We will focus especially on credal networks under epistemic irrelevance, which—as their name suggests—adopt epistemic irrelevance as their notion of independence. Relatively little is known about this particular type of credal network. This dissertation is intended to be a first detailed study of the topic.

Basically, a credal network under epistemic irrelevance is just a special kind of multivariate model, which models some subject's beliefs about a finite set of variables X_s, with $s \in G$. We will consider four different frameworks. Depending on the framework that is chosen, the multivariate model will either be a set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$, a coherent conditional lower prevision on $\mathscr{C}(\mathscr{X}_G)$, a set $\mathscr{M}_G$ of linear conditional previsions on $\mathscr{C}(\mathscr{X}_G)$ or a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\mathscr{X}_G)$.

5.1 DIRECTED ACYCLIC GRAPHS

The first step in constructing a credal network is to connect the variables of our multivariate model by means of a graph. In particular, a directed acyclic graph (DAG), which consists of a finite set of nodes (vertices), joined into a network by a set of directed edges, each edge connecting one node with another. Since this directed graph is assumed to be acyclic, it is not possible to follow a directed sequence of edges from node to node and end up at the node one started out from.

Formally, we will identify the elements s of the index set G with the nodes of such a DAG. Technically, the DAG connects these nodes. Loosely speaking, it connects the variables X_s. For two nodes s and t in G, if there is a directed edge from s to t, we denote this as $s \to t$ and say that s is a *parent* of t and t is a *child* of s. For any node s, its set of parents is denoted by $P(s)$ and its set of children by $C(s)$.[1] If a node s has no parents, that is, $P(s) = \emptyset$, then we call s a *root node*. We will use $Ro(G)$ to refer to the set of all root nodes. If $C(s) = \emptyset$, then we call s a *leaf*, or a *terminal node*.

Two nodes s and t are said to have a *path* between them if one can start from s, follow the edges of the DAG regardless of their direction and end up in t. In other words: one can find a sequence of nodes $s = s_1, \dots, s_n = t$, $n \geq 1$, such that for all $i \in \{1, \dots, n-1\}$ either $s_i \to s_{i+1}$ or $s_i \leftarrow s_{i+1}$. If this sequence is such that $s_i \to s_{i+1}$ for all $i \in \{1, \dots, n-1\}$ (all edges in the path point away from s), we say that there is a *directed path* from

[1] The symbol P is already used for linear previsions as well. However, due to the rather different meaning of these two concepts, and because we will mainly consider lower previsions—rather than linear previsions—anyway, we do not consider this overload of notation to be problematic.

$$G = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}$$

$$P(s_7) = \{s_4, s_5\}$$
$$C(s_7) = \{s_9, s_{10}\}$$
$$D(s_7) = \{s_9, s_{10}\}$$
$$N(s_7) = \{s_1, s_2, s_3, s_6, s_8\}$$
$$A(s_7) = \{s_1, s_2, s_3, s_4, s_5\}$$

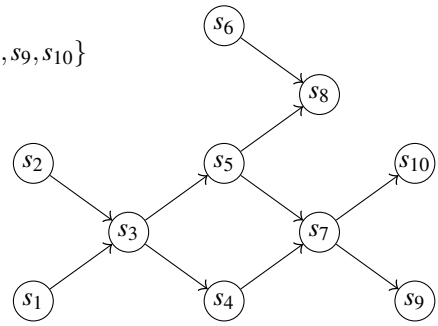


Figure 5.1: Example of a directed acyclic graph (DAG).

s to t and write $s \sqsubseteq t$. In that case we also say that s *precedes* t. If $s \sqsubseteq t$ and $s \neq t$, we say that s *strictly precedes* t and write $s \sqsubset t$. For any node s, we denote its set of *descendants* by $D(s) := \{t \in G \colon s \sqsubset t\}$, its set of *non-parent non-descendants* by $N(s) := G \setminus (P(s) \cup \{s\} \cup D(s))$ and its set of *ancestors* by $A(s) := \{t \in G \colon t \sqsubset s\}$. We also use the shorthand notation $PN(s) := P(s) \cup N(s) = G \setminus (\{s\} \cup D(s))$ to refer to the so-called *non-descendants* of s.

Example 3. *Consider the DAG in Figure 5.1, which consists of* 10 *nodes—the elements of G—and* 10 *edges. For the node $s_7 \in G$, the set of parents is $P(s_7) = \{s_4, s_5\}$, the set of children is $C(s_7) = \{s_9, s_{10}\}$, the set of ancestors is $A(s_7) = \{s_1, s_2, s_3, s_4, s_5\}$, the set of descendants is $D(s_7) = \{s_9, s_{10}\}$ and the set of non-parent non-descendants is $N(s_7) = \{s_1, s_2, s_3, s_6, s_8\}$.* ◊

5.2 ADDING LOCAL MODELS

The second step in constructing a credal network is to add *local uncertainty models* to each of the variables X_s, with $s \in G$. These local models are assumed to be given beforehand and serve as basic building blocks that we will combine into a global multivariate model.

For every node $s \in G$ and any instantiation $x_{P(s)} \in \mathscr{X}_{P(s)}$ of the variables associated with the parents of s, we are provided with a local belief model about X_s, conditional on $x_{P(s)}$. If we use the framework of sets of desirable gambles, this local model will be a coherent set of desirable gambles on $\mathscr{X}_s$, which we will denote by $\mathscr{D}_{s \rfloor x_{P(s)}}$. This local model represents our subject's beliefs about the variable X_s, contingent on the fact that its parents $X_{P(s)}$ assume the value $x_{P(s)}$. If s is a root node—has no parents—then $P(s) = \emptyset$, which implies that $X_{P(s)}$ can take only the single value $x_\emptyset$ and therefore, that $\mathscr{D}_{s \rfloor x_\emptyset}$

is effectively an unconditional model, which we will then simply denote by $\mathscr{D}_s := \mathscr{D}_{s\rfloor x_\emptyset}$.

These local models constitute direct assessments of desirability. Given that we do not enforce an exhaustive interpretation, they imply that, for all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$:

$$\mathscr{D}_{s\rfloor x_{P(s)}} \subseteq \mathrm{marg}_s(\mathscr{D}_G \rfloor x_{P(s)}) \tag{N1A}$$

It is not really essential for these local assessments to be coherent. Arbitrary assessments $\mathscr{A}_{s\rfloor x_{P(s)}} \subseteq \mathscr{X}_s$ can also be considered, as long as their natural extension exists. Without loss of generality, $\mathscr{D}_{s\rfloor x_{P(s)}}$ can then be taken to be equal to $\mathscr{E}(\mathscr{A}_{s\rfloor x_{P(s)}})$.

Alternatively, these local assessments can also be expressed in terms of other frameworks. Depending on the chosen framework, we assume that we are given, for all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, either a coherent lower prevision $\underline{P}_{s\rfloor x_{P(s)}}$ on $\mathscr{G}(\mathscr{X}_s)$, a closed and convex set $\mathscr{M}_{s\rfloor x_{P(s)}}$ of linear previsions on $\mathscr{G}(\mathscr{X}_s)$ or a closed and convex set $\mathscr{F}_{s\rfloor x_{P(s)}}$ of probability mass functions on $\mathscr{X}_s$—a credal set. These local assessment are again interpreted as constraints on a global model:

$$\underline{P}_s(\cdot\rfloor x_{P(s)}) \geq \underline{P}_{s\rfloor x_{P(s)}} \tag{N2A}$$

or

$$\mathrm{marg}_s(\mathscr{M}_G \rfloor x_{P(s)}) \subseteq \mathscr{M}_{s\rfloor x_{P(s)}} \tag{N3A}$$

or

$$\mathrm{marg}_s(\mathscr{F}_G \rfloor x_{P(s)}) \subseteq \mathscr{F}_{s\rfloor x_{P(s)}}. \tag{N4A}$$

Again, these local models do not need to be given as such, they can also be derived from local assessments.

For example, we can consider a—not necessarily coherent—lower prevision on some subset $\mathscr{K}$ of $\mathscr{G}(\mathscr{X}_s)$ and let $\underline{P}_{s\rfloor x_{P(s)}}$ be the unconditional part of its natural extension. In principle, we could also consider conditional lower previsions $\underline{P}(\cdot\rfloor\cdot)_{s\rfloor x_{P(s)}}$ on $\mathscr{C}(\mathscr{X}_s)$ as local models. In fact, if we were to do so, the theory that we are about to develop would still work—given some rather straightforward minor adaptations. Our only reason for not doing so is because it would be overkill. In practical modelling situations, local assessments are always 'unconditional', in the sense that they are only conditional on the value of the parent variables $X_{P(s)}$, and not conditional on an additional event $B_s \in \mathscr{P}_\emptyset(\mathscr{X}_s)$.

For that same reason, the elements of $\mathscr{M}_{s\rfloor x_{P(s)}}$ are taken to be unconditional linear previsions on $\mathscr{G}(\mathscr{X}_s)$ rather than conditional linear previsions on $\mathscr{C}(\mathscr{X}_s)$, and similarly for $\mathscr{F}_{s\rfloor x_{P(s)}}$. The requirement that $\mathscr{M}_{s\rfloor x_{P(s)}}$ and $\mathscr{F}_{s\rfloor x_{P(s)}}$ should be closed and convex could in principle be dropped. However, this would lead to non-trivial technical complications, especially with respect to the development of efficient algorithms. In any case, such an assumption of closedness and

convexity is standard in the context of credal networks.[2] It is also not that restrictive because many imprecise-probabilistic methods for elicitation—either directly from experts or indirectly by means of a statistical model and data—result in closed and convex sets of probability mass functions. For example, provided that it has at least one solution, a set of (non-strict) linear constraints has a closed and convex solution set. Furthermore, for any set of probability mass functions, the closure of its convex hull can always be used as a conservative outer approximation.

5.3 JUDGEMENTS OF EPISTEMIC IRRELEVANCE

In Bayesian networks, the graphical structure of the DAG of a network is taken to represent the following assessments: for any node s, conditional on its parent variables, the associated variable is independent of its non-parent non-descendant variables [64, Section 3.2.2]. When generalising this interpretation to credal networks, the classical notion of independence gets replaced by a more general, imprecise-probabilistic notion of independence. For credal networks under epistemic irrelevance, this notion of independence is epistemic irrelevance, as introduced in Section 4.3.3$_{117}$.[3]

Let us state this interpretation more formally. We assume that the graphical structure of the DAG embodies the following conditional epistemic irrelevance assessments which, by definition, turn the corresponding multivariate model into a *credal network under epistemic irrelevance*. Consider any node s in the network, its set of parents $P(s)$ and its set of non-parent non-descendants $N(s)$. Then *conditional on $X_{P(s)}$, $X_{N(s)}$ is assumed to be epistemically irrelevant to X_s*:

$$\mathrm{IR}(N(s), \{s\} \mid P(s)). \tag{NB}$$

Depending on the framework that is adopted, this means that, for all $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$:

$$\mathrm{marg}_s(\mathscr{D}_G \rfloor \{x_{P(s)}\} \times B_{N(s)}) = \mathrm{marg}_s(\mathscr{D}_G \rfloor x_{P(s)}) \tag{N1B}$$

or

$$\underline{P}_s(\cdot \rfloor \cdot \times \{x_{P(s)}\} \times B_{N(s)}) = \underline{P}_s(\cdot \rfloor \cdot \times \{x_{P(s)}\}) \tag{N2B}$$

or

$$\mathrm{marg}^{\mathrm{c}}_s(\mathscr{M}_G \rfloor | \{x_{P(s)}\} \times B_{N(s)}) = \mathrm{marg}^{\mathrm{c}}_s(\mathscr{M}_G \rfloor | x_{P(s)}) \tag{N3B}$$

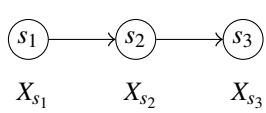
or

$$\mathrm{marg}^{\mathrm{c}}_s(\mathscr{F}_G \rfloor | \{x_{P(s)}\} \times B_{N(s)}) = \mathrm{marg}^{\mathrm{c}}_s(\mathscr{F}_G \rfloor | x_{P(s)}). \tag{N4B}$$

[2] See Section 5.5 for a short overview of the theory of credal networks.

[3] Recall that we take epistemic irrelevance to mean epistemic h-irrlevance.

$\mathrm{IR}(\{s_1\},\{s_3\} \mid \{s_2\})$



$\cancel{\mathrm{IR}(\{s_3\},\{s_1\} \mid \{s_2\})}$

Figure 5.2: Illustration of the non-symmetric irrelevance assessment that correspond to a DAG

An at first sight perhaps peculiar aspect of these assessments is that they are asymmetric. Consider for example the simple network in Figure 5.2. For that DAG, the only epistemic irrelevance assessment that follows from Requirement $(\mathrm{NB})_{\frown}$ is that X_{s_1} should be epistemically irrelevant to X_{s_3} conditional on X_{s_2}: $\mathrm{IR}(\{s_1\},\{s_3\} \mid \{s_2\})$. However, we do not impose that $\mathrm{IR}(\{s_3\},\{s_1\} \mid \{s_2\})$. We do not consider this aspect to be strange. In fact, we think that it is very natural, for the following reasons.

First of all, the assessments of epistemic irrelevance that we make—in the specific direction that we make them—have an intuitive meaning in terms of the local models that we introduced in the previous section. For example, Requirement $(\mathrm{N1B})_{\frown}$ is not just a structural assessment of irrelevance. In combination with Requirement $(\mathrm{N1A})_{131}$, it can also be interpreted as an additional collection of direct assessments: for any $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$, it leads us to impose that $\mathscr{D}_{s\rfloor x_{P(s)}}$ should be a subset of $\mathrm{marg}_s(\mathscr{D}_G\rfloor\{x_{P(s)}\} \times B_{N(s)})$. Similar things happen for the other frameworks. Basically, the effect of our epistemic irrelevance assessments is that the local assessments are duplicated and applied to a larger class of conditional models.

Secondly, there seems to be no fundamental reason why an assessment of epistemic irrelevance—say $\mathrm{IR}(\{s_1\},\{s_3\} \mid \{s_2\})$—should lead us to adopt the reverse version—say $\mathrm{IR}(\{s_3\},\{s_1\} \mid \{s_2\})$. As Dawid [23] put it: "The desirability of the symmetry property is not so obvious: if learning Y is irrelevant to X, must it follow that learning X is irrelevant to Y?". We agree with Dawid that this is indeed not obvious, and that therefore, there is no reason why it should be imposed by definition. Of course, this does not exclude that there might be instances where it makes sense to impose mutual irrelevance. In fact, we will sometimes do so: if a credal network under epistemic irrelevance consists of two disconnected nodes, the associated variables will be epistemically irrelevant to each other and therefore epistemically independent; Section 6.6_{171} discusses more general cases.

Thirdly, to a practitioner who constructs a credal—or Bayesian—network, the direction of the arrows matters intuitively. Suppose for example that X_{s_1} represents whether or not someone is a smoker, X_{s_2} represents whether or not someone has lung cancer and X_{s_3} represents whether or not some medical test indicates that the person has lung cancer. Then most people would put the edges in Figure 5.2 as we have put them, in that particular direction, and would

feel that changing the direction would alter the meaning of the assessment. Nevertheless, in a Bayesian network, because stochastic independence is symmetric, these two graphs—ours and the one where the arrows are reversed—correspond to the same structural assessments. In technical parlance: they are Markov-equivalent. For credal networks under epistemic irrelevance, no such equivalence is present: since epistemic irrelevance is an asymmetric notion of independence, graphs that are Markov-equivalent in the Bayesian network sense can lead to different structural assessments in a credal network under epistemic irrelevance.

The intuitive philosophical difference between Markov-equivalent graphs that is perceived by users of Bayesian networks is often associated with causality. As Shafer [90] put it: "we need a way to give mathematical and philosophical content to the differences between Markov-equivalent graphs, differences that are none the less real to practitioners for the fact that they are not expressed by conditional independence". The theory of causal networks [64, Chapter 21] does exactly that: although it is based on the theory of Bayesian networks, it differentiates between Markov equivalent graphs, both mathematically and philosophically. However, as explained above, it is not the only theory that is capable of doing so. It shares these features with the theory of credal networks under epistemic irrelevance.

The advantage that credal networks under epistemic irrelevance have over causal networks is that they are more generally applicable. The reason why—besides the fact that they allow for imprecision—is because causality is a property that is hard to define, and an assessment that is often too strong to make. Although it is natural to think of the edges in a credal network as some kind of causal links, the notion of 'causality' that people tend to associate with these links is usually rather vague and weak. For example, although one might think of the edge from X_{s_1} to X_{s_2} as a causal link between smoking and lung cancer, this is usually not taken to mean that smoking truly 'causes' lung cancer but rather that it influences it in some way, and similarly for the edge from X_{s_2} to X_{s_3}. Credal networks under epistemic irrelevance are fully compatible with this weak notion of causality. Furthermore, if the edges of the DAG in Figure 5.2$_{\frown}$ are truly causal—whatever that may mean—this will imply an assessment of epistemic irrelevance. If X_{s_1} causes X_{s_2} and X_{s_2} in turn causes X_{s_3}, then given X_{s_2}, X_{s_1} is causally irrelevant to X_{s_3}—whatever that may mean—and therefore definitely epistemically irrelevant. However, conversely, as mentioned in Section 4.3.3$_{117}$, an assessment of epistemic irrelevance does not require a causal interpretation, it is simply a statement about beliefs.

5.4 THE IRRELEVANT NATURAL EXTENSION

The Requirements (N1A)$_{131}$ and (N1B)$_{132}$—and their counterparts for frameworks other than sets of desirable gambles—usually do not determine a unique

global coherent model. However, as explained in Section 4.5$_{124}$, as soon as there is a single model that satisfies them, then among all solutions, there is a unique most conservative one. In the case of credal networks, we will call this unique most conservative solution the *irrelevant natural extension* of the network. The goal of this section is to define this irrelevant natural extension in terms of each of the four frameworks that we consider, to derive manageable expressions for them, and to establish connections between them. Finally, we will show that Bayesian networks correspond to the special case where the local models are precise.

5.4.1 *For sets of desirable gambles*

Within the framework of sets of desirable gambles, the local models of a credal network are coherent sets of desirable gambles $\mathscr{D}_{s\rfloor x_{P(s)}}$. The irrelevant natural extension is then a global coherent set of desirable gambles. In particular, the most conservative one that includes the local assessments and furthermore satisfies the irrelevancies that are encoded by the network. More formally: the irrelevant natural extension of the local models is the smallest coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$ that satisfies Requirements (N1A)$_{131}$ and (N1B)$_{132}$. As we are about to show, this global model can be constructed easily from the local ones. Basically, it suffices to extend the ideas of De Cooman and Miranda [45] to the context of credal networks; a published version of this material can be found in Reference [31].

We start by looking at a single given local model $\mathscr{D}_{s\rfloor z_{P(s)}}$ and investigate some of its implications for the global model $\mathscr{D}_G$. Consider any node s and fix values $z_{P(s)}$ and $z_{N(s)}$ for its parents and non-parent non-descendants. Then for any $f \in \mathscr{D}_{s\rfloor z_{P(s)}}$, we infer from Requirements (N1A)$_{131}$ and (N1B)$_{132}$ that $f \in \operatorname{marg}_s(\mathscr{D}_G\rfloor z_{PN(s)})$ and therefore, that $\mathbb{I}_{\{z_{PN(s)}\}}f \in \mathscr{D}_G$. Inspired by this observation, we introduce the following set of gambles on $\mathscr{X}_G$:

$$\mathscr{A}_G^{\mathrm{irr}} := \left\{\mathbb{I}_{\{z_{PN(s)}\}}f \colon s \in G,\ z_{PN(s)} \in \mathscr{X}_{PN(s)},\ f \in \mathscr{D}_{s\rfloor z_{P(s)}}\right\}. \tag{5.1}$$

It should be clear that $\mathscr{A}_G^{\mathrm{irr}}$ must be a subset of our joint model $\mathscr{D}_G$.

Proposition 34. *$\mathscr{A}_G^{\mathrm{irr}}$ is a subset of any global set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$ that satisfies Requirements* (N1A)$_{131}$ *and* (N1B)$_{132}$.

Proof of Proposition 34. Consider any $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$. As a consequence of Requirements (N1A)$_{131}$ and (N1B)$_{132}$, we see that $\operatorname{marg}_s(\mathscr{D}_G\rfloor z_{PN(s)})$ should be a superset of the given local model $\mathscr{D}_{s\rfloor z_{P(s)}}$. Hence, if we choose $f \in \mathscr{D}_{s\rfloor z_{P(s)}}$ and apply Equation (4.6)$_{107}$, it follows immediately that $\mathbb{I}_{\{z_{PN(s)}\}}f$ is an element of $\mathscr{D}_G$. □

Since the global set of desirable gambles $\mathscr{D}_G$ should also be coherent and therefore, in particular, a convex cone, we can derive the following corollary.

Corollary 35. $\mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}})$ *is a subset of any coherent set* $\mathscr{D}_G$ *of desirable gambles on* $\mathscr{X}_G$ *that satisfies Requirements* $(\mathrm{N1A})_{131}$ *and* $(\mathrm{N1B})_{132}$.

Proof of Corollary 35. We already know from Proposition $34_{\cap}$ that $\mathscr{A}_G^{\mathrm{irr}}$ is a subset of any joint model that satisfies Requirements $(\mathrm{N1A})_{131}$ and $(\mathrm{N1B})_{132}$: $\mathscr{A}_G^{\mathrm{irr}} \subseteq \mathscr{D}_G$. Applying the posi operator to both sides, we obtain that $\mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}}) \subseteq \mathrm{posi}(\mathscr{D}_G)$. If in addition to satisfying Requirements $(\mathrm{N1A})_{131}$ and $(\mathrm{N1B})_{132}$, $\mathscr{D}_G$ is also coherent, and thus in particular is a convex cone (satisfies properties $\mathrm{D3}_{39}$ and $\mathrm{D4}_{39}$), then $\mathrm{posi}(\mathscr{D}_G) = \mathscr{D}_G$ and we get that $\mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}}) \subseteq \mathscr{D}_G$. □

We now suggest the following expression for the irrelevant natural extension of the network:

$$\mathscr{D}_G^{\mathrm{irr}} := \mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}}). \tag{5.2}$$

Since we know from Corollary 35 that it is guaranteed to be a subset of the irrelevant natural extension, it is rather natural to propose it as a candidate for the irrelevant natural extension itself. In the remainder of this section, we will prove that $\mathscr{D}_G^{\mathrm{irr}}$ is indeed the unique smallest coherent set of desirable gambles on $\mathscr{X}_G$ that satisfies Requirements $(\mathrm{N1A})_{131}$ and $(\mathrm{N1B})_{132}$. We start by providing two alternative characterisations of $\mathscr{D}_G^{\mathrm{irr}}$, the first of which shows that it is not necessary to take positive linear combinations of elements of $\mathscr{A}_G^{\mathrm{irr}}$: a simple sum will do the job just fine.

Proposition 36. *A gamble* $f \in \mathscr{G}(\mathscr{X}_G)$ *is an element of* $\mathscr{D}_G^{\mathrm{irr}}$ *if and only if it can be written as:*

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}},$$

where, for all $s \in G$ *and* $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, *the gamble* $f_{s, z_{PN(s)}}$ *is either zero or an element of* $\mathscr{D}_{s \rfloor z_{P(s)}}$, *and at least one of them is non-zero.*

Proof of Proposition 36. Since $\mathscr{D}_G^{\mathrm{irr}} := \mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}})$, the 'if' part of this proof is trivial. For the 'only if' part, fix any $f \in \mathscr{D}_G^{\mathrm{irr}}$. By the definition of $\mathscr{D}_G^{\mathrm{irr}}$, we know that

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \sum_{i \in I(s, z_{PN(s)})} \lambda_{s, z_{PN(s)}, i} \mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}, i},$$

where, for all $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, $I(s, z_{PN(s)})$ is a (possibly empty) finite index set (but with at least one of them non-empty) and for all $s \in G$, $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ and $i \in I(s, z_{PN(s)})$, $\lambda_{s, z_{PN(s)}, i} \in \mathbb{R}_{>0}$ and $f_{s, z_{PN(s)}, i} \in \mathscr{D}_{s \rfloor z_{P(s)}}$.

We now construct, for every $s \in G$ and every $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, a gamble $f_{s,z_{PN(s)}} \in \mathscr{G}(\mathscr{X}_s)$. If $I(s, z_{PN(s)}) = \emptyset$, we let $f_{s,z_{PN(s)}} = 0$. If $I(s, z_{PN(s)}) \neq \emptyset$, we let

$$f_{s,z_{PN(s)}} = \sum_{i \in I(s,z_{PN(s)})} \lambda_{s,z_{PN(s)},i} f_{s,z_{PN(s)},i},$$

which is an element of $\mathscr{D}_{s\rfloor z_{P(s)}}$ (and thus different from zero) because $f_{s,z_{PN(s)},i} \in \mathscr{D}_{s\rfloor z_{P(s)}}$ for all $i \in I(s, z_{PN(s)})$ and because $\mathscr{D}_{s\rfloor z_{P(s)}}$ is coherent. It should now be clear that

$$f = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}},$$

in which the gambles $f_{s,z_{PN(s)}}$ are elements of $\mathscr{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero. □

Proposition 37. *$\mathscr{G}(\mathscr{X}_G)_{>0} \subseteq \mathscr{D}_G^{\text{irr}}$ and, consequently, $\mathscr{D}_G^{\text{irr}} = \mathscr{E}(\mathscr{A}_G^{\text{irr}})$.*

Proof of Proposition 37. The first step in the proof consists in showing that for any $z_G \in \mathscr{X}_G$, the indicator $\mathbb{I}_{\{z_G\}}$ is an element of $\mathscr{A}_G^{\text{irr}}$. To prove this, pick an arbitrary leaf $s \in G$. This is possible because a DAG with a finite amount of nodes always has at least one leaf. Since s is a leaf, it has no descendants and we therefore have that $G = \{s\} \cup PN(s)$. Due to the coherence of the local models, and in particular property D2$_{38}$, we know that the indicator $\mathbb{I}_{\{z_s\}}$ is an element of $\mathscr{D}_{s\rfloor z_{P(s)}}$. We can therefore apply Equation (5.1)$_{135}$ to see that $\mathbb{I}_{\{z_s\}}\mathbb{I}_{\{z_{PN(s)}\}} = \mathbb{I}_{\{z_{\{s\}\cup PN(s)}\}} = \mathbb{I}_{\{z_G\}}$ is an element of $\mathscr{A}_G^{\text{irr}}$. Since every $f \in \mathscr{G}(\mathscr{X}_G)_{>0}$ is a finite strictly positive linear combination of the indicators $\mathbb{I}_{\{z_G\}}$ that were constructed above, it follows that $\mathscr{G}(\mathscr{X}_G)_{>0} \subseteq \text{posi}(\mathscr{A}_G^{\text{irr}}) =: \mathscr{D}_G^{\text{irr}}$.

To prove the second part, notice that any gamble in $\mathscr{E}(\mathscr{A}_G^{\text{irr}})$ is a finite, strictly positive linear combination of gambles in $\mathscr{A}_G^{\text{irr}}$ and gambles in $\mathscr{G}(\mathscr{X}_G)_{>0}$. Since we have just shown that gambles in $\mathscr{G}(\mathscr{X}_G)_{>0}$ are themselves finite strictly positive linear combinations of specific indicators in $\mathscr{A}_G^{\text{irr}}$, this implies that $\mathscr{E}(\mathscr{A}_G^{\text{irr}}) \subseteq \text{posi}(\mathscr{A}_G^{\text{irr}})$. The converse inclusion is trivial and we thus find that $\mathscr{E}(\mathscr{A}_G^{\text{irr}}) = \text{posi}(\mathscr{A}_G^{\text{irr}}) =: \mathscr{D}_G^{\text{irr}}$. □

These two propositions serve as a first step towards the following important result, the proof of which can be found in Appendix 5.A$_{148}$.

Proposition 38. *$\mathscr{D}_G^{\text{irr}}$ is a coherent set of desirable gambles.*

The crucial step in our proof for this result is to consider a specific Bayesian network that has the same topology as our credal network and to use the corresponding joint probability mass function to construct a separating hyperplane argument. In this way, we are using existing coherence results for Bayesian

networks to prove their counterparts for credal networks. Since we now know that $\mathscr{D}_G^{\text{irr}}$ is coherent, Proposition $37_{\frown}$ can now be taken to mean that $\mathscr{D}_G^{\text{irr}}$ is the natural extension of $\mathscr{A}_G^{\text{irr}}$.

Next, we turn to a factorisation result that is essential in order to prove that $\mathscr{D}_G^{\text{irr}}$ extends the local models and expresses all conditional irrelevancies encoded in the network—satisfies Requirements $(\text{N1A})_{131}$ and $(\text{N1B})_{132}$. The proof for this result is given in Appendix $5.\text{A}_{148}$.

Proposition 39. *Fix any $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $g \in \mathscr{G}(\mathscr{X}_{N(s)})_{>0}$. Then for every $f \in \mathscr{G}(\mathscr{X}_s)$:*

$$g\mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}_G^{\text{irr}} \Leftrightarrow f \in \mathscr{D}_{s\rfloor x_{P(s)}}.$$

Corollary 40. *$\mathscr{D}_G^{\text{irr}}$ satisfies Requirements* $(\text{N1A})_{131}$ *and* $(\text{N1B})_{132}$. *Even stronger: it holds for all $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$ that:*

$$\text{marg}_s(\mathscr{D}_G^{\text{irr}} \rfloor \{x_{P(s)}\} \times B_{N(s)}) = \text{marg}_s(\mathscr{D}_G^{\text{irr}} \rfloor x_{P(s)}) = \mathscr{D}_{s\rfloor x_{P(s)}}.$$

Proof of Corollary 40. Fix any $s \in G$, $f \in \mathscr{G}(\mathscr{X}_s)$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and any $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$. Since $\mathbb{I}_{B_{N(s)}}$ and 1 are elements of $\mathscr{G}(\mathscr{X}_{N(s)})_{>0}$, we infer from Equation $(4.6)_{107}$ and Proposition 39 that

$$\begin{aligned} f \in \text{marg}_s(\mathscr{D}_G^{\text{irr}} \rfloor \{x_{P(s)}\} \times B_{N(s)}) &\Leftrightarrow \mathbb{I}_{B_{N(s)}} \mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}_G^{\text{irr}} \\ &\Leftrightarrow f \in \mathscr{D}_{s\rfloor x_{P(s)}} \Leftrightarrow \mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}_G^{\text{irr}} \\ &\Leftrightarrow f \in \text{marg}_s(\mathscr{D}_G^{\text{irr}} \rfloor x_{P(s)}). \end{aligned}$$

□

Interestingly, although $(\text{N1A})_{131}$ only requires $\text{marg}_s(\mathscr{D}_G^{\text{irr}} \rfloor x_{P(s)})$ to be a superset of the local model $\mathscr{D}_{s\rfloor x_{P(s)}}$, they turn out to be equal.

At this point, we already know that $\mathscr{D}_G^{\text{irr}}$ is coherent and that it satisfies Requirements $(\text{N1A})_{131}$ and $(\text{N1B})_{132}$. The final result of this section ensures that $\mathscr{D}_G^{\text{irr}}$ is the smallest set for which this is the case, or equivalently, that it is irrelevant natural extension that we have been looking for.

Theorem 41. *$\mathscr{D}_G^{\text{irr}}$ is the smallest coherent set of desirable gambles on $\mathscr{X}_G$ that satisfies Requirements* $(\text{N1A})_{131}$ *and* $(\text{N1B})_{132}$.

Proof of Theorem 41. We know from Proposition $38_{\frown}$ and Corollary 40 that $\mathscr{D}_G^{\text{irr}} := \text{posi}(\mathscr{A}_G^{\text{irr}})$ is a coherent set of desirable gambles on $\mathscr{X}_G$ that satisfies Requirements $(\text{N1A})_{131}$ and $(\text{N1B})_{132}$. Because of Corollary 35_{136}, it is furthermore the smallest coherent set of desirable gambles on $\mathscr{X}_G$ for which this is the case. □

By carefully going through the proof of Theorem 41$_{\frown}$ and the proofs of the results that it uses, one can furthermore see that if we were to replace Requirement (N1B)$_{132}$ with a weaker version that only imposes epistemic value-irrelevance, the resulting irrelevant natural extension would be the same set $\mathscr{D}_G^{\text{irr}}$ and would therefore still end up satisfying the strong version of Requirement (N1B)$_{132}$ that we are currently using. This is a nice example of how imposing epistemic value-irrelevance can lead to epistemic (h-)irrelevance being satisfied 'for free'.

5.4.2 *For conditional lower previsions*

Within the framework of lower previsions, the global model that corresponds to a credal network is a coherent conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ that satisfies Requirements (N2A)$_{131}$ and (N2B)$_{132}$. If such a model exists—and we will show that it does—then, since coherence and both of these properties are preserved under taking lower envelopes, there is a unique most conservative—pointwise smallest—coherent conditional lower prevision $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ that satisfies them both. We call this unique model the irrelevant natural extension of the local models $\underline{P}_{s\rfloor x_{P(s)}}$. The following results show that this concept is strongly connected with its counterpart for sets of desirable gambles.

Theorem 42. *Let $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot)$ be the irrelevant natural extension of local coherent lower previsions $\underline{P}_{s\rfloor x_{P(s)}}$ and let $\mathscr{D}_G^{\text{irr}}$ be the irrelevant natural extension of the corresponding local coherent sets of desirable gambles $\mathscr{D}_{s\rfloor x_{P(s)}} := \mathscr{D}_{\underline{P}_{s\rfloor x_{P(s)}}}$. Then $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}_G^{\text{irr}}}(\cdot\rfloor\cdot)$.*

Proof of Theorem 42. Since we know that $\mathscr{D}_G^{\text{irr}}$ is coherent and satisfies Requirements (N1A)$_{131}$ and (N1B)$_{132}$ [see Section 5.4.1$_{135}$] it follows that $\underline{P}_{\mathscr{D}_G^{\text{irr}}}(\cdot\rfloor\cdot)$ is also coherent and satisfies Requirements (N2A)$_{131}$ and (N2B)$_{132}$. This implies that $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot) \leq \underline{P}_{\mathscr{D}_G^{\text{irr}}}(\cdot\rfloor\cdot)$.

Now let $\mathscr{D}_G$ be any coherent set of desirable gambles on $\mathscr{X}_G$ such that $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}_G}(\cdot\rfloor\cdot)$ [there is such a set $\mathscr{D}_G$ because $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot)$ is coherent]. Choose any $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$. Since it follows from Requirements (N2A)$_{131}$ and (N2B)$_{132}$ that $\underline{P}_s^{\text{irr}}(f\rfloor x_{PN(s)}) = \underline{P}_s^{\text{irr}}(f\rfloor x_{P(s)}) \geq \underline{P}_{s\rfloor x_{P(s)}}(f)$ for all $f \in \mathscr{G}(\mathscr{X}_s)$, we have that

$$\mathscr{D}_{s\rfloor x_{P(s)}} = \mathscr{D}_{\underline{P}_{s\rfloor x_{P(s)}}} \subseteq \mathscr{D}_{\underline{P}_s^{\text{irr}}(\cdot\rfloor x_{PN(s)})} \subseteq \text{marg}_s(\mathscr{D}_G\rfloor x_{PN(s)}),$$

where the last inclusion holds because it follows from $\underline{P}_G^{\text{irr}}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}_G}(\cdot\rfloor\cdot)$ that the coherent set of desirable gambles $\text{marg}_s(\mathscr{D}_G\rfloor x_{PN(s)})$ has $\underline{P}_s^{\text{irr}}(\cdot\rfloor x_{PN(s)})$ as its lower prevision [see Section 4.2.3$_{105}$] and because $\mathscr{D}_{\underline{P}_s^{\text{irr}}(\cdot\rfloor x_{PN(s)})}$ is the smallest coherent set of desirable gambles for which this holds [see Equation (2.4)$_{41}$].

Hence, we find that $\mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}_G$ for all $f \in \mathscr{D}_{s\rfloor x_{P(s)}}$. Since this holds for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$, we find that $\mathscr{D}_G^{\mathrm{irr}} = \mathrm{posi}(\mathscr{A}_G^{\mathrm{irr}}) \subseteq \mathscr{D}_G$ because $\mathscr{D}_G$ is coherent, which in turn implies that $\underline{P}_{\mathscr{D}_G^{\mathrm{irr}}}(\cdot\rfloor\cdot) \leq \underline{P}_{\mathscr{D}_G}(\cdot\rfloor\cdot) = \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$. □

Proposition 43. *Let $\mathscr{D}_G^{\mathrm{irr}}$ be the irrelevant natural extension of local coherent sets of desirable gambles $\mathscr{D}_{s\rfloor x_{P(s)}}$ and let $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ be the irrelevant natural extension of the corresponding coherent lower previsions $\underline{P}_{s\rfloor x_{P(s)}} := \underline{P}_{\mathscr{D}_{s\rfloor x_{P(s)}}}$. Then $\underline{P}_{\mathscr{D}_G^{\mathrm{irr}}} = \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\mathscr{X}_G)$.*

Proof of Proposition 43. For each $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, consider the local set of desirable gambles $\mathscr{D}^*_{s\rfloor x_{P(s)}} := \mathscr{D}_{\underline{P}_{s\rfloor x_{P(s)}}} \subseteq \mathscr{D}_{s\rfloor x_{P(s)}}$. If we let $\mathscr{D}_G^{\mathrm{irr}*}$ be the irrelevant natural extension of these local models, then $\mathscr{D}_G^{\mathrm{irr}*}$ is clearly a subset of $\mathscr{D}_G^{\mathrm{irr}}$ [by monotonicity of irrelevant natural extension].

Consider any $f \in \mathscr{D}_G^{\mathrm{irr}}$ and $\delta \in \mathbb{R}_{>0}$. Due to Proposition 36_{136}, we know that

$$f = \sum_{s\in G} \sum_{z_{PN(s)}\in\mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}},$$

where, for all $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, the gamble $f_{s,z_{PN(s)}}$ is either zero or an element of $\mathscr{D}_{s\rfloor z_{P(s)}}$, and at least one of them is non-zero. It is now always possible to choose $\varepsilon \in \mathbb{R}_{>0}$ small enough to guarantee that

$$\sum_{s\in G} \sum_{z_{PN(s)}\in\mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}}[f_{s,z_{PN(s)}} + \varepsilon] \leq f + \delta. \tag{5.3}$$

For all $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, it follows from the coherence of $\mathscr{D}_{s\rfloor z_{P(s)}}$ and the fact that $f_{s,z_{PN(s)}} \in \mathscr{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ that

$$\underline{P}_{s\rfloor z_{P(s)}}(f_{s,z_{PN(s)}} + \varepsilon) := \underline{P}_{\mathscr{D}_{s\rfloor z_{P(s)}}}(f_{s,z_{PN(s)}} + \varepsilon) > 0,$$

which implies that $f_{s,z_{PN(s)}} + \varepsilon \in \mathscr{D}^*_{s\rfloor z_{P(s)}}$. Therefore, it follows from Equation (5.3), Proposition 36_{136} and the coherence of $\mathscr{D}_G^{\mathrm{irr}*}$ [see Proposition 38_{137}] that $f + \delta \in \mathscr{D}_G^{\mathrm{irr}*}$. Since this holds for all $\delta \in \mathbb{R}_{>0}$, we find that $f \in \mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}*})$. Since this holds for all $f \in \mathscr{D}_G^{\mathrm{irr}}$, we find that $\mathscr{D}_G^{\mathrm{irr}} \subseteq \mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}*})$, which in turn implies that $\mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}}) \subseteq \mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}*})$. Since we already know that $\mathscr{D}_G^{\mathrm{irr}*}$ is a subset of $\mathscr{D}_G^{\mathrm{irr}}$, this allows us to infer that $\mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}}) = \mathrm{cl}(\mathscr{D}_G^{\mathrm{irr}*})$. The proof now follows from Theorem $42_{\frown}$ and Equation $(2.8)_{41}$. □

Similarly to what we found for sets of desirable gambles, a closer look at the proof of Theorem $42_{\frown}$ reveals that replacing Requirement $(\mathrm{N2B})_{132}$ with a weaker version that only imposes epistemic value-irrelevance would result in the exact same notion of irrelevant natural extension.

The importance of Theorem $42_{\frown}$ is that it allows us to—often trivially—translate properties that we have proved—or will prove—for $\mathscr{D}_G^{\mathrm{irr}}$ into properties of $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$. As a first example, we translate Corollary 40_{138} into the language of lower previsions.

Corollary 44. *For all $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$, it holds that:*

$$\underline{P}^{\text{irr}}_s(\cdot\rfloor\{x_{P(s)}\} \times B_{N(s)}) = \underline{P}^{\text{irr}}_s(\cdot\rfloor x_{P(s)}) = \underline{P}_{s\rfloor x_{P(s)}}$$

and

$$\underline{P}^{\text{irr}}_s(\cdot\rfloor \cdot \times \{x_{P(s)}\} \times B_{N(s)}) = \underline{P}^{\text{irr}}_s(\cdot\rfloor \cdot \times \{x_{P(s)}\}).$$

Proof of Corollary 44. Immediate consequence of Theorem 42_{139} and Corollary 40_{138}. □

The second expression in this result is to be expected, as it simply means that $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$ satisfies Requirement $(\text{N2B})_{132}$. The first expression implies a strengthened version of Requirement $(\text{N2A})_{131}$.

We end this section by showing that the irrelevant natural extension of the local models is equal to the 'normal' natural extension of a set of global assessments, obtained by extending the local assessments using epistemic irrelevance.

Proposition 45. *Consider a conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ with domain $\mathscr{C} = \{(f, x_{PN(s)}) \colon s \in G, f \in \mathscr{G}(\mathscr{X}_s), x_{PN(s)} \in \mathscr{X}_{PN(s)}\}$, defined by*

$$\underline{P}_s(f\rfloor x_{PN(s)}) := \underline{P}_{s\rfloor x_{P(s)}}(f) \text{ for all } s \in G,\ f \in \mathscr{G}(\mathscr{X}_s) \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}.$$

Let $\underline{E}_G(\cdot\rfloor\cdot)$ be the natural extension of $\underline{P}_G(\cdot\rfloor\cdot)$. Then $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot) = \underline{E}_G(\cdot\rfloor\cdot)$.

Proof of Proposition 45. Since $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$ is coherent and, by Corollary 44, coincides with $\underline{P}_G(\cdot\rfloor\cdot)$ on its domain, Proposition 3_{45} implies that $\mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$ is coherent, which in turn implies that $\underline{E}_G(\cdot\rfloor\cdot)$ exists and is coherent and, by Proposition 2_{45}, that $\underline{E}_G(\cdot\rfloor\cdot) \leq \underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$.

For any $s \in G$ and any $x_{PN(s)} \in \mathscr{X}_{PN(s)}$, Equation $(2.12)_{44}$ now implies that $\underline{E}_s(f\rfloor x_{PN(s)}) \geq \underline{P}_s(f\rfloor x_{PN(s)}) = \underline{P}_{s\rfloor x_{P(s)}}(f)$ for all $f \in \mathscr{G}(\mathscr{X}_s)$. Hence, since $\underline{E}_G(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}}(\cdot\rfloor\cdot)$, an argument that is completely analogous to the one that was given in the second part of the proof of Theorem 42_{139} leads us to conclude that $\mathscr{D}^{\text{irr}}_G \subseteq \mathscr{E}_{\underline{P}_G(\cdot\rfloor\cdot)}$ and therefore, that $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot) \leq \underline{E}_G(\cdot\rfloor\cdot)$. □

5.4.3 *For sets of conditional linear previsions*

If we use the framework of sets of conditional linear previsions, the global model that corresponds to a credal network under epistemic irrelevance is a set $\mathscr{M}_G$ of conditional linear previsions $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ that satisfies Requirements $(\text{N3A})_{131}$ and $(\text{N3B})_{132}$. The largest—most conservative—such set is denoted by $\mathscr{M}^{\text{irr}}_G$ and is called the irrelevant natural extension of the local models $\mathscr{M}_{s\rfloor x_{P(s)}}$. The following result establishes that this concept is equivalent to the corresponding notion for lower previsions.

Theorem 46. *Let $\mathscr{M}_G^{\mathrm{irr}}$ be the irrelevant natural extension of the local closed and convex sets of linear previsions $\mathscr{M}_{s\rfloor x_{P(s)}}$ and let $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ be the irrelevant natural extension of the corresponding local coherent lower previsions $\underline{P}_{s\rfloor x_{P(s)}} := \underline{P}_{\mathscr{M}_{s\rfloor x_{P(s)}}}$. Then $\mathscr{M}_G^{\mathrm{irr}} = \mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)}$.*

Proof of Theorem 46. Since $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is coherent and satisfies Requirement (N2B)$_{132}$, we know from the discussion in Section 4.3.4$_{119}$ that $\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)}$ satisfies Requirement (N3B)$_{132}$. It also satisfies Requirement (N3A)$_{131}$, because

$$\mathrm{marg}_s(\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)}\rfloor x_{P(s)}) = \mathscr{M}_{\underline{P}_s^{\mathrm{irr}}(\cdot\rfloor x_{P(s)})} \subseteq \mathscr{M}_{\underline{P}_{s\rfloor x_{P(s)}}} = \mathscr{M}_{s\rfloor x_{P(s)}},$$

where the first equality is a consequence of Equation (4.11)$_{109}$, the inclusion follows from the fact that $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ satisfies Requirement (N2A)$_{131}$ and the last equality holds because of the one-to-one correspondence between unconditional coherent lower previsions and closed convex sets of unconditional linear previsions. Since $\mathscr{M}_G^{\mathrm{irr}}$ is by definition the largest set of conditional linear prevision on $\mathscr{C}(\mathscr{X}_G)$ for which this is the case, this implies that $\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)} \subseteq \mathscr{M}_G^{\mathrm{irr}}$.

Conversely, since $\mathscr{M}_G^{\mathrm{irr}}$ satisfies Requirements (N3A)$_{131}$ and (N3B)$_{132}$, its lower envelope $\underline{P}_{\mathscr{M}_G^{\mathrm{irr}}}(\cdot\rfloor\cdot)$ satisfies Requirements (N2A)$_{131}$ and (N2B)$_{132}$. Since $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is the pointwise smallest coherent conditional lower prevision on $\mathscr{C}(\mathscr{X}_G)$ for which this is the case, this implies that $\underline{P}_{\mathscr{M}_G^{\mathrm{irr}}}(\cdot\rfloor\cdot) \geq \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$, which in turn implies that $\mathscr{M}_G^{\mathrm{irr}} \subseteq \mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)}$. □

Using this connection, we can translate—current and future—results for $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ into analogous results for $\mathscr{M}_G^{\mathrm{irr}}$. The following intuitive characterisation of $\mathscr{M}_G^{\mathrm{irr}}$ is a first example, as it is basically a translation of Proposition 45$_{\frown}$ to the framework of sets of conditional linear previsions.

Proposition 47. *A coherent conditional linear prevision $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ belongs to $\mathscr{M}_G^{\mathrm{irr}}$ if and only if*

$$P_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{M}_{s\rfloor x_{P(s)}} \text{ for all } s \in G \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}. \tag{5.4}$$

Proof of Proposition 47. We define, for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$, $\underline{P}_{s\rfloor x_{P(s)}} := \underline{P}_{\mathscr{M}_{s\rfloor x_{P(s)}}}$. Since $\mathscr{M}_{s\rfloor x_{P(s)}}$ is assumed to be closed and convex, we know that $\mathscr{M}_{s\rfloor x_{P(s)}} = \mathscr{M}_{\underline{P}_{s\rfloor x_{P(s)}}}$, which implies that Equation (5.4) is equivalent to

$$P_s(\cdot\rfloor x_{PN(s)}) \geq \underline{P}_{s\rfloor x_{P(s)}} \text{ for all } s \in G \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}. \tag{5.5}$$

Consider now any coherent conditional linear prevision $P_G(\cdot\rfloor\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$. We need to prove that $P_G(\cdot\rfloor\cdot)$ satisfies Equation (5.5) if and only if $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_G^{\mathrm{irr}}$.

First assume that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_G^{\mathrm{irr}}$. By applying Theorem 46, we find that $P_G(\cdot\rfloor\cdot) \geq \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$. Due to Corollary 44$_{\frown}$, this implies that $P_G(\cdot\rfloor\cdot)$ satisfies Equation (5.5).

Next, assume that $P_G(\cdot\rfloor\cdot)$ satisfies Equation $(5.5)_{\frown}$. Let $\underline{P}_G(\cdot\rfloor\cdot)$ and $\underline{E}_G(\cdot\rfloor\cdot)$ be defined as in Proposition 45_{141}. Equation $(5.5)_{\frown}$ implies that $P_G(\cdot\rfloor\cdot)$ dominates $\underline{P}_G(\cdot\rfloor\cdot)$ and therefore, since $\underline{P}_G(\cdot\rfloor\cdot)$ is coherent, that $P_G(\cdot\rfloor\cdot)$ dominates $\underline{E}_G(\cdot\rfloor\cdot)$, which means that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{E}_G(\cdot\rfloor\cdot)}$ or, equivalently, by Proposition 45_{141}, that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)}$. By applying Theorem $46_{\frown}$, we find that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}^{\text{irr}}_G$. □

Our next result is the sets of linear previsions analogue to Corollary 40_{138}. It states—as is to be expected—that $\mathscr{M}^{\text{irr}}_G$ satisfies Requirement $(\text{N3B})_{132}$ and a strengthened version of Requirement $(\text{N3A})_{131}$.

Corollary 48. *For all $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$, it holds that:*

$$\text{marg}_s(\mathscr{M}^{\text{irr}}_G \rfloor \{x_{P(s)}\} \times B_{N(s)}) = \text{marg}_s(\mathscr{M}^{\text{irr}}_G \rfloor x_{P(s)}) = \mathscr{M}_{s\rfloor x_{P(s)}}$$

and

$$\text{marg}^{\text{c}}_s(\mathscr{M}^{\text{irr}}_G \rfloor | \{x_{P(s)}\} \times B_{N(s)}) = \text{marg}^{\text{c}}_s(\mathscr{M}^{\text{irr}}_G \rfloor | x_{P(s)}).$$

Proof of Corollary 48. Immediate consequence of Theorem $46_{\frown}$ and Corollary 44_{141}. □

5.4.4 *For sets of full conditional probability mass functions*

Within the framework of probability mass functions, the global model that corresponds to a credal network under epistemic irrelevance is a set $\mathscr{F}_G$ of full conditional probability mass functions on $\mathscr{C}_*(\Omega)$ that satisfies Requirements $(\text{N4A})_{131}$ and $(\text{N4B})_{132}$. The largest such set is called the irrelevant natural extension of the local credal sets $\mathscr{F}_{s\rfloor x_{P(s)}}$. It is connected to the corresponding notion for sets of conditional linear previsions in the following trivial way.

Proposition 49. *Let $\mathscr{F}^{\text{irr}}_G$ be the irrelevant natural extension of the credal sets $\mathscr{F}_{s\rfloor x_{P(s)}}$ and let $\mathscr{M}^{\text{irr}}_G$ be the irrelevant natural extension of the corresponding local sets of linear previsions $\mathscr{M}_{\mathscr{F}_{s\rfloor x_{P(s)}}}$. Then $\mathscr{F}^{\text{irr}}_G = \mathscr{F}_{\mathscr{M}^{\text{irr}}_G}$.*

Proof of Proposition 49. This is an immediate consequence of the one-to-one correspondence between (conditional) linear previsions and (full conditional) probability mass functions. □

The following two results are direct translations of Proposition $47_{\frown}$ and Corollary 48 to the language of sets of full conditional probability mass functions.

Corollary 50. *A full conditional probability mass function $p_G(\cdot\rfloor\cdot)$ on $\mathscr{C}_*(\mathscr{X}_G)$ belongs to $\mathscr{F}^{\text{irr}}_G$ if and only if*

$$p_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{F}_{s\rfloor x_{P(s)}} \text{ for all } s \in G \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}.$$

Proof of Corollary 50$_{\frown}$. Immediate consequence of Proposition 49$_{\frown}$ and Proposition 47$_{142}$. □

Corollary 51. *For all $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$, it holds that:*

$$\operatorname{marg}_s(\mathscr{F}_G^{\mathrm{irr}} \rfloor \{x_{P(s)}\} \times B_{N(s)}) = \operatorname{marg}_s(\mathscr{F}_G^{\mathrm{irr}} \rfloor x_{P(s)}) = \mathscr{F}_{s \rfloor x_{P(s)}}$$

and

$$\operatorname{marg}_s^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | \{x_{P(s)}\} \times B_{N(s)}) = \operatorname{marg}_s^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | x_{P(s)}).$$

Proof of Corollary 51. Immediate consequence of Proposition 49$_{\frown}$ and Corollary 48$_{\frown}$. □

5.4.5 *Bayesian networks as a special case*

Let us now focus on an important special case. Consider the framework of sets of full conditional probability mass functions and, for all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, let $\mathscr{F}_{s \rfloor x_{P(s)}} := \{p_{s \rfloor x_{P(s)}}\}$, where $p_{s \rfloor x_{P(s)}}$ is a probability mass function on $\mathscr{X}_s$. It then follows from Corollary 50$_{\frown}$ that a full conditional probability mass function $p_G(\cdot \rfloor \cdot)$ on $\mathscr{C}_*(\mathscr{X}_G)$ belongs to $\mathscr{F}_G^{\mathrm{irr}}$ if and only if

$$p_s(x_s \rfloor x_{PN(s)}) = p_{s \rfloor x_{P(s)}}(x_s) \text{ for all } s \in G, x_{PN(s)} \in \mathscr{X}_{PN(s)} \text{ and } x_s \in \mathscr{X}_s. \quad (5.6)$$

For readers who are familiar with the theory of Bayesian networks [64, 78, 82], this requirement should ring a bell. It is well-known to imply that

$$p_G(x_G) = \prod_{s \in G} p_s(x_s \rfloor x_{P(s)}) = \prod_{s \in G} p_{s \rfloor x_{P(s)}}(x_s) \text{ for all } x_G \in \mathscr{X}_G, \quad (5.7)$$

which is the celebrated factorised form of the joint probability mass function for a Bayesian network. In other words, in this particular case, the unconditional part $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ of $\mathscr{F}_G^{\mathrm{irr}}$ consists of a single probability mass function on $\mathscr{X}_G$, which is identical to the joint probability mass function of a Bayesian network with local models $p_{s \rfloor x_{P(s)}}$, as given by Equation (5.7). For this reason, the theory of credal networks under epistemic irrelevance can rightfully be referred to as a generalisation of the theory of Bayesian networks.

The main generalisation consists in introducing imprecision, but it is not the only one. In the particular case that we are discussing here, where the local models are precise, our approach is still more general than that of Bayesian networks, because we regard conditional models as primitive notions. For example, in our framework, for any $p_G(\cdot \rfloor \cdot) \in \mathscr{F}_G^{\mathrm{irr}}$, the local model $p_{s \rfloor x_{P(s)}}$ can always be recovered from the global model $p_G(\cdot \rfloor \cdot)$: it is equal to $p_s(\cdot \rfloor x_{P(s)})$. For Bayesian networks, because they do not consider full conditional probability mass functions, this is only possible if $p_{P(s)}(x_{P(s)}) > 0$. More generally, for every $B_G \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$, $p_G(\cdot \rfloor \cdot)$ provides us with a conditional probability mass

function $p_G(\cdot\rfloor B_G)$. For Bayesian networks, this is only the case if $p_G(B_G) > 0$. Of course, some of this extra conditional information may not be fully determined by Equation (5.6)$_{\cap}$. If that is the case, then $\mathscr{F}_G^{\text{irr}}$ will consist of multiple full conditional probability mass functions. Nevertheless, as discussed above, they will all have the same unconditional part. In the framework of Bayesian networks, this unconditional part is taken to constitute the complete model, whereas we consider the—possibly imprecise—conditional information to be part of the model as well.

5.5 OTHER APPROACHES AND HOW THEY ARE RELATED

Bayesian networks can be generalised to allow for imprecision in many different ways. The focus of this chapter has been on one specific approach—credal networks under epistemic irrelevance—and, in particular, on the corresponding irrelevant natural extension. In the next two chapters, we present a detailed study of the properties of this irrelevant natural extension and we show how these properties can be used to develop algorithms. However, before we do so, we now take a step back and compare our approach to other types of credal networks.

Let us start with the name 'credal network', which, according to our findings, first appeared in Reference [55], where it was used to refer to what we now prefer to call the complete extension of a credal network, as discussed below. Cozman [16, 17] extended the scope of this terminology and used it as a generic name for structures that associate convex sets of probability measures with DAGs or, in other words, to refer to generalisations of Bayesian networks that use convex sets of probabilities. Today, the theory of credal networks no longer restricts attention to sets of probabilities but considers other imprecise-probabilistic frameworks as well, such as sets of desirable gambles [31,76] and lower previsions [27, 42].

The most straightforward way to develop an imprecise-probabilistic version of a Bayesian network is to consider a set of Bayesian networks instead of a single one. The elements of such a set have the same graphical structure but their local conditional probabilities may differ. Equivalently, we consider a set $\mathscr{F}_G$ of joint probability mass functions on $\mathscr{F}_G$ such that, for every $s \in G$, $X_{PN(s)}$ is completely independent from X_s conditional on $X_{P(s)}$; see Section 4.4. For obvious reasons, we call this a credal network under complete independence. Starting from local credal sets $\mathscr{F}_{s\rfloor x_{P(s)}}$, the corresponding largest set of joint probability mass functions on $\mathscr{X}_G$ is given by [55]

$$\mathscr{F}_G^{\text{com}} := \left\{ \prod_{s\in G} p_{s\rfloor X_{P(s)}}(X_s) \colon (\forall s \in G)(\forall x_{P(s)} \in \mathscr{X}_{P(s)}) p_{s\rfloor x_{P(s)}} \in \mathscr{F}_{s\rfloor x_{P(s)}} \right\}.$$

We call it the complete extension of the local sets $\mathscr{F}_{s\rfloor x_{P(s)}}$. This approach is highly intuitive if the ideal of precision is adopted. The idea is then that there

is an underlying true Bayesian network for which, for some reason, the local models are only known to belong to a set of candidates $\mathscr{F}_{s\rfloor x_{P(s)}}$—are partially specified. It seems reasonable to model this type of uncertainty by means of a set of Bayesian models.

However, credal networks under complete independence are almost never considered. Instead, the large majority of results have been developed for credal networks under strong independence; Reference [2] provides a recent overview of related literature, containing numerous references to algorithmic developments, practical applications, and so on. On that approach, $\mathscr{F}_G$ is taken to be closed and convex and it is assumed that, for every $s \in G$, $X_{PN(s)}$ is strongly independent from X_s conditional on $X_{P(s)}$ or, equivalently, that the extreme points of $\mathscr{F}_G$ correspond to Bayesian networks with the same graphical structure. In other words: a credal network under strong independence is the convex hull of a credal network under complete independence. In particular: the strong extension $\mathscr{F}_G^{\text{strong}}$ of a collection of local credal sets $\mathscr{F}_{s\rfloor x_{P(s)}}$ is equal to the convex hull of their complete extension $\mathscr{F}_G^{\text{com}}$.

It is unclear to us why credal networks under strong independence are more popular than credal networks under complete independence, and similarly for the corresponding extensions. We presume that this is due to a desire to keep convexity, which, in many fields, is known to lead to more efficient computations. However, in the case of credal networks, it has already been shown in Reference [55] that this does not lead to computational savings: for many commonly considered parameters of interest—such as posterior lower and upper probabilities—it makes no difference whether we compute them with respect to the complete extension or its convex hull—the strong extension. Hence, in this case, enforcing convexity seems to serve no purpose. Combined with the fact that complete extensions have a clear and intuitive sensitivity analysis interpretation, which strong extensions do not have, we see no reason to keep on favouring the latter approach.

Although the complete and strong extensions are named after the corresponding imprecise-probabilistic notions of independence and—as we have done—can be defined in terms of them, this is usually regarded as being of minor importance. The crucial feature of credal networks under complete and strong independence is that they can be defined in terms of sets of Bayesian networks. Not only does this make these approaches familiar, it also helps in the development of theoretical results and algorithms. The corresponding concepts for Bayesian networks can be used as a starting point, and the ideas behind them can be adapted to an imprecise-probabilistic setting. In fact, much of the research on credal networks has been concerned with doing exactly that, and we believe this to be one of the main reasons for the popularity of strong extensions.

In his seminal paper [16], Cozman approached the topic of credal networks from a different angle. He moved the focus away from sets of Bayesian net-

works and instead suggested that imprecise-probabilistic notions of independence should play a more central role. In our opinion, one of his most important insights was that the generalised notion of natural extension that was discussed in Section 4.5_{124} can be applied in the context of credal networks; our developments in this chapter serve as a nice example of such an approach. Also, instead of restricting attention to imprecise-probabilistic notions of independence that are based on stochastic independence—such as complete and strong independence—Cozman advocated the use of epistemic irrelevance and epistemic independence. This led him to introduce two new types of credal networks: credal networks under epistemic irrelevance and credal networks under epistemic independence [16, 35].

The theory of credal networks under epistemic irrelevance that was put forward by Cozman [16] can be regarded as a precursor to the one that we presented in this chapter. The basic idea is exactly the same: for all $s \in G$, $X_{N(s)}$ is taken to be epistemically irrelevant to X_s conditional on $X_{P(s)}$. These structural assessments are combined with local direct assessments and the unique most conservative model that satisfies them both—the irrelevant natural extension[4]—is then considered. The difference between our approach and that of Cozman is that he only considers the framework of sets of probabilities and that he makes the simplifying assumption that every event has strictly positive lower probability.[5] The main contribution of this chapter consists in dropping these restrictions.[6]

The idea behind credal networks under epistemic independence is similar. The only difference is that epistemic irrelevance is replaced by epistemic independence: for all $s \in G$, $X_{N(s)}$ is taken to be epistemically independent to X_s conditional on $X_{P(s)}$. The most conservative global model that satisfies these independence assessments and that is furthermore compatible with a given collection of local models is called the independent natural extension.[7,8] A prob-

[4]Cozman simply refers to it as the natural extension; we prefer to distinguish it from other types of natural extension by adding the prefix 'irrelevant'.

[5]Another difference is that he uses epistemic value-irrelevance instead of epistemic h-irrelevance. However, as far as the definition of the irrelevant natural extension is concerned, this makes no difference; see the comments after Proposition 43_{140} and Theorem 42_{139}. The main advantage of epistemic h-irrelevance—besides its philosophical superiority—lies in the fact that it allows us to state more powerful properties for this irrelevant natural extension; see Chapter 6_{152}.

[6]Moral [76] already considered the framework of sets of desirable gambles. However, he uses a more stringent notion of epistemic irrelevance than we do. As a result, his theory can only be applied to a restricted subset of DAGs; see Section $6.5.4_{170}$ for more information. For the special case of credal networks under epistemic irrelevance that have a tree topology, the framework of lower previsions was already considered in Reference [42].

[7]Reference [16] refers to it as the natural extension. We prefer to distinguish it from other types of natural extension by adding the prefix 'independent'; see Footnote 4 as well. Reference [35] calls it 'the extension based on epistemic independence' or 'the epistemic extension'.

[8]This name is also used to refer to the most conservative so-called independent product of a

lematic feature of this independent natural extension is that, in general, it has no—known—closed-form expression. We believe this to be one of the main reasons why credal networks under epistemic independence have received almost no attention; to the best of our knowledge, References [16] and [35] are the only exceptions.

Credal networks under epistemic irrelevance have received considerably more attention [6,16,27,28,30,31,35,42,67,76], and—in our opinion—rightly so. On the one hand, from a philosophical point of view, credal networks under epistemic irrelevance share the advantages of epistemic irrelevance that we discussed in Section 4.4_{120}. Most importantly: the irrelevant natural extension has a clear and intuitive definition that does not require an assumption of ideal precision. On the other hand, from an algorithmic point of view, initial results indicate that the irrelevant natural extension allows for efficient computations. If the topology of the network is a tree, there is a polynomial-time updating algorithm that can compute posterior beliefs about a single target variable conditional on the observation of others [42]. We believe this to be promising, especially since the same inference problem is NP-hard for credal trees under strong independence [67]. Other promising algorithmic developments have also been made—in part by us—for the special case of imprecise hidden Markov models under epistemic irrelevance [6, 30].

Despite these advantages and promising algorithmic developments, the majority of research on credal networks is still concerned with credal networks under strong independence. This is nicely illustrated by the fact that a recent overview paper [2] deals almost exclusively with credal networks under strong independence; credal networks under epistemic irrelevance are only mentioned in passing. Nevertheless, they are regarded as a promising new subfield that requires—and deserves—further research [2, Section 10.6]. One of the main persisting problems is that—except for networks that are sufficiently small or have a tree topology—no efficient, exact or even approximate inference algorithm is available for the irrelevant natural extension. We believe that this is to a great extent due to a profound lack of known theoretical properties. In the next chapter, we start to remedy this situation by conducting a thorough theoretical study of the properties of the irrelevant natural extension.

5.A PROOF OF PROPOSITIONS 38 AND 39

Our proof for Proposition 38_{137} (and Theorem 53_{156}) uses the following convenient version of the separating hyperplane theorem. This result has been

collection of local models; see Section 6.6_{171} and Footnote 19_{171}. Given that this most conservative independent product is equal to the independent natural extension of a credal network under epistemic independence—or epistemic irrelevance—that consists of disconnected nodes, there is no clash in terminology.

proved in Reference [45, Lemma 2]; we repeat its statement here to make this dissertation more self-contained.

Lemma 52. *Consider a finite subset $\mathscr{A}$ of $\mathscr{G}(\mathscr{X})$. Then $0 \notin \mathscr{E}(\mathscr{A})$ if and only if there is a probability mass function p on $\mathscr{X}$ such that, for all $f \in \mathscr{A}$, $P(f) := \sum_{x \in \mathscr{X}} p(x) f(x) > 0$ and, for all $x \in \mathscr{X}$, $p(x) > 0$.*

Proof of Proposition 38$_{137}$. Since, by Proposition 37$_{137}$, $\mathscr{D}_G^{\mathrm{irr}} = \mathscr{E}(\mathscr{A}_G^{\mathrm{irr}})$, we know that $\mathscr{D}_G^{\mathrm{irr}}$ is coherent if and only if it satisfies D1, which states that 0 cannot be an element of $\mathscr{D}_G^{\mathrm{irr}}$. So assume *ex absurdo* that $0 \in \mathscr{D}_G^{\mathrm{irr}}$. We will show that this leads to a contradiction.

Since $0 \in \mathscr{D}_G^{\mathrm{irr}}$, we know from Proposition 36$_{136}$ that

$$0 = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}}, \tag{5.8}$$

where every $f_{s,z_{PN(s)}}$ is an element of $\mathscr{D}_{s \rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero. We now construct, for every $s \in G$ and $z_{P(s)} \in \mathscr{X}_{P(s)}$, a finite subset of the local model $\mathscr{D}_{s \rfloor z_{P(s)}}$:

$$\mathscr{A}^0_{s \rfloor z_{P(s)}} := \left\{ f_{s, x_{PN(s)}} : x_{PN(s)} \in \mathscr{X}_{PN(s)},\ x_{P(s)} = z_{P(s)} \text{ and } f_{s, x_{PN(s)}} \neq 0 \right\}.$$

Since $\mathscr{D}_{s \rfloor z_{P(s)}}$ is coherent, we have that $0 \notin \mathscr{D}_{s \rfloor z_{P(s)}} = \mathrm{posi}(\mathscr{D}_{s \rfloor z_{P(s)}})$. This in turn implies that $0 \notin \mathrm{posi}(\mathscr{A}^0_{s \rfloor z_{P(s)}} \cup \mathscr{G}(\mathscr{X}_s)_{>0}) =: \mathscr{E}(\mathscr{A}^0_{s \rfloor z_{P(s)}})$, because both $\mathscr{A}^0_{s \rfloor z_{P(s)}}$ and $\mathscr{G}(\mathscr{X}_s)_{>0}$ are subsets of $\mathscr{D}_{s \rfloor z_{P(s)}}$, and we can therefore apply Lemma 52. This yields, for every $s \in G$ and $z_{P(s)} \in \mathscr{X}_{P(s)}$, a probability mass function $p_s(\cdot \rfloor z_{P(s)})$ on $\mathscr{X}_s$ with associated linear prevision $P_s(\cdot \rfloor z_{P(s)})$ on $\mathscr{G}(\mathscr{X}_s)$ such that $p_s(z_s \rfloor z_{P(s)}) > 0$ for all $z_s \in \mathscr{X}_s$ and $P_s(g \rfloor z_{P(s)}) > 0$ for every $g \in \mathscr{A}^0_{s \rfloor z_{P(s)}}$.

The trick is now to create a Bayesian network that has the conditional probability mass functions $p_s(\cdot \rfloor z_{P(s)})$ as its local models and has the same graphical structure as our credal network under epistemic irrelevance. Let p_G be the global probability probability mass function of this Bayesian network and let P_G be the corresponding linear prevision. We then find that

$$\begin{aligned}
P_G(f) &= \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} P_G\big(\mathbb{I}_{\{z_{PN(s)}\}} f_{s, z_{PN(s)}}\big) \\
&= \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} P_G(z_{PN(s)}) P_s\big(f_{s, z_{PN(s)}} \rfloor z_{PN(s)}\big) \\
&= \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} P_G(z_{PN(s)}) P_s\big(f_{s, z_{PN(s)}} \rfloor z_{P(s)}\big),
\end{aligned}$$

where we have applied Bayes's rule and the conditional independencies encoded in the graph. Since all the local probabilities $p_s(\cdot \rfloor z_{P(s)})$ are strictly

positive, this is also true for the global ones and we therefore find that $P_G(z_{PN(s)}) > 0$. For the previsions $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)})$, there are two possibilities. The first possibility is that $f_{s,z_{PN(s)}} = 0$, in which case $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)}) = 0$. The second possibility is that $f_{s,z_{PN(s)}} \in \mathscr{A}^0_{s \rfloor z_{P(s)}}$, in which case $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)}) > 0$. Since at least one of the gambles $f_{s,z_{PN(s)}}$ in Equation $(5.8)_\frown$ has to be non-zero, it is not possible that $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)}) = 0$ for all gambles $f_{s,z_{PN(s)}}$ and we can therefore conclude that $P_G(0) > 0$. However, the coherence of P_G implies that $P_G(0) = 0$, a contradiction. □

Since Proposition 55_{157} generalises Proposition 39_{138} without building upon it, it is not necessary to provide Proposition 39_{138} with a separate proof. However, we feel that the complexity of the proof for Theorem 53_{156} (which is essential for the proof of Corollary 54_{156} and therefore also for the proof of Proposition 55_{157}) obscures the ease with which Proposition 39_{138} can be proved. We therefore choose to provide Proposition 39_{138} with a proof of its own. As it makes use of so-called maximal sets of desirable gambles, a concept that has not been introduced yet, we provide a short introduction here.

A coherent set $\mathscr{D}$ of desirable gambles on $\mathscr{X}$ is called *maximal* if it is not included in any other coherent set of desirable gambles on $\mathscr{X}$—in other words, if adding any gamble f to $\mathscr{D}$ makes sure we can no longer extend the resulting set $\mathscr{D} \cup \{f\}$ to a coherent set of desirable gambles. Maximal sets of desirable gambles have a number of useful properties. For example, a coherent set $\mathscr{D}$ of desirable gambles on $\mathscr{X}$ is always the intersection of all the maximal coherent sets $\mathscr{D}^*$ of desirable gambles on $\mathscr{X}$ that include it; see Reference [47]. In other words, $f \in \mathscr{D}$ if and only if $f \in \mathscr{D}^*$ for every maximal coherent set $\mathscr{D}^* \supseteq \mathscr{D}$. As a consequence, we have the following separation property: if a gamble $f \in \mathscr{G}(\mathscr{X})$ is not an element of $\mathscr{D}$, there is at least one maximal coherent set $\mathscr{D}^* \supseteq \mathscr{D}$ for which $f \notin \mathscr{D}^*$. Another useful property is that maximal sets of desirable gambles resolve points: for any maximal coherent set $\mathscr{D}$ and any non-zero gamble f in $\mathscr{G}(\mathscr{X})$, either f or $-f$ is an element of $\mathscr{D}$; see Reference [14].

Proof of Proposition 39$_{138}$. Fix any $s \in G$, $g \in \mathscr{G}(\mathscr{X}_{N(s)})_{>0}$, $f \in \mathscr{G}(\mathscr{X}_s)$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$.

We begin by proving that $f \in \mathscr{D}_{s \rfloor x_{P(s)}}$ implies $g\mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}^{\text{irr}}_G$. As explained in the main text of Section $5.4.1_{135}$ and in the proof of Proposition 34_{135}, it holds for any $x_{N(s)} \in \mathscr{X}_{N(s)}$ that $\mathbb{I}_{\{x_{PN(s)}\}} f$ is an element of $\mathscr{D}^{\text{irr}}_G$. Hence, since $g = \sum_{x_{N(s)} \in \mathscr{X}_{N(s)}} g(x_{N(s)}) \mathbb{I}_{\{x_{N(s)}\}}$, we get that

$$g\mathbb{I}_{\{x_{P(s)}\}} f = \sum_{x_{N(s)} \in \mathscr{X}_{N(s)}} g(x_{N(s)}) \mathbb{I}_{\{x_{PN(s)}\}} f$$

is a finite strictly positive linear combination of elements of $\mathscr{D}^{\text{irr}}_G$. Hence, since $\mathscr{D}^{\text{irr}}_G$ is coherent [see Proposition 38_{137}] we find that $g\mathbb{I}_{\{x_{P(s)}\}} f \in \mathscr{D}^{\text{irr}}_G$.

Next we prove that $f \notin \mathscr{D}_{s\rfloor x_{P(s)}}$ implies $g\mathbb{I}_{\{x_{P(s)}\}}f \notin \mathscr{D}_G^{\mathrm{irr}}$. The case $f = 0$ is trivial because $g\mathbb{I}_{\{x_{P(s)}\}}f$ is then equal to zero, which cannot be an element of $\mathscr{D}_G^{\mathrm{irr}}$ due to its coherence; see Proposition 38_{137}. If $f \neq 0$, we start by applying some of the properties of maximal coherent sets of desirable gambles that were introduced in the text preceding this proof. Due to the first property, we can infer from $f \notin \mathscr{D}_{s\rfloor x_{P(s)}}$ that there is at least one maximal set of desirable gambles $\mathscr{D}^*_{s\rfloor x_{P(s)}} \supseteq \mathscr{D}_{s\rfloor x_{P(s)}}$ for which $f \notin \mathscr{D}^*_{s\rfloor x_{P(s)}}$. Due to the second property and the fact that $f \neq 0$, this in turn implies that $-f \in \mathscr{D}^*_{s\rfloor x_{P(s)}}$. We will now denote by $\mathscr{D}_G^{\mathrm{irr}*}$ the set that is obtained by Equation $(5.2)_{136}$ if we replace the local model $\mathscr{D}_{s\rfloor x_{P(s)}}$ by the specific maximal superset $\mathscr{D}^*_{s\rfloor x_{P(s)}}$. It should be clear that $\mathscr{D}_G^{\mathrm{irr}*} \supseteq \mathscr{D}_G^{\mathrm{irr}}$. Next, since $-f \in \mathscr{D}^*_{s\rfloor x_{P(s)}}$, it follows from a similar argument as the one that was used in the first part of this proof that $g\mathbb{I}_{\{x_{P(s)}\}}(-f) \in \mathscr{D}_G^{\mathrm{irr}*}$. Hence, due to the coherence of $\mathscr{D}_G^{\mathrm{irr}*}$, $g\mathbb{I}_{\{x_{P(s)}\}}f \notin \mathscr{D}_G^{\mathrm{irr}*}$ and therefore, since $\mathscr{D}_G^{\mathrm{irr}*} \supseteq \mathscr{D}_G^{\mathrm{irr}}$, we find that $g\mathbb{I}_{\{x_{P(s)}\}}f \notin \mathscr{D}_G^{\mathrm{irr}}$. $\square$

6

THEORETICAL PROPERTIES

"One cannot really argue with a mathematical theorem."

Stephen Hawking

Some of the results in the previous chapter already illustrate that the properties of the irrelevant natural extension are not limited to the ones that are needed and/or used to define it; see for example Proposition 39_{138} and Corollaries 40_{138} and 44_{141}. However, so far, we have merely scratched the surface of what can be done. As we are about to show, the irrelevant natural extension satisfies numerous other, and often surprisingly strong, theoretical properties.

Our main technical achievement is a separating hyperplane result. As we will see, it can be used to establish various connections between the irrelevant natural extension of a network and those of its subnetworks, including marginalisation, factorisation and external additivity properties. Another important consequence is an analogon of the classical result for Bayesian networks that d-separation implies independence. In our case, the symmetric notion of d-separation is replaced by an asymmetric version, called AD-separation, and epistemic irrelevance takes the place of independence. We also establish connections between the irrelevant natural extension and the notions of marginal extension and independent natural extension and discuss some properties of the updated models that result from applying regular extension to the irrelevant natural extension. Published versions of some of this material, be it in a less developed and/or general form, can be found in Reference [31].

6.1 CLOSED SETS AND THEIR PROPERTIES

The different attributes of nodes $s \in G$ that were introduced in Section 5.1$_{129}$, such as its parents $P(s)$, its descendants $D(s)$, etcetera, can also be defined for subsets of G. For any subset K of G, we define its set of parents as $P(K) := (\bigcup_{s \in K} P(s)) \setminus K$ and its set of children as $C(K) := (\bigcup_{s \in G} C(s)) \setminus K$. We also let $D(K) := (\bigcup_{s \in K} D(s)) \setminus K$ be its set of descendants and define its set of ancestors as $A(K) := (\bigcup_{s \in K} A(s)) \setminus K$. The non-parent non-descendants of K are given by[1] $N(K) := G \setminus (P(K) \cup K \cup D(K)) = \bigcap_{s \in K} N(s)$ and we also define $PN(K) := P(K) \cup N(K)$.

In general, $PN(K)$ cannot be referred to as the non-descendants of K since $P(K)$ and $D(K)$ are not necessarily disjoint. We call those subsets of G for which they are disjoint *closed*: a set $K \subseteq G$ is closed if, for all $s,t \in K$ and any $k \in G$ such that $s \sqsubseteq k \sqsubseteq t$, it holds that $k \in K$. For these closed subsets K of G, we find that[2] $P(K) \cap D(K) = \emptyset$ and therefore that $PN(K) = G \setminus (K \cup D(K))$. Hence, for closed sets $K \subseteq G$, $PN(K)$ can rightfully be referred to as the non-descendants of K.

A set $K \subseteq G$ is called *ancestral* if $A(K) = \emptyset$—if it includes the ancestors of each of its elements—or, equivalently,[3] if $P(K) = \emptyset$—if it includes the parents of each of its elements. An ancestral set is always closed and, for any closed set $K \subseteq G$, $PN(K)$ is an ancestral set.[4]

Example 4. *We illustrate these notions in Figure 6.1$_{\frown}$. It depicts the same DAG as the one that was shown in Figure 5.1$_{130}$, but now focuses on the closed subset $K = \{s_5, s_7, s_9\}$ of G—depicted in boldface. For this subset, we find that $P(K) = \{s_3, s_4\}$, $D(K) = \{s_8, s_{10}\}$ and $N(K) = \{s_1, s_2, s_6\}$, implying that $P(K) \cap D(K) = \emptyset$. The set $PN(K)$—the shaded nodes—is an ancestral set.* $\Diamond$

With a subset K of G, we can also associate a so-called *sub-DAG* of the DAG that is associated with G. The nodes of this sub-DAG are the elements of K and the directed edges of this sub-DAG are those edges in the original DAG that connect elements in K. For a sub-DAG that is associated with some subset K of G, we will use similar definitions as those for the original DAG, adding the subset K as an index. As an example: for all $k \in K$, we denote by $P_K(k)$ the parents of k in the sub-DAG that is associated with the nodes in K. It is not hard to see that[5] $P_K(k) = P(k) \cap K$ and $P(k) \setminus P_K(k) = P(k) \cap P(K)$. Similar definitions can be given for $D_K(k)$, $N_K(k)$, $PN_K(k)$ and $A_K(k)$. In the same way, for subsets S of K, we consider the sets $P_K(S)$, $D_K(S)$, etcetera.

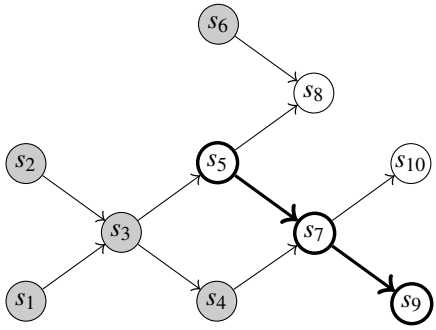
[1] A proof for the last equality can be found in Lemma 77(ii)$_{182}$ in Appendix 6.A$_{182}$.

[2] See Lemma 78(i)$_{183}$ in Appendix 6.A$_{182}$.

[3] See Lemma 77(iv)$_{182}$ in Appendix 6.A$_{182}$.

[4] See Lemma 77(v)$_{182}$ and Lemma 78(iv)$_{183}$ in Appendix 6.A$_{182}$, respectively.

[5] See Lemma 76$_{182}$ in Appendix 6.A$_{182}$.



$$K = \{s_5, s_7, s_9\}$$

$$P(K) = \{s_3, s_4\}$$
$$D(K) = \{s_8, s_{10}\}$$
$$N(K) = \{s_1, s_2, s_6\}$$

$$P_K(s_7) = \{s_5\}$$
$$D_K(s_7) = \{s_9\}$$
$$N_K(s_7) = \emptyset$$

Figure 6.1: Example of a sub-DAG of the DAG in Figure 5.1$_{130}$.

Finally, $Ro(K) := \{s \in K \colon P_K(s) = \emptyset\}$ is the set of root nodes of the sub-DAG that corresponds to K.

Example 5. *Consider again the DAG in Figure 6.1 and consider the sub-DAG that is associated with $K = \{s_5, s_7, s_9\} \subset G$—depicted in boldface. We find that $P_K(s_7) = \{s_5\}$, $D_K(s_7) = \{s_9\}$ and $N_K(s_7) = \emptyset$.* $\Diamond$

These concepts we have just introduced have many properties, which, if the actual graph is drawn, are often very intuitive. However, in general, for an abstract DAG, it can be rather cumbersome to check them or work with them; Appendix 6.A$_{182}$ gathers some technical lemmas that provide proofs for numerous useful results related to DAGs.

6.2 A FUNDAMENTAL SEPARATING HYPERPLANE RESULT

As explained in the previous section, a subset K of G can be associated with a so-called sub-DAG of the original DAG. Similarly to what we have done for the original DAG, we can consider the irrelevant natural extension of this subnetwork. All that we need to be able to do so is local models. One particular way of providing these local models is to derive them from the ones of the original DAG. We start by explaining how this works within the framework of sets of desirable gambles.

In that case, we need to provide, for every $s \in K$ and $z_{P_K(s)} \in \mathscr{X}_{P_K(s)}$, a local set of desirable gambles $\mathscr{D}_{s\rfloor z_{P_K(s)}}$, whereas the original local models are of the form $\mathscr{D}_{s\rfloor z_{P(s)}}$. Hence, if we want to identify the new local models with the original ones, we need to come up with a method for choosing the value $z_{P(s)\setminus P_K(s)}$ of the remaining variables. We will do this in a specific way: by fixing a value $x_{P(K)} \in \mathscr{X}_{P(K)}$ for the variables that are associated with the parents $P(K)$ of K. Since $P(s) \setminus P_K(s)$ is a subset of $P(K)$, this provides us, for every $s \in K$, with

$$P(s_5) = \{s_3\} \qquad K = \{s_5, s_7, s_9\}$$
$$P_K(s_5) = \emptyset \qquad P(K) = \{s_3, s_4\}$$
$$P(s_5) \setminus P_K(s_5) = \{s_3\}$$

$$P(s_7) = \{s_4, s_5\}$$
$$P_K(s_7) = \{s_5\}$$
$$P(s_7) \setminus P_K(s_7) = \{s_4\}$$

$$P(s_9) = \{s_7\}$$
$$P_K(s_9) = \{s_7\}$$
$$P(s_9) \setminus P_K(s_9) = \emptyset$$

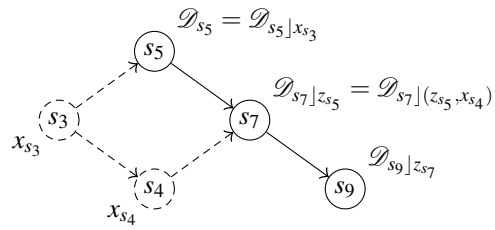


Figure 6.2: Local models for a sub-DAG: an illustration of Equation (6.1).

the value $z_{P(s)\setminus P_K(s)} := x_{P(s)\setminus P_K(s)}$ that was missing, and thereby allows us to identify the local model $\mathscr{D}_{s \rfloor z_{P_K(s)}}$ of the sub-DAG with the local model $\mathscr{D}_{s \rfloor z_{P(s)}}$ of the original DAG, by letting $z_{P(s)} := (z_{P_K(s)}, x_{P(s)\setminus P_K(s)})$.

This approach can be briefly summarised as follows: for any subset K of G and any instantiation $x_{P(K)} \in \mathscr{X}_{P(K)}$ of $X_{P(K)}$, we construct a collection of local models for the sub-DAG that corresponds to K, by defining, for all $s \in K$ and $z_{P_K(s)} \in \mathscr{X}_{P_K(s)}$:

$$\mathscr{D}_{s \rfloor z_{P_K(s)}} = \mathscr{D}_{s \rfloor (z_{P_K(s)}, x_{P(s)\setminus P_K(s)})}. \tag{6.1}$$

The following example illustrates this method for constructing the local models of a sub-DAG.

Example 6. *Consider again the DAG in Figure 6.1$_\frown$ and the sub-DAG that corresponds to the closed subset $K = \{s_5, s_7, s_9\}$ of G. In order to construct a collection of local models for this sub-DAG, we fix some value $x_{P(K)} \in \mathscr{X}_{P(K)}$ for the variables that are associated with the parents of K. Equivalently, since $P(K) = \{s_3, s_4\}$, we fix values $x_{s_3} \in \mathscr{X}_{s_3}$ and $x_{s_4} \in \mathscr{X}_{s_4}$. We can now construct local models by means of Equation* (6.1)*. For the node s_5, we obtain an unconditional local model $\mathscr{D}_{s_5} = \mathscr{D}_{s_5 \rfloor x_{s_3}}$. For the node s_7, this yields, for every $z_{s_5} \in \mathscr{X}_{s_5}$, a conditional local model $\mathscr{D}_{s_7 \rfloor z_{s_5}} = \mathscr{D}_{s_7 \rfloor (z_{s_5}, x_{s_4})}$. Finally, for the node s_9, we obtain, for every $z_{s_7} \in \mathscr{X}_{s_7}$, a conditional local model $\mathscr{D}_{s_9 \rfloor z_{s_7}}$. Figure 6.2 provides a graphical representation of the construction of these local models.* ◊

For every $K \subseteq G$ and $x_{P(K)} \in \mathscr{X}_{P(K)}$, we can now use these local models to construct an irrelevant natural extension for the subnetwork that corresponds

to the set K. By applying Equation $(5.2)_{136}$, we find that the resulting model is given by

$$\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}} := \text{posi}(\mathscr{A}^{\text{irr}}_{K\rfloor x_{P(K)}}), \tag{6.2}$$

with

$$\mathscr{A}^{\text{irr}}_{K\rfloor x_{P(K)}} := \left\{ \mathbb{I}_{\{z_{PN_K(s)}\}} f \colon s \in K, z_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}, f \in \mathscr{D}_{s\rfloor(z_{PN_K(s)}, x_{P(s)\setminus PN_K(s)})} \right\}. \tag{6.3}$$

If K is an ancestral graph—if $P(K) = \emptyset$—then we do not need to fix a value $x_{P(K)} \in \mathscr{X}_{P(K)}$ for $X_{P(K)}$, as it is deterministic. We will then use $\mathscr{D}^{\text{irr}}_K$ and $\mathscr{A}^{\text{irr}}_K$ as shorthand notations for $\mathscr{D}^{\text{irr}}_{K\rfloor x_\emptyset}$ and $\mathscr{A}^{\text{irr}}_{K\rfloor x_\emptyset}$.

A question that now naturally arises is whether the irrelevant natural extension of these subnetworks, as given by Equation (6.2), can be related to the irrelevant natural extension $\mathscr{D}^{\text{irr}}_G$ of the original network. The following separating hyperplane result establishes that, for subsets K of G that are closed, this is indeed the case.

Theorem 53. *Let K be a closed subset of G and consider any $x_{P(K)} \in \mathscr{X}_{P(K)}$ and nonzero $f \in \mathscr{G}(\mathscr{X}_K)$ and $h \in \mathscr{G}(\mathscr{X}_{PN(K)})$ such that $f \notin \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ and $h \notin \mathscr{D}^{\text{irr}}_{PN(K)}$. Then for all $f_* \in \mathscr{D}^{\text{irr}}_G$ and $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$, there is a linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$ such that $P_G(f_*) > 0$, $P_G(h) < 0$ and $P_G(g\mathbb{I}_{\{x_{P(K)}\}} f) < 0$.*

Our proof for this result is long and complicated, and has therefore been moved to Appendix $6.\text{B}_{188}$. Nevertheless, the main idea is very simple. Similarly to what we have done in the proof of Proposition 38_{137}, we construct a joint probability mass function p_G to perform the separation. The corresponding linear prevision P_G characterises a hyperplane $\{f' \in \mathscr{G}(\mathscr{X}_G) \colon P_G(f') = 0\}$, with normal p_G, that separates f_* from h and $g\mathbb{I}_{\{x_{P(K)}\}} f$. However, in contrast with the proof of Proposition 38_{137}, a factorising probability mass function is not sufficient to perform this separation, and this renders constructing p_G a complex and extremely elaborate task.

We consider Theorem 53 to be one of the main technical achievements of this dissertation. Almost all the results in the present chapter can ultimately be traced back to this single theorem. However, in its current form, it is rather technical. Theorem 53 clearly establishes some connection between the irrelevant natural extension of the original network and that of its subnetworks, but it is hard to get a feeling for what it actually means. The following rather straightforward corollary expresses this connection more intuitively, and serves as a first—and crucial—step towards even stronger—and clearer—connections, such as the ones in Propositions $55_{\frown}$ and 56_{158}.

Corollary 54. *Let K be a closed subset of G and consider any $f \in \mathscr{G}(\mathscr{X}_K)$, $h \in \mathscr{G}(\mathscr{X}_{PN(K)})$, $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{\geq 0}$, $x_{P(K)} \in \mathscr{X}_{P(K)}$. Then*

$$f \notin \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}} \text{ and } h \notin \mathscr{D}^{\text{irr}}_{PN(K)} \Rightarrow h + g\mathbb{I}_{\{x_{P(K)}\}} f \notin \mathscr{D}^{\text{irr}}_G.$$

Proof of Corollary 54. Assume that $f \notin \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and that $h \notin \mathscr{D}^{\mathrm{irr}}_{PN(K)}$ and let $f_* := h + g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathscr{G}(\mathscr{X}_G)$. Assume *ex absurdo* that $f_* \in \mathscr{D}^{\mathrm{irr}}_G$. If we let

$$f' := \begin{cases} f & \text{if } f \neq 0 \\ -1 & \text{if } f = 0, \end{cases} \quad h' := \begin{cases} h & \text{if } h \neq 0 \\ -1 & \text{if } h = 0 \end{cases} \text{ and } g' := \begin{cases} g & \text{if } g \neq 0 \\ 1 & \text{if } g = 0, \end{cases}$$

then $f' \neq 0$, $h' \neq 0$ and $g' \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$. The coherence of $\mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $\mathscr{D}^{\mathrm{irr}}_{PN(K)}$ implies that $f' \notin \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $h' \notin \mathscr{D}^{\mathrm{irr}}_{PN(K)}$. Therefore, Theorem $53_{\frown}$ provides us with a linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$ such that $P_G(f_*) > 0$, $P_G(h') < 0$ and $P_G(g'\mathbb{I}_{\{x_{P(K)}\}} f') < 0$. If $h = 0$, then $P_G(h) = P_G(0) = 0$. Otherwise, $P_G(h) = P_G(h') < 0$. In any case: $P_G(h) \leq 0$. If $f = 0$ or $g = 0$, then $P_G(g\mathbb{I}_{\{x_{P(K)}\}} f) = P_G(0) = 0$. Otherwise, $P_G(g\mathbb{I}_{\{x_{P(K)}\}} f) = P_G(g'\mathbb{I}_{\{x_{P(K)}\}} f') < 0$. In any case, $P_G(g\mathbb{I}_{\{x_{P(K)}\}} f) \leq 0$. Due to the linearity of P_G, this implies that $P_G(f_*) \leq 0$, a contradiction. □

6.3 FACTORISATION AND EXTERNAL ADDITIVITY

The first important types of consequences of Theorem $53_{\frown}$ are factorisation and (external) additivity properties, which are crucial for the development of efficient algorithms.[6] The following generalisation of Proposition 39_{138} [with $K = \{s\}$] is a first example of a factorisation property; it will turn out to constitute the basis for many of the later results in this chapter.

Proposition 55. *Let K be a closed subset of G and consider any $f \in \mathscr{G}(\mathscr{X}_K)$, $x_{P(K)} \in \mathscr{X}_{P(K)}$ and $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$. Then*

$$g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathscr{D}^{\mathrm{irr}}_G \Leftrightarrow f \in \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}.$$

Proof of Proposition 55. Due to Corollary $54_{\frown}$ [with $h := 0$] we already know that $f \notin \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ implies $g\mathbb{I}_{\{x_{P(K)}\}} f \notin \mathscr{D}^{\mathrm{irr}}_G$. Hence, we only need to prove that $f \in \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ implies $g\mathbb{I}_{\{x_{P(K)}\}} f \in \mathscr{D}^{\mathrm{irr}}_G$.

By the coherence of $\mathscr{D}^{\mathrm{irr}}_G$ and the definition of $\mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $\mathscr{G}(\mathscr{X}_{N(K)})_{>0}$, we can assume, without loss of generality, that $f \in \mathscr{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ [Equation $(6.3)_{\frown}$] and $g = \mathbb{I}_{\{z_{N(K)}\}}$, with $z_{N(K)} \in \mathscr{X}_{N(K)}$. Since $f \in \mathscr{A}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, we know that $f = \mathbb{I}_{\{z_{PN_K(s)}\}} f'$ for some $s \in K$, $z_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}$ and $f' \in \mathscr{D}_{s\rfloor z_{P(s)}}$, with $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$. We need to prove that $\mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{PN_K(s)}\}} f' \in \mathscr{D}^{\mathrm{irr}}_G$. By Lemma 76_{182},

$$\mathbb{I}_{\{z_{P(s)}\}} = \mathbb{I}_{\{z_{P(s)\setminus P_K(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}} = \mathbb{I}_{\{x_{P(s)\setminus P_K(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}} = \mathbb{I}_{\{x_{P(K)\cap P(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}},$$

[6] See for example the algorithms in Sections $7.5.3_{249}$–$7.5.6_{264}$.

and therefore

$$\begin{aligned}\mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{PN_K(s)}\}}f' &= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{z_{P_K(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}f' \\ &= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{x_{P(K)\cap P(s)}\}}\mathbb{I}_{\{z_{P_K(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}f' \\ &= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}f' \\ &= \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}\mathbb{I}_{\{z_{P(s)}\}}f' = g'\mathbb{I}_{\{z_{P(s)}\}}f',\end{aligned}$$

where $g' := \mathbb{I}_{\{z_{N(K)}\}}\mathbb{I}_{\{x_{P(K)\setminus P(s)}\}}\mathbb{I}_{\{z_{N_K(s)}\}}$. This means that we are left to prove that $g'\mathbb{I}_{\{z_{P(s)}\}}f' \in \mathscr{D}_G^{\text{irr}}$. We have already explained in Section 5.4.1$_{135}$ and in the proof of Proposition 34$_{135}$ that, for all $y_{N(s)} \in \mathscr{X}_{N(s)}$, $\mathbb{I}_{\{y_{N(s)}\}}\mathbb{I}_{\{z_{P(s)}\}}f' \in \mathscr{D}_G^{\text{irr}}$ because $f' \in \mathscr{D}_{s\rfloor z_{P(s)}}$. Therefore, the desired result follows from the coherence of $\mathscr{D}_G^{\text{irr}}$ because, since Lemma 79(iv)$_{185}$ implies that $N(K)$, $P(K)\setminus P(s)$ and $N_K(s)$ are pairwise disjoint subsets of $N(s)$, we know that g' is a finite (and non-empty) sum of indicators $\mathbb{I}_{\{y_{N(s)}\}}$, $y_{N(s)} \in \mathscr{X}_{N(s)}$. □

Before we can formulate similar results in terms of other frameworks, we first need to extend our method for constructing the irrelevant natural extension of a subnetwork.

The main idea is identical. We fix a value $x_{P(K)} \in \mathscr{X}_{P(K)}$ for the variable $X_{P(K)}$ that corresponds to the parents of some subset K of G, and we use this value to define a collection of local models for the sub-DAG that corresponds to K. For every $s \in K$ and $z_{P_K(s)} \in \mathscr{X}_{P_K(s)}$, the corresponding local model is defined as

$$\underline{P}_{s\rfloor z_{P_K(s)}} = \underline{P}_{s\rfloor(z_{P_K(s)},x_{P(s)\setminus P_K(s)})} \tag{6.4}$$

or

$$\mathscr{M}_{s\rfloor z_{P_K(s)}} = \mathscr{M}_{s\rfloor(z_{P_K(s)},x_{P(s)\setminus P_K(s)})} \tag{6.5}$$

or

$$\mathscr{F}_{s\rfloor z_{P_K(s)}} = \mathscr{F}_{s\rfloor(z_{P_K(s)},x_{P(s)\setminus P_K(s)})}, \tag{6.6}$$

depending on the framework that is adopted. Once we have these local models, we can consider their irrelevant natural extensions, which we will denote by $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}$, $\mathscr{M}^{\text{irr}}_{K\rfloor x_{P(K)}}$ and $\mathscr{F}^{\text{irr}}_{K\rfloor x_{P(K)}}$, respectively.

The following result is a first example of how these irrelevant natural extensions of subnetworks are related to those of the original network. It can be regarded as a translation of Corollary 54$_{156}$ and Proposition 55$_{\frown}$ to the framework of coherent lower previsions. The proof can be found in Appendix 6.C$_{199}$.

Proposition 56. *Let K be a closed subset of G and consider any $f \in \mathscr{G}(\mathscr{X}_K)$, $x_{P(K)} \in \mathscr{X}_{P(K)}$, $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{\geq 0}$ and $h \in \mathscr{G}(\mathscr{X}_{PN(K)})$. Then*

$$\underline{P}^{\mathrm{irr}}_{G}(h+g\mathbb{I}_{\{x_{P(K)}\}}f)=\underline{P}^{\mathrm{irr}}_{PN(K)}\big(h+g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f)\big).$$

The practical significance of this result is that it allows us to split a global computational problem—evaluating $\underline{P}^{\mathrm{irr}}_{G}$—into two smaller ones—evaluating $\underline{P}^{\mathrm{irr}}_{PN(K)}$ and $\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$. For $h=0$, we obtain factorisation as a special case.

Corollary 57 (factorisation)**.** *Let K be a closed subset of G and consider any $f\in\mathscr{G}(\mathscr{X}_K)$, $x_{P(K)}\in\mathscr{X}_{P(K)}$ and $g\in\mathscr{G}(\mathscr{X}_{N(K)})_{\geq 0}$. Then*

$$\underline{P}^{\mathrm{irr}}_{G}(g\mathbb{I}_{\{x_{P(K)}\}}f)=\begin{cases}\underline{P}^{\mathrm{irr}}_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f) & \text{if } \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f)\geq 0\\ \overline{P}^{\mathrm{irr}}_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f) & \text{if } \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f)\leq 0.\end{cases}$$

Proof of Corollary 57. This is an immediate consequence of Proposition 56$_{\curvearrowleft}$ [with $h=0$], the coherence of $\underline{P}^{\mathrm{irr}}_{PN(K)}$ [C2$_{48}$] and conjugacy. □

The following result—and its proof—nicely illustrates how Corollary 57 allows us to reduce a global optimisation problem into smaller subproblems.

Proposition 58. *Consider any $x_G\in\mathscr{X}_G$. Then*

$$\underline{P}^{\mathrm{irr}}_{G}(x_G)=\prod_{s\in G}\underline{P}_{s\rfloor x_{P(s)}}(x_s)\ \textit{and}\ \overline{P}^{\mathrm{irr}}_{G}(x_G)=\prod_{s\in G}\overline{P}_{s\rfloor x_{P(s)}}(x_s)$$

Proof of Proposition 58. For every $s\in G$, it follows from Corollary 44$_{141}$ and conjugacy that $\underline{P}^{\mathrm{irr}}_{s\rfloor x_{P(s)}}(x_s)=\underline{P}_{s\rfloor x_{P(s)}}(x_s)$ and $\overline{P}^{\mathrm{irr}}_{s\rfloor x_{P(s)}}(x_s)=\overline{P}_{s\rfloor x_{P(s)}}(x_s)$. If G contains only a single node s—if $G:=\{s\}$—this already establishes the result.

For $|G|>1$, we provide a proof by induction. The induction hypothesis is that the result holds for all credal networks with less than $|G|$ nodes. Let ℓ be an arbitrary leaf of the DAG of the network [every DAG has at least one leaf]. Then $D(\ell)=\emptyset$ and therefore also $PN(\ell)=G\setminus\{\ell\}$. Since coherence [C1$_{48}$] implies that $\underline{P}^{\mathrm{irr}}_{s\rfloor x_{P(s)}}(\mathbb{I}_{\{x_s\}})\geq 0$, we infer from Corollary 57 that

$$\begin{aligned}\underline{P}^{\mathrm{irr}}_{G}(x_G)=\underline{P}^{\mathrm{irr}}_{G}(\mathbb{I}_{\{x_G\}})&=\underline{P}^{\mathrm{irr}}_{G\setminus\{\ell\}}(\mathbb{I}_{\{x_{G\setminus\{\ell\}}\}})\underline{P}^{\mathrm{irr}}_{s\rfloor x_{P(s)}}(\mathbb{I}_{\{x_s\}})\\&=\underline{P}^{\mathrm{irr}}_{G\setminus\{\ell\}}(x_{G\setminus\{\ell\}})\underline{P}^{\mathrm{irr}}_{s\rfloor x_{P(s)}}(x_s)\\&=\underline{P}^{\mathrm{irr}}_{G\setminus\{\ell\}}(x_{G\setminus\{\ell\}})\underline{P}_{s\rfloor x_{P(s)}}(x_s).\end{aligned}$$

The first equation of this corollary now follows from the induction hypothesis. The proof of the second equation is completely analogous—after applying conjugacy—and is therefore omitted. □

For $P(K) = \emptyset$, we obtain *external additivity*[7] as another important special case of Proposition 56_{158}.

Corollary 59 (external additivity)**.** *Let K be an ancestral set—a closed subset of G such that $P(K) = \emptyset$—and consider any $f \in \mathscr{G}(\mathscr{X}_K)$ and $h \in \mathscr{G}(\mathscr{X}_{N(K)})$. Then*

$$\underline{P}^{\mathrm{irr}}_G(h+f) = \underline{P}^{\mathrm{irr}}_{N(K)}(h) + \underline{P}^{\mathrm{irr}}_K(f).$$

Proof of Corollary 59. This is an immediate consequence of Proposition 56_{158} [with $g = 1$] and the coherence of $\underline{P}^{\mathrm{irr}}_{N(K)}$ [$\mathrm{C8}_{49}$]. □

Analogous results can be obtained in terms of sets of linear previsions and sets of probability mass functions, in a trivial way, by interpreting $\underline{P}^{\mathrm{irr}}_G$ as the lower envelope of $\mathscr{M}^{\mathrm{irr}}_G$ or $\mathscr{M}_{\mathscr{F}^{\mathrm{irr}}_G}$, and similarly for $\underline{P}^{\mathrm{irr}}_{PN(K)}$, $\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$, $\underline{P}^{\mathrm{irr}}_{N(K)}$, $\underline{P}^{\mathrm{irr}}_K$ and $\underline{P}_{s\rfloor x_{P(K)}}$; see Theorem 46_{142} and Proposition 49_{143}.

6.4 CONDITIONING, MARGINALISATION AND IRRELEVANCE

The factorisation and external additivity properties in the previous section already illustrate that for closed subsets K of G, the irrelevant natural extensions of the corresponding subnetworks—one for every $x_{P(K)} \in \mathscr{X}_{P(K)}$—are related to that of the original network. The following result makes this connection even clearer: the irrelevant natural extension of these subnetworks can be obtained by conditioning and marginalising the irrelevant natural extension of the original network.

Corollary 60. *Let K be a closed subset of G and consider any $x_{P(K)} \in \mathscr{X}_{P(K)}$ and $B_{N(K)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(K)})$. Then*

$$\mathrm{marg}_K(\mathscr{D}^{\mathrm{irr}}_G \rfloor \{x_{P(K)}\} \times B_{N(K)}) = \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$$

and

$$\underline{P}^{\mathrm{irr}}_K(\cdot\rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$$

and

$$\mathrm{marg}^{\mathrm{c}}_K(\mathscr{M}^{\mathrm{irr}}_G \rfloor\!| \{x_{P(K)}\} \times B_{N(K)}) = \mathscr{M}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$$

and

$$\mathrm{marg}^{\mathrm{c}}_K(\mathscr{F}^{\mathrm{irr}}_G \rfloor\!| \{x_{P(K)}\} \times B_{N(K)}) = \mathscr{F}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}.$$

[7] We use this terminology because this property is a generalisation of the (strong) external additivity property discussed in Reference [46]; see Section 6.6_{171} and Footnote 21_{172} as well.

Proof of Corollary 60$_{\frown}$. Consider any $f \in \mathscr{G}(\mathscr{X}_K)$. Since $\mathbb{I}_{B_{N(K)}}$ is an element of $\mathscr{G}(\mathscr{X}_{N(K)})_{>0}$, we find that

$$f \in \operatorname{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor \{x_{P(K)}\} \times B_{N(K)}) \Leftrightarrow \mathbb{I}_{B_{N(K)}} \mathbb{I}_{\{x_{P(K)}\}} f \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow f \in \mathscr{D}_K^{\mathrm{irr}} \rfloor x_{P(K)},$$

where the two equivalences follow from Equation (4.6)$_{107}$ and Proposition 55$_{157}$, respectively. This already implies the first equality of this corollary. The second equality is a consequence of the first, Theorem 42$_{139}$ and the discussion after Equation (4.13)$_{110}$. The third equality follows from the second and Theorem 46$_{142}$ and 33$_{110}$. The final equality is a consequence of the third equality, Proposition 49$_{143}$ and the discussion at the end of Section 4.2.4$_{109}$. □

In order to illustrate the generality of this result, let us consider the special case where K is an ancestral set. It then follows from Corollary 60$_{\frown}$ that, for example, $\operatorname{marg}_K^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}})$ and $\mathscr{F}_K^{\mathrm{irr}}$ are equal, which implies that their unconditional parts—$\operatorname{marg}_K^{\mathrm{c}}(\mathscr{F}_G) \rfloor \mathscr{X}_K = \operatorname{marg}_K(\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G)$ and $\mathscr{F}_K^{\mathrm{irr}} \rfloor \mathscr{X}_K$, respectively—are equal as well. This result has already been proved by Cozman in Reference [16, Theorem 15],[8] under the assumption that all lower probabilities are strictly positive. Here, we recover it as a particular special case.

As illustrated by our next result, Corollary 60$_{\frown}$ also implies a number of irrelevancies. For now, this can be regarded as being of minor importance: we will show in the next section [see Corollary 66$_{168}$] that the irrelevant natural extension satisfies many more irrelevancies than the ones in the following corollary.

Corollary 61. *For all closed sets $K \subseteq G$, the irrelevant natural extension satisfies the following irrelevance statement:* $\mathrm{IR}(N(K), K \mid P(K))$. *In other words, for any $x_{P(K)} \in \mathscr{X}_{P(K)}$ and $B_{N(K)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(K)})$, we have that*

$$\operatorname{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor \{x_{P(K)}\} \times B_{N(K)}) = \operatorname{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_{P(K)})$$

and

$$\underline{P}_K^{\mathrm{irr}}(\cdot \rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}_K^{\mathrm{irr}}(\cdot \rfloor \cdot \times \{x_{P(K)}\})$$

and

$$\operatorname{marg}_K^{\mathrm{c}}(\mathscr{M}_G^{\mathrm{irr}} \rfloor | \{x_{P(K)}\} \times B_{N(K)}) = \operatorname{marg}_K^{\mathrm{c}}(\mathscr{M}_G^{\mathrm{irr}} \rfloor | x_{P(K)})$$

and

$$\operatorname{marg}_K^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | \{x_{P(K)}\} \times B_{N(K)}) = \operatorname{marg}_K^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | x_{P(K)}).$$

[8] Reference [16] refers to an ancestral set as a top-subnetwork.

Proof of Corollary 61$_{\cap}$. By applying Corollary 60_{160} twice, we find that

$$\mathrm{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor \{x_{P(K)}\} \times B_{N(K)}) = \mathscr{D}_{K \rfloor x_{P(K)}}^{\mathrm{irr}} = \mathrm{marg}_K(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_{P(K)}),$$

where we let $B_{N(K)} = \mathscr{X}_{N(K)}$ for the second equality. This already establishes the first equation of this corollary. The other three equations now follow from Theorems 42_{139} and 46_{142}, Proposition 49_{143} and the discussion in Section $4.3.4_{119}$. □

6.5 SEPARATION PROPERTIES

In probabilistic graphical networks that are defined by means of a symmetrical independence concept, d-separation [57, 78] is a very powerful tool. On the one hand, this separation concept is purely graphical, as it is defined solely in terms of the DAG of the network. On the other hand, for Bayesian networks, it is guaranteed to imply probabilistic independence of the variables that are associated with the separated nodes. This is rather remarkable, as it allows us to verify independencies without resorting to numerical computations.

In graphical networks that adopt an asymmetric notion of independence, such as credal networks under epistemic irrelevance, d-separation cannot be expected to have a similar property, because d-separation is symmetric. If d-separation were to imply epistemic irrelevance, it would imply that epistemic irrelevance is—in this context—symmetric, which is clearly not the case.

Nevertheless, separation properties that are similar to those of Bayesian networks can still be obtained. We will replace d-separation by an asymmetric alternative, called AD-separation, and will show that for the irrelevant natural extension, this new notion of separation implies epistemic irrelevance.

6.5.1 *Asymmetric D-separation*

Asymmetric versions of d-separation have already been proposed in the literature. Moral [76] speaks of *asymmetric D-separation*[9] (AD-separation) and Vantaggi [100–103] has introduced the very similar *L-separation* criterion. Here, we will not use any of these existing concepts, but choose to introduce a slightly modified version. Borrowing Moral's terminology, we will call it *AD-separation* (asymmetric d-separation).[10] We prefer our version because

[9] Judging by the references he provides [78, 79], Moral actually seems to mean asymmetric d-separation rather than asymmetric D-separation; D-separation is an enhanced version of d-separation that allows for deterministic nodes [58]. However, since d-separation is a special case of D-separation, the term asymmetric D-separation (AD-separation) does not produce a conflict in terminology and we choose to adopt it as well.

[10] We prefer Moral's terminology over the one by Vantaggi because (i) we think that the term asymmetric d-separation really captures the meaning of the concept and (ii) the L in L-separation

our definition is weaker than—in the sense that it is implied by—Moral's AD-separation, slightly more general[11] than Vantaggi's L-separation and yet, it has stronger properties than both of these other concepts.

Consider any path $s_1, \ldots, s_n$ in G, with $n \geq 1$. We say that this path is *blocked* by a set of nodes $C \subseteq G$ whenever at least one of the following four conditions holds:

B1. $s_1 \in C$;

B2. there is a node s_i, with $1 < i < n$, such that $s_i \rightarrow s_{i+1}$ and $s_i \in C$;

B3. there is a node s_i, with $1 < i < n$, such that $s_{i-1} \rightarrow s_i \leftarrow s_{i+1}$, $s_i \notin C$ and $D(s_i) \cap C = \emptyset$;

B4. $s_n \in C$.

In Moral's version of AD-separation, the notion of a blocked path is very similar. The only difference is condition B1, which he strengthens by requiring that $s_1 \rightarrow s_2$. Clearly, our condition is implied by Moral's. Vantaggi uses the same notion of blocked path as we do,[12] but leaves out conditions B1 and B4. They are redundant in her case, because she does not need to consider cases where s_1 or s_n are elements of C.[13]

Example 7. *Figure 6.3$_{\frown}$ illustrates how each of the blocking conditions* B1–B4 *can block a path. The examples for* B1 *and* B4 *are straightforward. Note that in the example of* B2*, the crucial point is the arrow between s_3 and s_5. If that arrow were reversed, the path would no longer be blocked. In the example of* B3*, it is essential that s_5, s_6 and s_7 are not elements of C. If any of them were, the path would not be blocked. Notice also that the path in the example for* B1 *is not blocked according to Moral's version of AD-separation, the reason being that the arrow between s_3 and s_2 is pointing in the wrong direction.*

We are now ready to define the important concept of AD-separation.

---

refers to the logical constraints that can be imposed in Vantaggi's framework, which do not seem relevant for our current purposes. It is however important to keep in mind that our notion of AD-separation is different than the one by Moral. We consider the resulting conflict in terminology to be minor, especially since our version of AD-separation is implied by Moral's version.

[11] At least as far as the sets on which it can defined is concerned: L-separation is only defined for pairwise disjoint sets. We should however mention that, if one restricts oneself to pairwise disjoint sets, Vantaggi's L-separation criterion is more general than ours because it also includes the possibility to include logical constraints, which our notion of AD-separation does not.

[12] At first sight, it might seem as if she does not; loosely speaking, the confusion arises because she applies her definition to the reversed path.

[13] Because L-separation is defined for pairwise disjoint sets only; see Definition 3$_{\frown}$.

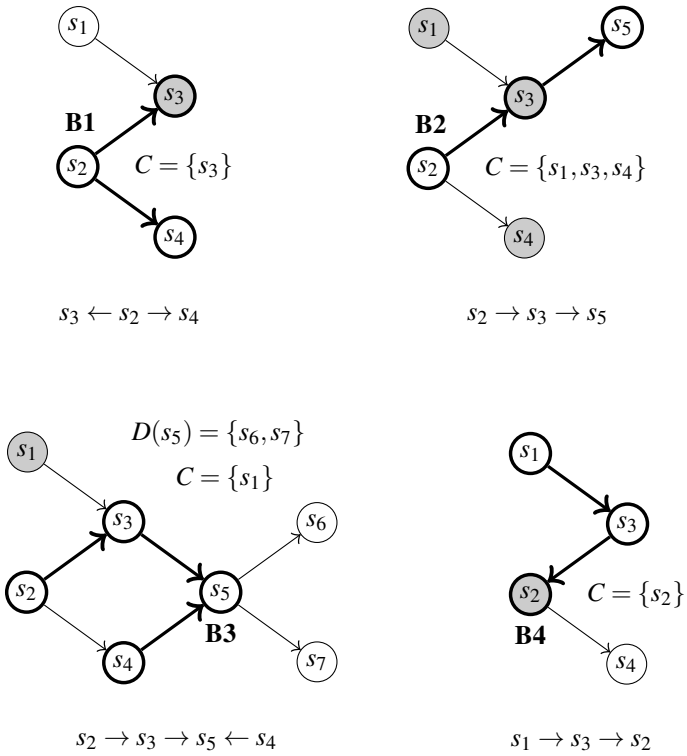


Figure 6.3: Illustration of paths that are blocked by conditions $\mathrm{B1}_{\frown}$–$\mathrm{B4}_{\frown}$.

Definition 3 (AD-separation)**.** *Consider (not necessarily pairwise disjoint) subsets I, S and C of G. Then I is* AD-separated *from S by C, denoted as* $\mathrm{AD}(I, S \mid C)$*, if every path* $i = s_1, \ldots, s_n = s$*,* $n \geq 1$*, from any node* $i \in I$ *to any node* $s \in S$*, is blocked by C.*

Figure $6.4_{\frown}$ provides an example of AD-separated sets.

Moral and Vantaggi define their separation criteria in much the same way. The only difference with Moral's version of AD-separation is his notion of a blocked path, as explained earlier. Clearly, AD-separation in Moral's sense implies AD-separation in our sense. The difference with Vantaggi's criterion is that L-separation is defined for pairwise disjoint sets only. Notice that if we restrict ourselves to pairwise disjoint sets, AD-separation (both our version and the one by Moral) is identical to L-separation.[14]

[14]Making abstraction of the logical component of L-separation.

$I = \{s_3, s_6, s_{12}, s_{14}\}$ $\qquad$ $\mathrm{AD}(I, S \mid C)$

$S = \{s_1, s_7, s_9\}$

$C = \{s_1, s_5, s_6\}$ $\qquad$ $G = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}, s_{11}, s_{12}, s_{13}, s_{14}, s_{15}\}$

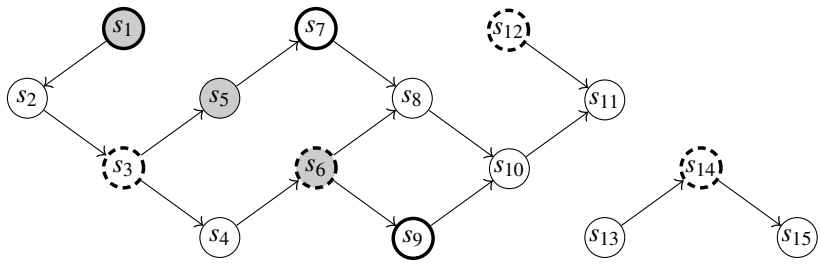


Figure 6.4: Illustration of AD-separation

Like d-separation, AD-separtion satisfies a number of graphoid properties [78], which can be used to infer new separations from separations that were already known. Since symmetry is not satisfied—$\mathrm{AD}(I, S \mid C)$ is not equivalent to $\mathrm{AD}(S, I \mid C)$—the other graphoid properties [see Theorem 62] now come in different versions, depending on the order of the first two arguments in $\mathrm{AD}(\cdot, \cdot \mid \cdot)$ [22]. Redundancy has two versions. Decomposition and Weak union come in four different versions, and contraction and intersection can even be stated in eight different ways. However, for each graphoid property, only two versions make sense. The 'direct' version, which is how they are usually stated, and the 'reverse' version, where the order of the two first arguments has been reversed in every instance of $\mathrm{AD}(\cdot, \cdot \mid \cdot)$. We will refer to the resulting ten properties as the *asymmetric graphoid properties*.

Theorem 62. *AD-separation satisfies all asymmetric graphoid properties. For any subsets I, S, W and C of G:*[15]

Direct redundancy: $\mathrm{AD}(I, S \mid I)$

Reverse redundancy: $\mathrm{AD}(I, S \mid S)$

[15] We follow Reference [22] in naming these properties. Moral [76] uses almost the same terminology; the only difference is that he interchanges the meaning of direct and reverse intersection; see Footnote 18_{169} as well. Vantaggi [101] uses a different terminology: for example, her notion of reverse decomposition refers to a property denoted as $(I \cup W, S|C)^I_G \Rightarrow (I, S|C)^I_G$, which seems similar to our notion of reverse decomposition, but actually, corresponds to what we call direct decomposition, since, loosely speaking, Vantaggi reverses the order in which I and S occur in the notation. Care should therefore be taken in comparing results.

Direct decomposition: $\mathrm{AD}(I, S \cup W \mid C) \Rightarrow \mathrm{AD}(I, S \mid C)$

Reverse decomposition: $\mathrm{AD}(I \cup W, S \mid C) \Rightarrow \mathrm{AD}(I, S \mid C)$

Direct weak union: $\mathrm{AD}(I, S \cup W \mid C) \Rightarrow \mathrm{AD}(I, S \mid C \cup W)$

Reverse weak union: $\mathrm{AD}(I \cup W, S \mid C) \Rightarrow \mathrm{AD}(I, S \mid C \cup W)$

Direct contraction: $\mathrm{AD}(I, S \mid C)$ & $\mathrm{AD}(I, W \mid C \cup S) \Rightarrow \mathrm{AD}(I, S \cup W \mid C)$

Reverse contraction: $\mathrm{AD}(I, S \mid C)$ & $\mathrm{AD}(W, S \mid C \cup I) \Rightarrow \mathrm{AD}(I \cup W, S \mid C)$

Direct intersection: *if* $S \cap W = \emptyset$*, then*

$$\mathrm{AD}(I, S \mid C \cup W) \text{ \& } \mathrm{AD}(I, W \mid C \cup S) \Rightarrow \mathrm{AD}(I, S \cup W \mid C)$$

Reverse intersection: *if* $I \cap W = \emptyset$*, then*

$$\mathrm{AD}(I, S \mid C \cup W) \text{ \& } \mathrm{AD}(W, S \mid C \cup I) \Rightarrow \mathrm{AD}(I \cup W, S \mid C)$$

This result—and our proof for it in Appendix $6.\mathrm{D}_{201}$—is very similar to, and heavily inspired by, the work of Vantaggi [101, Theorem 7.1].[16] The main difference is that Vantaggi does not include the two redundancy properties, since L-separation is defined only for *pairwise disjoint* subsets I, S and C of G. Moral's version of AD-separation does not require I, S and C to be pairwise disjoint, but it does not satisfy direct redundancy, and proofs for a number of other properties are not given [76, Theorem 4].

The following two results provide an alternative characterisation for AD-separation. The first one shows that, as far as checking AD-separation is concerned, we can restrict attention to pairwise disjoint sets $I, S, C \subseteq G$. The second result establishes that, for such pairwise disjoint sets, AD-separation can be characterised in terms of closed subsets of G. The proof for Theorem $64_{\frown}$ requires a few additional technical results and has therefore been relegated to Appendix $6.\mathrm{E}_{203}$.

Proposition 63. *Consider any subsets I, S and C of G. Then*

$$\mathrm{AD}(I, S \mid C) \Leftrightarrow \mathrm{AD}(I \setminus C, S \setminus C \mid C) \textit{ and } (I \setminus C) \cap (S \setminus C) = \emptyset.$$

[16] We provide a direct proof for Theorem $62_{\frown}$. However, as suggested to us by Barbara Vantaggi, our result can probably be derived as a corollary of Reference [101, Theorem 7.1] as well. A possible way of doing so could be to first prove (direct and reverse) redundancy and decomposition (which is trivial) and to use these properties to try and infer (direct and reverse) weak union, contraction and intersection from their 'pairwise disjoint' versions (which were proved by Vantaggi in Reference [101, Theorem 7.1]).

Proof of Proposition 63$_{\cap}$. First assume that $\mathrm{AD}(I,S \mid C)$. This clearly implies that $\mathrm{AD}(I \setminus C, S \setminus C \mid C)$ because AD-separation satisfies direct and reverse decomposition [see Theorem 62$_{165}$]. This in turn implies that $I \setminus C$ and $S \setminus C$ are disjoint, because otherwise, any element $t \in (I \setminus C) \cap (S \setminus C)$ would be a trivial path from $I \setminus C$ to $S \setminus C$—containing only the single node t—that is not blocked by C, thereby contradicting $\mathrm{AD}(I \setminus C, S \setminus C \mid C)$.

Next, assume that $\mathrm{AD}(I \setminus C, S \setminus C \mid C)$ and that $I \setminus C$ and $S \setminus C$ are disjoint. Consider any path from $i \in I$ to $s \in S$. If $i \in C$ or $s \in C$, this path is trivially blocked [due to B1$_{163}$ or B4$_{163}$, respectively]. If $i \notin C$ and $s \notin C$, then $i \in I \setminus C$ and $s \in S \setminus C$, which implies that the path is blocked by C because we know that $\mathrm{AD}(I \setminus C, S \setminus C \mid C)$. □

Theorem 64. *Consider pairwise disjoint subsets I, S and C of G. Then $\mathrm{AD}(I,S \mid C)$ if and only if there is some closed subset K of G such that $S \subseteq K$, $P(K) \subseteq C$, $I \subseteq N(K)$ and $D(K) \cap C = \emptyset$.*

6.5.2 *Graphoid properties of credal networks*

The main reason why we have introduced AD-separation is because credal networks under epistemic irrelevance satisfy a property that is very similar to the classical d-separation result in Bayesian networks: for the irrelevant natural extension of a credal network, AD-separation implies epistemic irrelevance. In order to prove this, the starting point is the following factorisation property.

Theorem 65. *Consider any pairwise disjoint $I,S,C \subseteq G$ such that $\mathrm{AD}(I,S \mid C)$. Then for all $x_C \in \mathscr{X}_C$, $g \in \mathscr{G}(\mathscr{X}_I)_{>0}$ and $f \in \mathscr{G}(\mathscr{X}_S)$:*

$$g\mathbb{I}_{\{x_C\}}f \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow \mathbb{I}_{\{x_C\}}f \in \mathscr{D}_G^{\mathrm{irr}}.$$

Proof of Theorem 65. It follows from Theorem 64 that there is a closed subset K of G such that $S \subseteq K$, $P(K) \subseteq C$, $I \subseteq N(K)$ and $D(K) \cap C = \emptyset$. By combining Lemma 78(ii)$_{183}$ with the fact that $P(K) \subseteq C$ and $D(K) \cap C = \emptyset$, we infer that $\mathbb{I}_{\{x_C\}} = \mathbb{I}_{\{x_{C\cap N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{x_{C\cap K}\}}$. Now let $g' := g\mathbb{I}_{\{x_{C\cap N(K)}\}}$, $g'' := \mathbb{I}_{\{x_{C\cap N(K)}\}}$ and $f' := \mathbb{I}_{\{x_{C\cap K}\}}f$, then

$$g\mathbb{I}_{\{x_C\}}f = g\mathbb{I}_{\{x_{C\cap N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{x_{C\cap K}\}}f = g'\mathbb{I}_{\{x_{P(K)}\}}f'$$

and

$$\mathbb{I}_{\{x_C\}}f = \mathbb{I}_{\{x_{C\cap N(K)}\}}\mathbb{I}_{\{x_{P(K)}\}}\mathbb{I}_{\{x_{C\cap K}\}}f = g''\mathbb{I}_{\{x_{P(K)}\}}f'.$$

Since $f' \in \mathscr{G}(\mathscr{X}_K)$, $g' \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$ and $g'' \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$, we can now apply Proposition 55$_{157}$ twice to find that

$$\begin{aligned} g\mathbb{I}_{\{x_C\}}f \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow g'\mathbb{I}_{\{x_{P(K)}\}}f' \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow f' \in \mathscr{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}} \\ \Leftrightarrow g''\mathbb{I}_{\{x_{P(K)}\}}f' \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow \mathbb{I}_{\{x_C\}}f \in \mathscr{D}_G^{\mathrm{irr}}. \end{aligned}$$

□

Using this factorisation property, it is now fairly easy to prove that AD-separation implies epistemic irrelevance.

Corollary 66. *Consider any pairwise disjoint sets $I,S,C \subseteq G$ such that* $\mathrm{AD}(I,S \mid C)$. *The irrelevant natural extension then satisfies the irrelevance statement* $\mathrm{IR}(I,S \mid C)$*: for any $x_C \in \mathscr{X}_C$ and $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$, we have that*

$$\mathrm{marg}_S(\mathscr{D}_G^{\mathrm{irr}} \rfloor \{x_C\} \times B_I) = \mathrm{marg}_S(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_C)$$

and

$$\underline{P}_S^{\mathrm{irr}}(\cdot \rfloor \cdot \times \{x_C\} \times B_I) = \underline{P}_S^{\mathrm{irr}}(\cdot \rfloor \cdot \times \{x_C\})$$

and

$$\mathrm{marg}_S^{\mathrm{c}}(\mathscr{M}_G^{\mathrm{irr}} \rfloor | \{x_C\} \times B_I) = \mathrm{marg}_S^{\mathrm{c}}(\mathscr{M}_G^{\mathrm{irr}} \rfloor | x_C)$$

and

$$\mathrm{marg}_S^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | \{x_C\} \times B_I) = \mathrm{marg}_S^{\mathrm{c}}(\mathscr{F}_G^{\mathrm{irr}} \rfloor | x_C).$$

Proof of Corollary 66. Consider any $f \in \mathscr{G}(\mathscr{X}_S)$. Since $\mathbb{I}_{B_I}$ is an element of $\mathscr{G}(\mathscr{X}_I)_{>0}$, and because $\mathrm{AD}(I,S \mid C)$, we can apply Theorem $65_{\cap}$ and Equation $(4.6)_{107}$ to find that

$$\begin{aligned} f \in \mathrm{marg}_S(\mathscr{D}_G^{\mathrm{irr}} \rfloor \{x_C\} \times B_I) &\Leftrightarrow \mathbb{I}_{B_I}\mathbb{I}_{\{x_C\}} f \in \mathscr{D}_G^{\mathrm{irr}} \\ &\Leftrightarrow \mathbb{I}_{\{x_C\}} f \in \mathscr{D}_G^{\mathrm{irr}} \Leftrightarrow f \in \mathrm{marg}_S(\mathscr{D}_G^{\mathrm{irr}} \rfloor x_C). \end{aligned}$$

This concludes the proof for the first equation of this corollary. The other three equalities now follow from Theorem 42_{139} and 46_{142}, Proposition 49_{143} and the discussion in Section $4.3.4_{119}$. □

This result can be regarded as a generalisation of Corollary 61_{161} [see Lemma 82_{205}]. From a practical point of view, the significance of Corollary 66 is that it allows us to detect epistemic irrelevancies in the joint model in a purely graphical way, without resorting to numerical computations; all we have to do is check for AD-separation.

6.5.3 *On the relevance of graphoid axioms*

Graphoid properties are not only associated with graphical separation criteria. They can also be applied to notions of independence. In that context, these properties are often regarded as axioms that a notion of independence should satisfy, and it has become common practice to compare different notions of independence by means of the graphoid axioms that they satisfy.

For example, for epistemic irrelevance, one could wonder whether the operator $\mathrm{IR}(\cdot,\cdot \mid \cdot)$ satisfies the properties that were proved to hold for the operator $\mathrm{AD}(\cdot,\cdot \mid \cdot)$ in Theorem 62_{165}. It turns out that some of them are indeed satisfied. For example, it follows easily from the definition of epistemic h-irrelevance that this notion of independence satisfies decomposition and weak

union, in both their direct and reverse form. It could be argued—and we would agree—that this is a reason for preferring epistemic h-irrelevance over epistemic value- or subset-irrelevance, because, if the conditioning events have zero lower probability, epistemic value- and subset-irrelevance may start to fail some of these properties.[17]

However, we think that care should be taken in regarding graphoid properties as axioms. Although we agree that it is reasonable to require any notion of independence to satisfy decomposition and weak union, this is no longer obvious for some of the other graphoid properties. Consider for example direct contraction. If we were to impose that epistemic irrelevance should satisfy this graphoid property, this would mean that

$$\mathrm{IR}(I,S \mid C) \text{ and } \mathrm{IR}(I,W \mid C \cup S) \Rightarrow \mathrm{IR}(I,S \cup W \mid C) \tag{6.7}$$

Consider now the framework of sets of desirable gambles and, for simplicity, let $C = \emptyset$. The left-hand side of the above implication then holds if and only if, for all $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$:

$$\mathrm{marg}_S(\mathscr{D}_G \rfloor B_I) = \mathrm{marg}_S(\mathscr{D}_G) \text{ and } \mathrm{marg}_W(\mathscr{D}_G \rfloor \{x_S\} \times B_I) = \mathrm{marg}_W(\mathscr{D}_G \rfloor x_S).$$

In other words, if we let $\mathscr{D}_{S\cup W \rfloor B_I} := \mathrm{marg}_{S\cup W}(\mathscr{D}_G \rfloor B_I)$, then $\mathrm{marg}_S(\mathscr{D}_{S\cup W \rfloor B_I})$ and $\mathrm{marg}_W(\mathscr{D}_{S\cup W \rfloor B_I} \rfloor x_S)$ do not depend on B_I. According to Equation (6.7), this should then imply that $\mathscr{D}_{S\cup W \rfloor B_I}$ is independent of B_I. We do not consider it reasonable to enforce this, because it would mean that $\mathscr{D}_{S\cup W \rfloor B_I}$ should be fully determined by $\mathrm{marg}_S(\mathscr{D}_{S\cup W \rfloor B_I})$ and $\mathrm{marg}_W(\mathscr{D}_{S\cup W \rfloor B_I} \rfloor x_S)$ or, more generally, that for any combination of a belief model for X_S and a belief model for X_W conditional on X_S, there is only one compatible belief model for $X_{S\cup W}$. It is well known that, in general, this is not true for imprecise-probabilistic belief models. We think that this is perfectly normal, and that there is no fundamental reason why such a property should hold. For that reason, we consider it unreasonable to regard direct contraction as an axiom. A similar argument can be used to question direct intersection as a reasonable axiom. It is therefore not surprising that many notions of epistemic irrelevance and epistemic independence do not satisfy these graphoid properties; see for example Reference [22]. Moral's version of epistemic irrelevance in Reference [76] is an important exception because it does satisfy direct contraction.[18] However, this comes at a serious cost. We discuss this further in Section 6.5.4$_\cap$; for now, it suffices to say that on Moral's version of epistemic irrelevance, it is impossible for two variables to be mutually irrelevant, except in some degenerate cases.

[17] Reference [22] provides some examples for epistemic value-irrelevance.

[18] Moral also proves that it satisfies direct intersection, but his terminology differs from ours as well as from that in Reference [22]. Moral's notion of direct intersection is what we call reverse intersection. We prefer our terminology because of the obvious similarity between what we call direct contraction and direct intersection.

In any case, regardless of whether or not we regard graphoid properties as axioms that a notion of independence should satisfy, they are definitely useful from a computational point of view. For example, in Bayesian networks, the proof for the counterpart to Corollary 66_{168}—with AD-separation replaced by d-separation and epistemic irrelevance replaced by stochastic independence—makes use of the fact that—given some positivity assumptions—stochastic independence satisfies all graphoid properties [78]. By applying these properties to the independence assessments that are used to define a Bayesian network, one can infer new independencies, and those are exactly the ones that correspond to the d-separations in the DAG of a Bayesian network.

If one tries to mimic this approach in our context, then since epistemic irrelevance can fail some of the graphoid properties, one might (erroneously) be led to suspect that Corollary 66_{168} cannot be proved. In fact, we believe that this might be one of the main reasons why a result such as Corollary 66_{168} has thus far not appeared in the literature on credal networks. However, as our proof for Corollary 66_{168} illustrates, it is not necessary to make use of the graphoid properties of epistemic irrelevance: our proof for Theorem 65_{167}—of which Corollary 66_{168} is a straightforward consequence—only uses Proposition 55_{157} and a number of properties of AD-separation. At no point does it invoke graphoid properties of epistemic irrelevance.

We conclude from all this that there is no need to focus on notions of independence that satisfy as many graphoid properties as possible, such as, for example, complete independence. It is neither compelling from a philosophical point of view—because some of the graphoid properties do not seem to make reasonable sense as axioms of independence—nor necessary from a practical point of view—because it might be, and actually is, possible to provide alternative proofs. In principle, any intuitive notion of independence can be used to construct an uncertainty model. From a practical point of view, the mathematical properties of the resulting model are more important than the graphoid properties of the notion of independence that was used to construct it. Graphoid properties can serve as useful tools for proving such mathematical properties, but—as we have illustrated—other approaches can be equally successful.

6.5.4 *A crucial difference with earlier work by Moral*

Readers who are familiar with Moral's results in Reference [76] may have noticed the similarity between Reference [76, Theorem 5] and the first equation in Corollary 66_{168}. The main difference between our approach and Moral's approach, besides the fact that we use a slightly different separation criterion, is that he enforces a more stringent version of epistemic irrelevance than we do. He calls X_I epistemically irrelevant to X_S if and only if $\mathrm{marg}_{I\cup S}(\mathscr{D}_G)$ is the unique smallest set of desirable gambles on $\mathscr{X}_{I\cup S}$ that has $\mathrm{marg}_I(\mathscr{D}_G)$ and $\mathrm{marg}_S(\mathscr{D}_G)$ as its marginal models and according to which X_I is irrelevant to

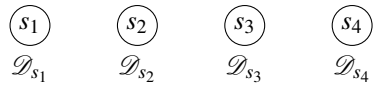


Figure 6.5: Example of a credal network consisting of disconnected nodes

X_S in our sense. He refers to our concept of irrelevance as 'weak' epistemic irrelevance. Consequently, *Moral's results in Reference [76] are not applicable to all directed acyclic networks*. As a simple example: his concept of irrelevance does not allow for two variables to be mutually irrelevant, except in some degenerate uninformative cases. Therefore, his results cannot be applied to a network consisting of two unconnected nodes. More generally speaking, we believe that his results can only be applied to networks where every pair of nodes is connected by means of a directed path.

6.6 THE INDEPENDENT NATURAL EXTENSION

Let us now consider the special case of a credal network for which the underlying DAG has no edges or, equivalently, consists of disconnected nodes only; Figure 6.5 provides an example with four nodes. Every node has then neither parents nor descendants. Consequently, for every $s \in G$, the local model $\mathscr{D}_s$—or $\underline{P}_s$, $\mathscr{M}_s$ or $\mathscr{F}_s$—is unconditional and all the other nodes are non-parent non-descendants: $N(s) = G \setminus \{s\}$. Hence, in this particular case, it follows from Equation $(5.2)_{136}$ that

$$\mathscr{D}_G^{\mathrm{irr}} = \mathrm{posi}(\{\mathbb{I}_{\{z_{G\setminus\{s\}}\}} f \colon s \in G,\ z_{G\setminus\{s\}} \in \mathscr{X}_{G\setminus\{s\}},\ f \in \mathscr{D}_s\}).$$

The right-hand side of this equation is equal to the *independent natural extension*[19] of the local models $\mathscr{D}_s$, $s \in G$, denoted by $\otimes_{s\in G}\mathscr{D}_s$ [45, Equation (18)], which was defined in Reference [45] as the most conservative independent product of the local models $\mathscr{D}_s$, $s \in G$, meaning that it is the smallest coherent set $\mathscr{D}_G$ of desirable gambles on $\mathscr{X}_G$ such that $\mathrm{marg}_s(\mathscr{D}_G) = \mathscr{D}_s$ for all $s \in G$ and such that X_I is epistemically value-irrelevant to X_S for all disjoint subsets I and S of G. We conclude that for the particular case of a credal network that has no edges, the irrelevant natural extension $\mathscr{D}_G^{\mathrm{irr}}$ is equal to the independent natural extension $\otimes_{s\in G}\mathscr{D}_s$. Therefore, the properties that we have established—or will establish—for $\mathscr{D}_G^{\mathrm{irr}}$ can be translated—trivially—to properties of $\otimes_{s\in G}\mathscr{D}_s$.[20]

[19] Not to be confused with the notion of independent natural extension that was discussed in Section 5.5_{145}. However, given that the notion of independent natural extension that we consider here is a special case of the one in Section 5.5_{145}, there is no conflict in terminology.

[20] In fact, many of our results for $\mathscr{D}_G^{\mathrm{irr}}$ can be regarded as generalisations of results that were proved to hold for $\otimes_{s\in G}\mathscr{D}_s$ in Reference [45]. For example, Proposition 15 in Reference [45]

Since the ideas in Section 5.4.1$_{135}$ are basically an extension of those in Reference [45], this is of course not that surprising.

A similar connection can be established in terms of lower previsions. Again, consider a credal network that consists of disconnected nodes. For every $s \in G$, the local model is then a coherent lower prevision $\underline{P}_s$ on $\mathscr{G}(\mathscr{X}_s)$. Due to the coherence of the local lower previsions, we can assume without loss of generality that, for all $s \in G$, $\underline{P}_s = \underline{P}_{\mathscr{D}_s}$, with $\mathscr{D}_s$ a coherent set of desirable gambles on $\mathscr{X}_s$.

The *independent natural extension* $\otimes_{s\in G}\underline{P}_s$ of the local lower previsions $\underline{P}_s$, $s \in G$, is defined as the most conservative coherent lower prevision $\underline{P}_G$ on $\mathscr{G}(\mathscr{X}_G)$ that has $\underline{P}_s$ as its marginal models and that satisfies a specific set of independence assessments; see Reference [46] for an exact definition. Loosely speaking, X_I should be epistemically value-irrelevant to X_S for all disjoint subsets I and S of G. However, since $\underline{P}_G$—and $\otimes_{s\in G}\underline{P}_s$—is taken to be an unconditional lower prevision on $\mathscr{G}(\mathscr{X}_G)$ instead of a conditional lower prevision on $\mathscr{C}(\mathscr{X}_G)$, our notion of epistemic value-irrelevance needs to be replaced by a more cumbersome version.

In any case, for our present purposes, it suffices to know that $\otimes_{s\in G}\underline{P}_s$ is equal to $\underline{P}_{\otimes_{s\in G}\mathscr{D}_s}$ [45, Theorem 21]. Combined with Theorem 42$_{139}$ and the fact that $\mathscr{D}_G^{\mathrm{irr}} = \otimes_{s\in G}\mathscr{D}_s$, this implies that, for credal networks that consist of disconnected nodes, the unconditional part $\underline{P}_G^{\mathrm{irr}}(\cdot) := \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\mathscr{X}_G)$ of the irrelevant natural extension $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is equal to the independent natural extension $\otimes_{s\in G}\underline{P}_s$ of the local models $\underline{P}_s$, $s \in G$. Consequently, our results for $\underline{P}_G^{\mathrm{irr}}$ can be used—often trivially—to obtain new properties—or alternative proofs for 'old' properties—of $\otimes_{s\in G}\underline{P}_s$.[21] The following two properties are important examples. For all $s \in G$, consider a gamble $f_s \in \mathscr{G}(\mathscr{X}_s)$. Then

$$\otimes_{s\in G}\underline{P}_s\left(\sum_{s\in G} f_s\right) = \underline{P}_G^{\mathrm{irr}}\left(\sum_{s\in G} f_s\right) = \sum_{s\in G}\underline{P}_s(f_s)$$

and, if there is some $t \in G$ such that $f_s \geq 0$ for all $s \in G \setminus \{t\}$, then also

is a special case of Proposition 38$_{137}$, Proposition 17 and 18 in Reference [45] are both implied by Theorem 60$_{160}$, and by combining Theorem 41$_{138}$ with Corollary 61$_{161}$ we can generalise Theorem 19 in Reference [45]. Furthermore, the associativity result in Reference [45, Theorem 20] can be regarded as a special case of Proposition 67$_{\frown}$: it suffices to apply Proposition 67$_{\frown}$ to a DAG consisting of two separate, disconnected sub-DAGs, each of which consists of disconnected nodes only.

[21] For example, Proposition 13 in Reference [46] is basically a consequence of Corollary 44$_{141}$ and the fact that $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is coherent, Theorem 18 in Reference [46] is a consequence of Corollary 59$_{160}$ [with $h = 0$], and the (strong) factorisation and (strong) external additivity properties in Reference [46, Theorems 22 and 24 and Proposition 27] can be derived from Corollaries 57$_{159}$ and 59$_{160}$, respectively. Furthermore, the associativity result in Reference [46, Theorem 23] can be regarded as a special case of Proposition 67$_{\frown}$: it suffices to apply Proposition 67$_{\frown}$ to a DAG consisting of two separate, disconnected sub-DAGs, each of which consists of disconnected nodes only.

$$\otimes_{s\in G}\underline{P}_s\left(\prod_{s\in G} f_s\right) = \underline{P}_G^{\mathrm{irr}}\left(\prod_{s\in G} f_s\right) = \begin{cases} \underline{P}_t(f_t)\prod_{s\in G\setminus\{t\}}\underline{P}_s(f_s) & \text{if } \underline{P}_t(f_t)\geq 0; \\ \underline{P}_t(f_t)\prod_{s\in G\setminus\{t\}}\overline{P}_s(f_s) & \text{if } \underline{P}_t(f_t)\leq 0. \end{cases}$$

Due to Corollary 44_{141}, these properties are trivial consequences of Corollaries 57_{159} and 59_{160}. They are extremely closely related—almost identical—to properties that were referred to as external additivity and factorisation in Reference [46].

The definition of the independent natural extension can also be applied to 'local' models that depend on more than one variable. Consider a partition $G_1,\dots,G_m$ of G and, for every $i\in\{1,\dots,m\}$, a coherent set $\mathscr{D}_i$ of desirable gambles on $\mathscr{G}(\mathscr{X}_{G_i})$ or a coherent lower prevision $\underline{P}_i$ on $\mathscr{G}(\mathscr{X}_{G_i})$. Again, without loss of generality, we can assume that $\underline{P}_i = \underline{P}_{\mathscr{D}_i}$ for all $i\in\{1,\dots,m\}$. If we regard $\mathscr{D}_i$—or $\underline{P}_i$—as the local model of a variable $X_i := X_{G_i}$ that takes values in $\mathscr{X}_i := \mathscr{X}_{G_i}$, we can apply the same concepts as above. In this way, the independent natural extension of the sets of desirable gambles $\mathscr{D}_i$, $i\in\{1,\dots,m\}$, is equal to

$$\otimes_{i=1}^m\mathscr{D}_i := \operatorname{posi}\left(\{\mathbb{I}_{\{z_{G\setminus G_i}\}}f\colon i\in\{1,\dots,m\},\ z_{G\setminus G_i}\in\mathscr{X}_{G\setminus G_i},\ f\in\mathscr{D}_i\}\right)$$

and the independent natural extension $\otimes_{i=1}^m\underline{P}_i$ of the lower previsions $\underline{P}_i$, $i\in\{1,\dots,m\}$ is given by $\otimes_{i=1}^m\underline{P}_i := \underline{P}_{\otimes_{i=1}^m\mathscr{D}_i}$. Furthermore, if $f_i\in\mathscr{G}(\mathscr{X}_{G_i})$ for all $i\in\{1,\dots,m\}$, then

$$\otimes_{i=1}^m\underline{P}_i\left(\sum_{i=1}^m f_i\right) = \sum_{i=1}^m\underline{P}_i(f_i) \tag{6.8}$$

and, if there is some $j\in\{1,\dots,m\}$ such that $f_i\geq 0$ for all $i\neq j$ in $\{1,\dots,m\}$, then also

$$\otimes_{i=1}^m\underline{P}_i\left(\prod_{i=1}^m f_i\right) = \begin{cases} \underline{P}_j(f_j)\displaystyle\prod_{\substack{1\leq i\leq m\\ i\neq j}}\underline{P}_i(f_i) & \text{if } \underline{P}_j(f_j)\geq 0; \\ \underline{P}_j(f_j)\displaystyle\prod_{\substack{1\leq i\leq m\\ i\neq j}}\overline{P}_i(f_i) & \text{if } \underline{P}_j(f_j)\leq 0. \end{cases} \tag{6.9}$$

Our next two results show that the connection between the irrelevant natural extension of a credal network and this notion of independent natural extension goes much further than the simple case that was discussed in the beginning of this section. The following proposition is proved in Appendix $6.\mathrm{F}_{206}$.

Proposition 67. *Consider a partition $G_1,\dots,G_m$ of G such that $P(G_i)=\emptyset$ for all $i\in\{1,\dots,m\}$ or, equivalently, let the DAG of the complete network consist of m separate, disconnected sub-DAGs, each of which has G_i as its set of nodes, for $i\in\{1,\dots,m\}$. Then $\mathscr{D}_G^{\mathrm{irr}} = \otimes_{i=1}^m\mathscr{D}_{G_i}^{\mathrm{irr}}$ and $\underline{P}_G^{\mathrm{irr}}(\cdot) = \otimes_{i=1}^m\underline{P}_{G_i}^{\mathrm{irr}}(\cdot)$.*

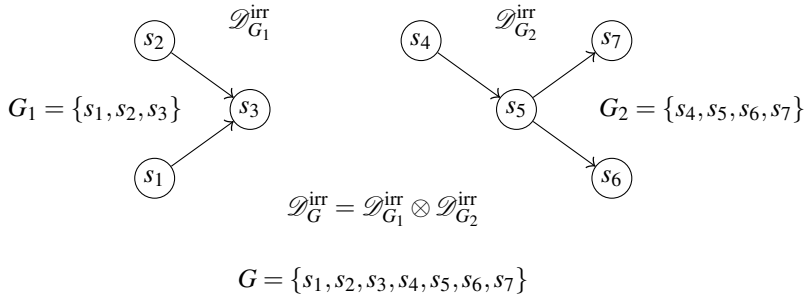


Figure 6.6: A simple illustration of Proposition $67_{\frown}$

Figure 6.6 illustrates this result with a simple example. It should be clear that the connections that were established earlier in this section—$\mathscr{D}^{\text{irr}}_G = \otimes_{s \in G} \mathscr{D}_s$ and $\underline{P}^{\text{irr}}_G = \otimes_{s \in G} \underline{P}_s$—correspond to special cases of Proposition $67_{\frown}$.[22] The following corollary generalises this proposition even further.

Corollary 68. *Consider a closed subset K of G and a partition $K_1, \ldots, K_m$ of K such that $P_K(K_i) = \emptyset$ for all $i \in \{1, \ldots, m\}$ or, equivalently, let the sub-DAG that corresponds to the set K consist of m separate, disconnected sub-DAGs, each of which has K_i as its set of nodes, for $i \in \{1, \ldots, m\}$. Then $P(K_i) \subseteq P(K)$ for all $i \in \{1, \ldots, m\}$ and, for all $x_{P(K)} \in \mathscr{X}_{P(K)}$: $\mathscr{D}^{\text{irr}}_{K \rfloor x_{P(K)}} = \otimes_{i=1}^m \mathscr{D}^{\text{irr}}_{K_i \rfloor x_{P(K_i)}}$ and $\underline{P}^{\text{irr}}_{K \rfloor x_{P(K)}}(\cdot) = \otimes_{i=1}^m \underline{P}^{\text{irr}}_{K_i \rfloor x_{P(K_i)}}(\cdot)$.*

Proof of Corollary 68. Fix $i \in \{1, \ldots, m\}$ and consider any $s \in P(K_i)$. Then there is some $q \in K_i \subseteq K$ such that $s \in P(q)$ and $s \notin K_i$. Due to our assumption that $P_K(K_i) = \emptyset$, this implies that $s \notin K$. Since $q \in K$, this allows us to infer that $s \in P(K)$. Hence, for all $i \in \{1, \ldots, m\}$, we have that $P(K_i) \subseteq P(K)$.

The rest of the proof is now a direct consequence of Proposition $67_{\frown}$. It suffices to apply Proposition $67_{\frown}$ to a credal network that has the sub-DAG associated with K as its graphical structure and whose local models are given by Equation $(6.1)_{155}$ or $(6.4)_{158}$. □

Figure $6.7_{\frown}$ illustrates this result with a simple example.

Results that are analogous to Proposition $67_{\frown}$ and Corollary 68 can also be obtained in terms of sets of linear previsions and sets of probability mass functions, in a trivial way, by defining $\otimes_{i=1}^m \mathscr{M}_i := \mathscr{M}_{\otimes_{i=1}^m \underline{P}_{\mathscr{M}_i}}$ and $\otimes_{i=1}^m \mathscr{F}_i := \mathscr{F}_{\otimes_{i=1}^m \mathscr{M}_{\mathscr{F}_i}} = \mathscr{F}_{\otimes_{i=1}^m \underline{P}_{\mathscr{F}_i}}$.

[22] Choose $m = |G|$ and let the sets G_i be singletons, each of which contains a different $s \in G$.

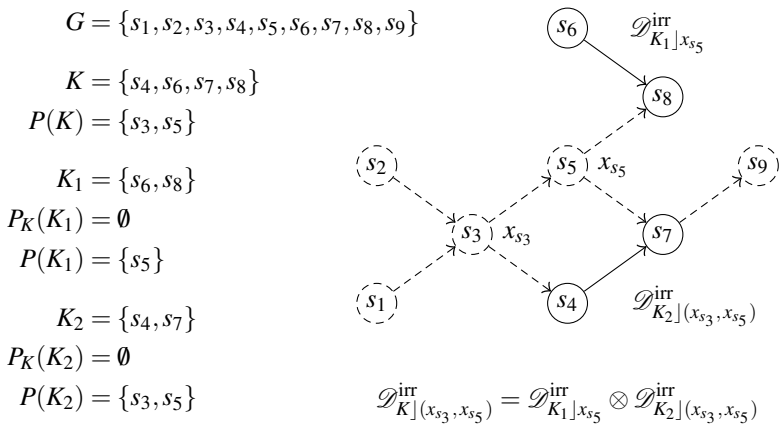


Figure 6.7: A simple illustration of Corollary 68$_{\frown}$

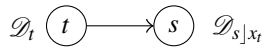


Figure 6.8: Example of a credal network with two nodes and one edge

6.7 A CONNECTION WITH MARGINAL EXTENSION

Another important special case are credal networks whose graphical structure consist of two nodes—$G = \{t, s\}$—and a single edge, as depicted in Figure 6.8. In this particular case, we use $\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$ as an alternative notation for $\mathscr{D}^{\mathrm{irr}}_G$. It follows from Proposition 36$_{136}$ that

$$\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t} = \Big\{ f \in \mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s) \setminus \{0\} \colon f = f_t + \textstyle\sum_{x_t \in \mathscr{X}_t} \mathbb{I}_{\{x_t\}} f_{s\rfloor x_t}, \quad f_t \in \mathscr{D}_t \cup \{0\}, (\forall x_t \in \mathscr{X}_t)\, f_{s\rfloor x_t} \in \mathscr{D}_{s\rfloor x_t} \cup \{0\} \Big\}. \tag{6.10}$$

In its current form, this expression does not seem all that useful. The reason why we introduce a special notation for it is because the corresponding lower prevision can be evaluated very easily.

Proposition 69. *Consider a coherent set $\mathscr{D}_t$ of desirable gambles on $\mathscr{X}_t$ and, for all $x_t \in \mathscr{X}_t$, a coherent set $\mathscr{D}_{s\rfloor x_t}$ of desirable gambles on $\mathscr{X}_s$. Let $\underline{P}_t := \underline{P}_{\mathscr{D}_t}$ and, for all $x_t \in \mathscr{X}_t$, let $\underline{P}_{s\rfloor x_t} := \underline{P}_{\mathscr{D}_{s\rfloor x_t}}$. It then holds for all $f \in \mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s)$ that*

$$\underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f) = \underline{P}_t(\underline{P}_{s\rfloor X_t}(f)),$$

where $\underline{P}_{s\rfloor X_t}(f)$ *is a gamble on* $\mathscr{X}_t$ *that is defined by*

$$\underline{P}_{s\rfloor X_t}(f)(x_t) := \underline{P}_{s\rfloor x_t}(f(x_t, X_s)) \text{ for all } x_t \in \mathscr{X}_t.$$

Proof of Proposition 69$_{\frown}$. Fix $f \in \mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s)$.

First consider any $\alpha \in \mathbb{R}$ such that $f - \alpha \in \mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$. It then follows from Equation (6.10)$_{\frown}$ that

$$f = \alpha + f_t + \sum_{x_t \in \mathscr{X}_t} \mathbb{I}_{\{x_t\}} f_{s\rfloor x_t},$$

where $f_t \in \mathscr{D}_t \cup \{0\}$ and, for all $x_t \in \mathscr{X}_t$, $f_{s\rfloor x_t} \in \mathscr{D}_{s\rfloor x_t} \cup \{0\}$. Since this implies that

$$\underline{P}_{s\rfloor X_t}(f) = \alpha + f_t + \sum_{x_t \in \mathscr{X}_t} \mathbb{I}_{\{x_t\}} \underline{P}_{s\rfloor x_t}(f_{s\rfloor x_t}) \geq \alpha + f_t,$$

it follows from the coherence of $\underline{P}_t$ that $\underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) \geq \alpha$. Since this is true for any $\alpha \in \mathbb{R}$ such that $f - \alpha \in \mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$, we infer from Equation (2.3)$_{40}$ that $\underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f) \leq \underline{P}_t(\underline{P}_{s\rfloor X_t}(f))$.

Next, consider any $\alpha \in \mathbb{R}$ such that $\underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f) < \alpha$. Equation (2.3)$_{40}$ then implies that $f - \alpha \notin \mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$. Fix $\varepsilon \in \mathbb{R}_{>0}$. Let $f_t := \underline{P}_{s\rfloor X_t}(f) - \alpha - \varepsilon$ and, for all $x_t \in \mathscr{X}_t$:

$$f_{s\rfloor x_t} := f(x_t, X_s) - \underline{P}_{s\rfloor x_t}(f(x_t, X_s)) + \varepsilon \in \mathscr{D}_{s\rfloor x_t},$$

where the inclusion follows from Equations (2.5)$_{41}$ and (2.6)$_{41}$ and the fact that $\underline{P}_{s\rfloor x_t}(f(x_t, X_s) - \underline{P}_{s\rfloor x_t}(f(x_t, X_s))) = 0$. Assume *ex absurdo* that $f_t \in \mathscr{D}_t$. Since

$$f - \alpha = \underline{P}_{s\rfloor X_t}(f) - \alpha - \varepsilon + f + \underline{P}_{s\rfloor X_t}(f) + \varepsilon = f_t + \sum_{x_t \in \mathscr{X}_t} \mathbb{I}_{\{x_t\}} f_{s\rfloor x_t},$$

Equation (6.10)$_{\frown}$ and the fact that $f - \alpha \notin \mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$ then allows us to infer that $f - \alpha = 0$. This implies that $\underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f) = \alpha$, a contradiction. Hence, we may conclude that $f_t \notin \mathscr{D}_t$ and therefore also that $\underline{P}_t(f_t) \leq 0$. Since

$$\underline{P}_t(f_t) = \underline{P}_t(\underline{P}_{s\rfloor X_t}(f) - \alpha - \varepsilon) = \underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) - \alpha - \varepsilon,$$

this implies that $\underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) \leq \alpha + \varepsilon$. Since this is true for every $\varepsilon \in \mathbb{R}_{>0}$, we infer that $\underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) \leq \alpha$. Since this inequality holds for every $\alpha \in \mathbb{R}$ such that $\underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f) < \alpha$, it follows that $\underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) \leq \underline{P}_{\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}}(f)$. □

Inspired by this result, we are led to introduce the lower prevision $\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}$ on $\mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s)$, defined by

$$\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}(f) := \underline{P}_t(\underline{P}_{s\rfloor X_t}(f)) \text{ for all } f \in \mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s). \tag{6.11}$$

It follows from Proposition 69_{175}, Theorem 42_{139} and our definition for $\mathscr{D}_t \oslash \mathscr{D}_{s\rfloor X_t}$ that $\underline{P}_t \oslash \underline{P}_{s\rfloor X_t} = \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor \mathscr{X}_G)$, which also implies that this lower prevision is coherent. The coherent lower prevision $\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}$ is called the *marginal extension* of $\underline{P}_t$ and $\underline{P}_{s\rfloor X_t}$ [70]. It was first considered by Walley [106, Section 6.7.2], who showed that it is the most conservative coherent lower prevision on $\mathscr{G}(\mathscr{X}_t \times \mathscr{X}_s)$ that is jointly coherent with the local models that are used to define it.

A similar concept can be introduced in terms of sets of linear previsions as well, simply by applying Equation $(6.11)_{\frown}$ elementwise. This leads us to define

$$\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t} := \{P_t \oslash P_{s\rfloor X_t} \colon P_t \in \mathscr{M}_t, (\forall x_t \in \mathscr{X}_t)\, P_{s\rfloor x_t} \in \mathscr{M}_{s\rfloor x_t}\} \tag{6.12}$$

The following result establishes that $\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$ is closely related to the corresponding concept for lower previsions.

Proposition 70. *Consider a coherent lower prevision $\underline{P}_t$ on $\mathscr{G}(\mathscr{X}_t)$ and, for all $x_t \in \mathscr{X}_t$, consider a coherent lower prevision $\underline{P}_{s\rfloor x_t}$ on $\mathscr{G}(\mathscr{X}_s)$. Let $\mathscr{M}_t := \mathscr{M}_{\underline{P}_t}$ and, for all $x_t \in \mathscr{X}_t$, let $\mathscr{M}_{s\rfloor x_t} := \mathscr{M}_{\underline{P}_{s\rfloor x_t}}$. It then holds that $\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t} = \mathscr{M}_{\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}}$.*[23]

Proof of Proposition 70. We only prove that $\mathscr{M}_{\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}} \subseteq \mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$; the other direction is trivial. So consider any $P_G \in \mathscr{M}_{\underline{P}_t \oslash \underline{P}_{s\rfloor X_t}}$. We will prove that $P_G \in \mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$.

Let $G := \{t, s\}$ and consider a coherent conditional lower prevision $\underline{P}_G^*(\cdot|\cdot)$ on $\mathscr{C}(\mathscr{X}_G)$ such that $\underline{P}_G^*(\cdot) = \underline{P}_t(\underline{P}_{s\rfloor X_t}(\cdot))$, $\underline{P}_t^*(\cdot) = \underline{P}_t(\cdot)$ and, for all $x_t \in \mathscr{X}_t$, $\underline{P}_s^*(\cdot\rfloor x_t) = \underline{P}_t(\cdot)$ [it was proved by Walley that this is always possible [106, Section 6.7.2]; see Reference [70, Theorem 1] as well]. It then follows from Proposition 8_{53} that there is some $P_G^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$ such that $P_G^*(\cdot) = P_G(\cdot)$. Since $P_G^*(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}_G^*(\cdot\rfloor\cdot)}$, we find that $P_t^* \in \mathscr{M}_{\underline{P}_t}$ and that, for all $x_t \in \mathscr{X}_t$, $P^*_{t\rfloor x_t}(\cdot) := P_s^*(\cdot\rfloor x_t) \in \mathscr{M}_{\underline{P}_{s\rfloor x_t}}$. Since the law of iterated prevision [see Section $4.2.6_{113}$] tells us that $P_G^*(\cdot) = P_t^*(P_s^*(\cdot\rfloor X_t)) = P_t^*(P^*_{s\rfloor X_t}(\cdot))$, we find that $P_G^*(\cdot) \in \mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$ and therefore also that $P_G \in \mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$. □

Combined with Theorem 46_{142} and the fact that $\underline{P}_t \oslash \underline{P}_{s\rfloor X_t} = \underline{P}_G^{\mathrm{irr}}(\cdot\rfloor \mathscr{X}_G)$, this result implies that $\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t} = \mathscr{M}_G^{\mathrm{irr}}\rfloor \mathscr{X}_G$.

Finally, we let $\mathscr{F}_t \oslash \mathscr{F}_{s\rfloor X_t}$ be the set of all probability mass functions p on $\mathscr{X}_t \times \mathscr{X}_s$ for which there are $p_t \in \mathscr{F}_t$ and, for all $x_t \in \mathscr{X}_t$, $p_{s\rfloor x_t} \in \mathscr{F}_{s\rfloor x_t}$ such that

$$p(z_t, z_s) = p_t(z_t)\, p_{s\rfloor z_t}(z_s) \ \text{ for all } (z_t, z_s) \in \mathscr{X}_t \times \mathscr{X}_s. \tag{6.13}$$

[23] A weaker version of this result was already proved by Walley [106, Section 6.7.4]. He showed that $\underline{P}_t \oslash \underline{P}_{s\rfloor X_t} = \underline{P}_{\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}}$; a direct consequence of our result.

It should be clear that if $\mathscr{F}_t = \mathscr{F}_{\mathscr{M}_t}$ and, for all $x_t \in \mathscr{X}_t$, $\mathscr{F}_{s\rfloor x_t} = \mathscr{F}_{\mathscr{M}_{s\rfloor x_t}}$, then $\mathscr{F}_t \oslash \mathscr{F}_{s\rfloor X_t}$ is the set of probability mass functions that corresponds to $\mathscr{M}_t \oslash \mathscr{M}_{s\rfloor X_t}$. Consequently, it follows from Proposition 49_{143} that—again in the particular case of Figure 6.8_{175}—$\mathscr{F}_t \oslash \mathscr{F}_{s\rfloor X_t} := \mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$. Equation $(6.13)_{\cap}$ also clearly shows that, in essence, marginal extension is just an elementwise constructive application of Bayes's rule.

We now extend these concepts to a more general setting, where the 'local' models are multivariate themselves. Instead of restricting attention to the case $G = \{t, s\}$, we consider an arbitrary set of nodes G, a binary partition T, S of G and a subset C of T.

We start with the framework of sets of desirable gambles. Let $\mathscr{D}_T$ be a coherent set of desirable gambles on $\mathscr{X}_T$ and, for all $x_C \in \mathscr{X}_C$, let $\mathscr{D}_{S\rfloor x_C}$ be a coherent set of desirable gambles on $\mathscr{X}_S$. Then

$$\mathscr{D}_T \oslash \mathscr{D}_{S\rfloor X_C} := \Big\{ f \in \mathscr{G}(\mathscr{X}_G)) \setminus \{0\} \colon f = f_T + \textstyle\sum_{x_T \in \mathscr{X}_T} \mathbb{I}_{\{x_T\}} f_{S\rfloor x_T}, \\ f_T \in \mathscr{D}_T \cup \{0\}, (\forall x_T \in \mathscr{X}_T)\, f_{S\rfloor x_T} \in \mathscr{D}_{S\rfloor x_C} \cup \{0\} \Big\}.$$

is a coherent set of desirable gambles on $\mathscr{X}_G$. Although it might not seem so at first sight, this is just a special case of Equation $(6.10)_{175}$. This should be obvious if $C = T$; it then suffices to identify X_t and X_s with X_T and X_S, respectively. For $C \subset T$, we also need to identify $\mathscr{D}_{S\rfloor X_C}$ with $\mathscr{D}_{S\rfloor X_T}$: for all $x_T \in \mathscr{X}_T$, we let $\mathscr{D}_{S\rfloor x_T} := \mathscr{D}_{S\rfloor x_C}$, where x_C is fully determined by x_T because C is a subset of T.

For the framework of lower previsions, as in the simple case above, this turns into a very convenient expression. Consider a coherent lower prevision $\underline{P}_T$ on $\mathscr{G}(\mathscr{X}_T)$ and, for all $x_C \in \mathscr{X}_C$, a coherent lower prevision $\underline{P}_{S\rfloor x_C}$ on $\mathscr{G}(\mathscr{X}_S)$. For all $f \in \mathscr{G}(\mathscr{X}_G)$, we then let

$$\underline{P}_T \oslash \underline{P}_{S\rfloor X_C}(f) := \underline{P}_T(\underline{P}_{S\rfloor X_C}(f)), \tag{6.14}$$

where—more or less as in Section $4.2.6_{113}$—$\underline{P}_{S\rfloor X_C}(f)$ is a gamble on $\mathscr{X}_T$, defined by

$$(\underline{P}_{S\rfloor X_C}(f))(x_T) := \underline{P}_{S\rfloor x_C}(f(x_T, X_S)) \text{ for all } x_T \in \mathscr{X}_T.$$

The operator $\underline{P}_T \oslash \underline{P}_{S\rfloor X_C}$ is a coherent lower prevision on $\mathscr{G}(\mathscr{X}_G)$ and—similarly to what we found for $\mathscr{D}_T \oslash \mathscr{D}_{S\rfloor X_C}$—is a special case of the notion of marginal extension that was considered above,[24] with $X_t := X_T$, $X_s := X_S$ and, for all $x_T \in \mathscr{X}_T$, $\underline{P}_{S\rfloor x_T} := \underline{P}_{S\rfloor x_C}$. Proposition 69_{175} can therefore be applied to this special case as well: if $\underline{P}_T = \underline{P}_{\mathscr{D}_T}$ and, for all $x_C \in \mathscr{X}_C$, $\underline{P}_{S\rfloor x_C} = \underline{P}_{\mathscr{D}_{S\rfloor x_C}}$, then $\underline{P}_T \oslash \underline{P}_{S\rfloor X_C} = \underline{P}_{\mathscr{D}_T \oslash \mathscr{D}_{S\rfloor X_C}}$.

[24] Also, if $C = \emptyset$, then $\underline{P}_T \oslash \underline{P}_S$ is equal to the forward irrelevant natural extension of $\underline{P}_T$ and $\underline{P}_S$, as studied in Reference [44].

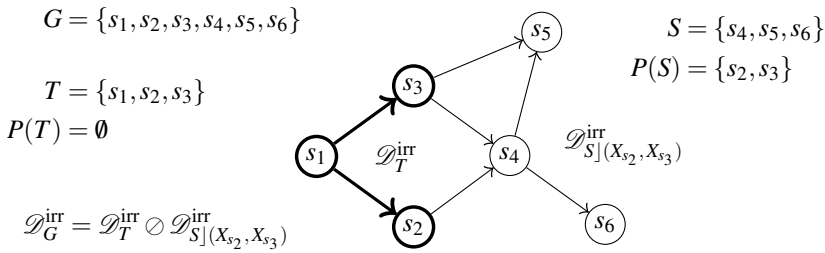


Figure 6.9: A simple illustration of Proposition 71

A similar concept can be introduced for sets of linear previsions. Consider a set $\mathscr{M}_T$ of linear previsions on $\mathscr{G}(\mathscr{X}_T)$ and, for all $x_C \in \mathscr{X}_C$, a set $\mathscr{M}_{S\rfloor x_C}$ of linear previsions on $\mathscr{G}(\mathscr{X}_S)$. We then define

$$\mathscr{M}_T \oslash \mathscr{M}_{S\rfloor X_C} := \{P_T \oslash P_{S\rfloor X_T} : P_T \in \mathscr{M}_T, (\forall x_T \in \mathscr{X}_T) P_{S\rfloor x_T} \in \mathscr{M}_{S\rfloor x_C}\}.$$

Since this is again a special case of the simple version that was discussed above, Proposition 70_{177} applies here as well: if $\mathscr{M}_T = \mathscr{M}_{\underline{P}_T}$ and, for all $x_C \in \mathscr{X}_C$, $\mathscr{M}_{S\rfloor x_C} = \mathscr{M}_{\underline{P}_{S\rfloor x_C}}$, then $\mathscr{M}_T \oslash \mathscr{M}_{S\rfloor X_C} := \mathscr{M}_{\underline{P}_T \oslash \underline{P}_{S\rfloor X_C}}$.

Finally, we let $\mathscr{F}_T \oslash \mathscr{F}_{S\rfloor X_C}$ be the set of all probability mass functions p_G on $\mathscr{X}_G$ for which there are $p_T \in \mathscr{F}_T$ and, for all $x_T \in \mathscr{X}_T$, $p_{S\rfloor x_T} \in \mathscr{F}_{S\rfloor x_C}$ such that

$$p_G(z_G) = p_T(z_T) p_{S\rfloor z_T}(z_S) \text{ for all } z_G \in \mathscr{X}_G.$$

If $\mathscr{F}_T = \mathscr{F}_{\mathscr{M}_T}$ and, for all $x_C \in \mathscr{X}_C$, $\mathscr{F}_{S\rfloor x_C} = \mathscr{F}_{\mathscr{M}_{S\rfloor x_C}}$, then $\mathscr{F}_T \oslash \mathscr{F}_{S\rfloor X_C}$ is the set of probability mass functions that corresponds to $\mathscr{M}_T \oslash \mathscr{M}_{S\rfloor X_C}$.

Our next result shows that the connection between credal networks under epistemic irrelevance and the concept of marginal extension—and its generalisations—goes far beyond the simple case that was discussed in the beginning of this section. The proof of this proposition can be found in Appendix $6.G_{207}$.

Proposition 71. *Consider a set $S \subseteq G$ such that $T := G \setminus S$ is an ancestral set—$P(T) = \emptyset$—and such that, for all $t \in T$, $S \subseteq D(t)$. Then $\mathscr{D}_G^{\mathrm{irr}} = \mathscr{D}_T^{\mathrm{irr}} \oslash \mathscr{D}_{S\rfloor X_{P(S)}}^{\mathrm{irr}}$ and therefore also*

$$\underline{P}_G^{\mathrm{irr}}(\cdot) = \underline{P}_T^{\mathrm{irr}}(\cdot) \oslash \underline{P}_{S\rfloor X_{P(S)}}^{\mathrm{irr}}(\cdot)$$

and

$$\mathscr{M}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G = (\mathscr{M}_T^{\mathrm{irr}} \rfloor \mathscr{X}_T) \oslash (\mathscr{M}_{S\rfloor X_{P(S)}}^{\mathrm{irr}} \rfloor \mathscr{X}_S)$$

and

$$\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G = (\mathscr{F}_T^{\mathrm{irr}} \rfloor \mathscr{X}_T) \oslash (\mathscr{F}_{S\rfloor X_{P(S)}}^{\mathrm{irr}} \rfloor \mathscr{X}_S).$$

The example in Figure 6.9$_{\frown}$ provides a simple illustration. The following corollary establishes a similar result for sub-DAGs; see Figure 6.10$_{\frown}$ for an example.

Corollary 72. *Let K be a closed subset of G and consider a binary partition S,T of K such that $P_K(T) = \emptyset$ and, for all $t \in T$, $S \subseteq D_K(t)$. Then $P(S) \setminus P_K(S) \subseteq P(K)$ and, for all $x_{P(K)} \in \mathscr{X}_{P(K)}$:*

$$\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}} = \mathscr{D}^{\text{irr}}_{T\rfloor x_{P(T)}} \oslash \mathscr{D}^{\text{irr}}_{S\rfloor (x_{P(S)\setminus P_K(S)}, X_{P_K(S)})}$$

and

$$\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot) = \underline{P}^{\text{irr}}_{T\rfloor x_{P(T)}}(\cdot) \oslash \underline{P}^{\text{irr}}_{S\rfloor (x_{P(S)\setminus P_K(S)}, X_{P_K(S)})}(\cdot)$$

and

$$\mathscr{M}^{\text{irr}}_{K\rfloor x_{P(K)}} \rfloor \mathscr{X}_K = (\mathscr{M}^{\text{irr}}_{T\rfloor x_{P(T)}} \rfloor \mathscr{X}_T) \oslash (\mathscr{M}^{\text{irr}}_{S\rfloor (x_{P(S)\setminus P_K(S)}, X_{P_K(S)})} \rfloor \mathscr{X}_S)$$

and

$$\mathscr{F}^{\text{irr}}_{K\rfloor x_{P(K)}} \rfloor \mathscr{X}_K = (\mathscr{F}^{\text{irr}}_{T\rfloor x_{P(T)}} \rfloor \mathscr{X}_T) \oslash (\mathscr{F}^{\text{irr}}_{S\rfloor (x_{P(S)\setminus P_K(S)}, X_{P_K(S)})} \rfloor \mathscr{X}_S).$$

Proof of Corollary 72. Consider any $t \in P(S) \setminus P_K(S)$. Since $t \in P(S)$, there is some $s \in S$ such that $t \in P(s) \setminus S$. Therefore, we can infer from $t \notin P_K(S)$ that $t \notin K$, which implies that $t \in P(s) \setminus K$. Since $s \in S \subseteq K$, it follows that $t \in P(K)$. Hence, we find that $P(S) \setminus P_K(S) \subseteq P(K)$.

The rest of the proof is now a direct consequence of Proposition 71$_{\frown}$. It suffices to apply Proposition 71$_{\frown}$ to a credal network that has the sub-DAG associated with K as its graphical structure and whose local models are given by Equation (6.1)$_{155}$, (6.4)$_{158}$, (6.5)$_{158}$ or (6.6)$_{158}$. □

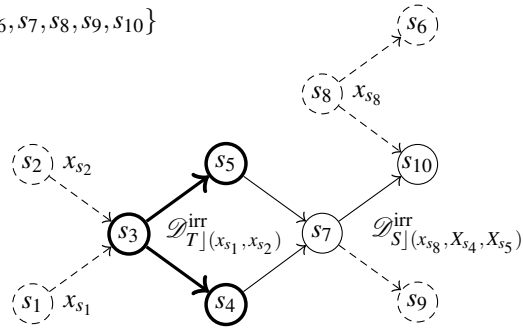
6.8 WHAT IF WE APPLY REGULAR EXTENSION?

We conclude this chapter by taking a quick look at the regular extension $\underline{R}^{\text{irr}}_G(\cdot\rfloor\cdot)$ that corresponds to $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$, in the sense of Section 3.4.7$_{90}$. Since $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$ is coherent and therefore coincides with its natural extension, we know from Equation (3.10)$_{90}$ that $\underline{R}^{\text{irr}}_G(\cdot\rfloor\cdot)$ is defined by

$$\underline{R}^{\text{irr}}_G(f\rfloor O) := \begin{cases} \underline{R}^{\text{irr}*}_G(f\rfloor O) & \text{if } \overline{P}^{\text{irr}}_G(O) > 0 \\ \underline{P}^{\text{irr}}_G(f\rfloor O) & \text{if } \overline{P}^{\text{irr}}_G(O) = 0 \end{cases} \quad \text{for all } (f,O) \in \mathscr{C}(\mathscr{X}_G), \quad (6.15)$$

where $\underline{R}^{\text{irr}*}_G(\cdot\rfloor\cdot)$ is the regular extension of the unconditional lower prevision $\underline{P}^{\text{irr}}_G(\cdot) := \underline{P}^{\text{irr}}_G(\cdot\rfloor\mathscr{X}_G)$, as defined by Equation (2.29)$_{63}$. The following result shows that $\underline{R}^{\text{irr}}_G(\cdot\rfloor\cdot)$ satisfies a marginalisation property that is very similar to that of $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$ [see Corollary 60$_{160}$]. The proof is rather elaborate and has therefore been moved to Appendix 6.H$_{208}$.

$$G = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}$$
$$K = \{s_3, s_4, s_5, s_7, s_{10}\}$$
$$P(K) = \{s_1, s_2, s_8\}$$
$$T = \{s_3, s_4, s_5\}$$
$$P_K(T) = \emptyset$$
$$P(T) = \{s_1, s_2\}$$
$$S = \{s_7, s_{10}\}$$
$$P_K(S) = \{s_4, s_5\}$$
$$P(S) = \{s_4, s_5, s_8\}$$



$$\mathscr{D}^{\text{irr}}_{K\rfloor(x_{s_1},x_{s_2},x_{s_8})} = \mathscr{D}^{\text{irr}}_{T\rfloor(x_{s_1},x_{s_2})} \oslash \mathscr{D}^{\text{irr}}_{S\rfloor(x_{s_8},X_{s_4},X_{s_5})}$$

Figure 6.10: A simple illustration of Corollary 72$_{\curvearrowleft}$

Theorem 73. *Let K be a closed subset of G and consider any $x_{P(K)} \in \mathscr{X}_{P(K)}$ and $B_{N(K)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(K)})$. Then*

$$\underline{R}^{\text{irr}}_K(\cdot\rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \begin{cases} \underline{R}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot) & \text{if } \overline{P}^{\text{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) > 0; \\ \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot) & \text{if } \overline{P}^{\text{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) = 0. \end{cases}$$

This result simplifies if we are only interested in the 'unconditional' part of $\underline{R}^{\text{irr}}_K(\cdot\rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)})$. In that case, we find that $\underline{R}^{\text{irr}}_K(\cdot\rfloor \{x_{P(K)}\} \times B_{N(K)})$ is equal to $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot)$.

Corollary 74. *Let K be a closed subset of G. Then for every $x_{P(K)} \in \mathscr{X}_{P(K)}$ and $B_{N(K)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(K)})$, it holds that $\underline{R}^{\text{irr}}_K(\cdot\rfloor \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot)$.*

Proof of Corollary 74. Since the unconditional parts of $\underline{R}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$ and $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$ are trivially equal, this follows directly from Theorem 73. □

By applying this property to sets of the form $K = \{s\}$, we find that regular extension recovers the local models that were used to construct $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$.

Corollary 75. *Consider any $s \in G$, $x_{P(s)} \in \mathscr{X}_{P(s)}$ and $B_{N(s)} \in \mathscr{P}_{\emptyset}(\mathscr{X}_{N(s)})$. Then $\underline{R}^{\text{irr}}_s(\cdot\rfloor \{x_{P(s)}\} \times B_{N(s)}) = \underline{R}^{\text{irr}}_s(\cdot\rfloor x_{P(s)}) = \underline{P}_{s\rfloor x_{P(s)}}$.*

Proof of Corollary 75. Since Corollary 44$_{141}$ implies that $\underline{P}^{\text{irr}}_{s\rfloor x_{P(s)}}(\cdot) = \underline{P}_{s\rfloor x_{P(s)}}$, this is an immediate consequence of Corollary 74 [with $K := \{s\}$]. □

6.A TECHNICAL LEMMAS ABOUT PROPERTIES OF DAGS

Lemma 76. *Fix any $K \subseteq G$ and $k \in K$. Then $P_K(k) = P(k) \cap K = P(k) \setminus P(K)$ and $P(k) \setminus P_K(k) = P(k) \cap P(K)$.*

Proof of Lemma 76. We start by proving that $P_K(k) = P(k) \cap K$. An element $q \in P_K(k)$ is by definition a parent of k according to the sub-DAG that corresponds to K, therefore q is also a parent of k in the original DAG: $q \in P(k)$. Since q is an element of the sub-DAG, we have $q \in K$ and therefore $q \in P(k) \cap K$. Conversely, if $q \in P(k) \cap K$, then q is clearly a parent of k in the sub-DAG that corresponds to K and therefore $q \in P_K(k)$.

Next we show that $P(k) \cap K = P(k) \setminus P(K)$. By definition of $P(K)$ and since $k \in K$, we know that $q \in P(k)$ implies $q \in P(K) \cup K$ and we therefore have that $P(k) \subseteq P(K) \cup K$. Since $P(K)$ and K are disjoint by definition, we infer that $P(k) \cap K = P(k) \setminus P(K)$.

The final property is a direct consequence of the previous equality:

$$P(k) \setminus P_K(k) = P(k) \setminus (P(k) \setminus P(K)) = P(k) \cap P(K). \qquad \square$$

Lemma 77. *Consider any $K \subseteq G$, $k \in K$, $s \in G$ and $t \in PN(s)$. Then the following statements hold:*

(i) $P(t) \subseteq PN(s)$;

(ii) $N(K) = \bigcap_{q \in K} N(q)$;

(iii) $N(K) \subseteq N(k)$;

(iv) $A(K) = \emptyset$ *if and only if* $P(K) = \emptyset$;

(v) *If* $P(K) = \emptyset$, *then* K *is a closed set;*

(vi) *If* $P(K) = \emptyset$, *then* $G \setminus K$ *is a closed set and* $D(G \setminus K) = \emptyset$.

(vii) *If* $D(K) = \emptyset$, *then* $G \setminus K \subseteq PN(k)$.

Proof of Lemma 77. (i). Consider any $q \in P(t)$. By the definition of $PN(s)$, it suffices to show that $q \notin \{s\} \cup D(s)$. Assume *ex absurdo* that $s \sqsubseteq q$, then we derive from $q \sqsubset t$ (since $q \in P(t)$) that $s \sqsubset t$, meaning that $t \in D(s)$, contradicting $t \in PN(s)$.

(ii). This follows at once from:

$$\begin{aligned} N(K) := G \setminus (P(K) \cup K \cup D(K)) &= G \setminus \left(\bigcup_{q \in K} P(q) \cup K \cup \bigcup_{q \in K} D(q) \right) \\ &= G \setminus \left(\bigcup_{q \in K} (P(q) \cup \{q\} \cup D(q)) \right) \\ &= \bigcap_{q \in K} (G \setminus (P(q) \cup \{q\} \cup D(q)) = \bigcap_{q \in K} N(q). \end{aligned}$$

(iii)$_{\frown}$. Trivial due to (ii)$_{\frown}$.

(iv)$_{\frown}$. Since $P(K) \subseteq A(K)$, it suffices to prove that $P(K) = \emptyset$ implies $A(K) = \emptyset$ or, equivalently, that $A(K) \neq \emptyset$ implies $P(K) \neq \emptyset$. So, assume that $A(K) \neq \emptyset$. This means that there is some $s \in K$ such that $A(s) \setminus K \neq \emptyset$, which in turn implies that there is some $t \sqsubset s$ such that $t \notin K$. This means that there is a directed sequence of nodes $t = r_1, \ldots, r_i, \ldots, r_n = s$, $n > 1$, in G. Let i be the first index in $\{1, \ldots, n\}$ for which $r_i \in K$. Since $s \in K$, we know that i always exists and, since $t \notin K$, we know that $i > 1$. Since $r_{i-1} \in P(r_i)$ and $r_{i-1} \notin K$, we find that $P(r_i) \setminus K \neq \emptyset$ and therefore, since $r_i \in K$, that $P(K) \neq \emptyset$.

(v)$_{\frown}$. Assume that $P(K) = \emptyset$. We then know from (iv)$_{\frown}$ that $A(K) = \emptyset$. Consider now any $s, t \in K$ and any $k \in G$ such that $s \sqsubseteq k \sqsubseteq t$. We need to prove that $k \in K$. If $k = t$, this is trivial. Otherwise, we know that $k \sqsubset t$, which implies that $k \in A(t)$ and therefore, since $t \in K$, that $k \in A(K) \cup K$. Since $A(K) = \emptyset$, this implies that $k \in K$.

(vi)$_{\frown}$. Assume that $P(K) = \emptyset$. We then know from (iv)$_{\frown}$ that $A(K) = \emptyset$.

We first prove that $G \setminus K$ is a closed set. So consider any $s, t \in G \setminus K$ and any $k \in G$ such that $s \sqsubseteq k \sqsubseteq t$. We need to prove that $k \in G \setminus K$. Assume *ex absurdo* that $k \in K$. Since $s \sqsubseteq k$ and $s \notin K$, this implies that $s \in A(K)$. This is a contradiction because $A(K) = \emptyset$.

Next, we prove that $D(G \setminus K) = \emptyset$. Assume *ex absurdo* that $D(G \setminus K) \neq \emptyset$. This implies that there is some $k \in K$ and $s \in G \setminus K$ such that $s \sqsubseteq k$ and therefore $s \in A(k)$. This is a contradiction because $A(K) = \emptyset$.

(vii)$_{\frown}$. Assume that $D(K) = \emptyset$. This implies that $D(k) \subseteq K$ and therefore also that $\{k\} \cup D(k) \subseteq K$. Since $PN(k) = G \setminus (\{k\} \cup D(k))$, this in turn implies that $G \setminus K \subseteq PN(k)$. □

Lemma 78. *Consider any closed $K \subseteq G$ and any $k \in K$. Then the following statements hold:*

(i) $P(K) \cap D(K) = \emptyset$;

(ii) $P(K)$, $N(K)$, K *and* $D(K)$ *constitute a partition*[25] *of* G;

(iii) $PN(K)$ *and* $D(K)$ *are closed subsets of* G;

(iv) $P(PN(K)) = \emptyset$;

(v) $PN(D(K)) = PN(K) \cup K$;

(vi) $P(K) \subseteq PN(k)$;

(vii) $PN(K) \subseteq PN(k)$;

[25] We use the term 'partition' in a somewhat looser sense than is usual, as we do not exclude that some of its elements may be empty.

(viii) $P(K) \setminus P(k) \subseteq N(k)$.

Proof of Lemma 78$_{\frown}$. (i)$_{\frown}$. Assume *ex absurdo* that $q \in P(K)$ and $q \in D(K)$. Then $q \in D(K)$ implies the existence of some $r_1 \in K$ such that $r_1 \sqsubseteq q$ and $q \in P(K)$ implies the existence of some $r_2 \in K$ such that $q \sqsubseteq r_2$. We find that $r_1 \sqsubseteq q \sqsubseteq r_2$, with $r_1, r_2 \in K$. Since K is closed, this implies that $q \in K$, contradicting both $q \in P(K)$ and $q \in D(K)$.

(ii)$_{\frown}$. Direct consequence of (i)$_{\frown}$ and the definition of $P(K)$, $D(K)$ and $N(K)$.

(iii)$_{\frown}$. To prove that $PN(K)$ is closed, consider $q_1, q_2 \in PN(K)$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$ and assume *ex absurdo* that $r \notin PN(K)$, implying, due to (ii)$_{\frown}$, that $r \in K \cup D(K)$. This in turn implies that there is some $u \in K$ such that $u \sqsubseteq r$ and therefore $u \sqsubseteq q_2$, which implies that $q_2 \in K \cup D(K)$, contradicting $q_2 \in PN(K)$ due to (ii)$_{\frown}$.

To prove that $D(K)$ is closed, consider $q_1, q_2 \in D(K)$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$ and assume *ex absurdo* that $r \notin D(K)$. $q_1 \in D(K)$ implies that there is some $u \in K$ such that $u \sqsubseteq q_1$ and therefore $u \sqsubseteq r$, implying that $r \in K \cup D(K)$ and, since $r \notin D(K)$, that $r \in K$. We thus find that $u \sqsubseteq q_1 \sqsubseteq r$, with $u, r \in K$. Because K is closed, this tells us that $q_1 \in K$, contradicting $q_1 \in D(K)$.

(iv)$_{\frown}$. Assume *ex absurdo* that $q \in P(PN(K))$, so there is some $r \in PN(K)$ such that $q \in P(r)$. By definition of $P(PN(K))$, this implies that $q \notin PN(K)$, which in turn implies, due to (ii)$_{\frown}$, that $q \in K \cup D(K)$. By definition of $D(K)$, this implies that there is some $u \in K$ such that $u \sqsubseteq q$. Since $q \sqsubset r$ (because $q \in P(r)$), we find that $u \sqsubset r$, implying that $r \in K \cup D(K)$. Due to (ii)$_{\frown}$, this contradicts $r \in PN(K)$.

(v)$_{\frown}$. First notice that it suffices to show that $D(D(K)) = \emptyset$. Indeed, this implies $PN(D(K)) = G \setminus D(K) = PN(K) \cup K$ by applying (ii)$_{\frown}$ once for the closed $D(K)$ and once for the closed K. So assume *ex absurdo* that $q \in D(D(K))$, implying that there is some $r \in D(K)$ such that $r \sqsubset q$. Since $r \in D(K)$ in turn implies that there is some $u \in K$ such that $u \sqsubset r$, we find that $u \sqsubset q$, implying that $q \in K \cup D(K)$. But $q \notin D(K)$ because we know that $q \in D(D(K))$, and therefore $q \in K$. Since $u, q \in K$ and $u \sqsubset r \sqsubset q$, we derive from K being closed that $r \in K$, contradicting $r \in D(K)$.

(vi)$_{\frown}$. Choose $q \in P(K)$ and assume, *ex absurdo*, that $q \notin PN(k)$. This implies that $q \in \{k\} \cup D(k)$, or equivalently, that $k \sqsubseteq q$, and therefore that $q \in K \cup D(K)$, contradicting $q \in P(K)$ because of (ii)$_{\frown}$.

(vii)$_{\frown}$. Direct consequence of (vi)$_{\frown}$ and Lemma 77(iii)$_{182}$.

(viii). Choose $q \in P(K) \setminus P(k)$ and assume *ex absurdo* that $q \notin N(k)$, implying that $q \in P(k) \cup \{k\} \cup D(k)$ or, since $q \notin P(k)$, that $q \in \{k\} \cup D(k)$ and therefore $k \sqsubseteq q$. This in turn implies that $q \in K \cup D(K)$, contradicting $q \in P(K)$ because of (ii)$_{\frown}$. □

Lemma 79. *Consider any closed set $K \subseteq G$ and any $k \in K$, $s \in G \setminus K$ and $t \in PN(s) \cap K$. Then the following statements hold:*

(i) $D_K(k) = D(k) \cap K$;

(ii) $N_K(k) = N(k) \cap K$;

(iii) $PN_K(k) = PN(k) \cap K$;

(iv) $N(K)$, $P(K) \setminus P(k)$ *and* $N_K(k)$ *are pairwise disjoint subsets of* $N(k)$;

(v) $P_K(t) \subseteq PN(s) \cap K$;

(vi) $P(t) \setminus P_K(t) \subseteq PN(s) \cap P(K)$.

Proof of Lemma 79$_{\frown}$. (i). An element $q \in D_K(k)$ is by definition a descendant of k according to the sub-DAG that corresponds to K, therefore q is also a descendant of k in the original DAG: $q \in D(k)$. Since q is an element of the sub-DAG, we have $q \in K$ and therefore $q \in D(k) \cap K$.

Conversely, if $q \in D(k) \cap K$, then $q \in D(k)$ implies the existence of a directed sequence of nodes $k = r_1, \ldots, r_n = q$, $n \geq 1$, in G. Since $k \in K$ and $q \in K$, we can derive from K being closed that for all $i \in \{1, \ldots, n\}$, $r_i \in K$, implying that in the sub-DAG that corresponds to K, q is also a descendant of k: $q \in D_K(k)$.

(ii). This follows at once from the definitions of $N_K(k)$ and $N(k)$:

$$\begin{aligned} N_K(k) = K \setminus (P_K(k) \cup \{k\} \cup D_K(k)) &= K \setminus ((P(k) \cap K) \cup \{k\} \cup (D(k) \cap K)) \\ &= K \setminus (P(k) \cup \{k\} \cup D(k)) = N(k), \end{aligned}$$

where the second equality is due to Lemma 76_{182} and (i).

(iii). This follows at once from the definitions of $PN_K(k)$ and $PN(k)$:

$$\begin{aligned} PN_K(k) = P_K(k) \cup N_K(k) &= (P(k) \cap K) \cup (N(k) \cap K) \\ &= (P(k) \cup N(k)) \cap K =: PN(k) \cap K, \end{aligned}$$

where the second equality is due to Lemma 76_{182} and (ii).

(iv). $N(K)$, $P(K)$ and K are pairwise disjoint subsets of G because of Lemma 78(ii)$_{183}$. Since $P(K) \setminus P(k) \subseteq P(K)$ and $N_K(k) \subseteq K$, this implies that $N(K)$, $P(K) \setminus P(k)$ and $N_K(k)$ are pairwise disjoint as well. It only remains to show that $N(K)$, $P(K) \setminus P(k)$ and $N_K(k)$ are subsets of $N(k)$. For $N(K)$, this is due to Lemma 77(iii)$_{182}$. For $P(K) \setminus P(k)$, this is due to Lemma 78(vi)$_{183}$ and for $N_K(k)$, this follows from (ii).

(v). By definition, $P_K(t) \subseteq K$. To show that $P_K(t) \subseteq PN(s)$, use Lemma 76_{182} to find that $P_K(t) \subseteq P(t)$, and use Lemma 77(i)$_{182}$ to infer that $P(t) \subseteq PN(s)$.

(vi). We know from Lemma 76_{182} that $P(t) \setminus P_K(t) \subseteq P(t) \cap P(K)$. Lemma 77(i)$_{182}$ implies that $P(t) \subseteq PN(s)$ and therefore also that $P(t) \cap P(K) \subseteq PN(s) \cap P(K)$. Hence, $P(t) \setminus P_K(t) \subseteq PN(s) \cap P(K)$. □

Lemma 80. *Consider any closed $K \subseteq G$ and $s \in PN(K)$ and let $P_1(K)$ and $P_2(K)$ be an arbitrary partition of $P(K)$.*[26] *Let $K_2 := K \cap D(P_2(K))$ and $K_1 := K \setminus K_2 = K \setminus D(P_2(K))$ and choose any $k_1 \in K_1$ and $k_2 \in K_2$. The following statements hold:*

(i) *K_2 is a closed subset of G;*

(ii) *$P(K_1) \subseteq P_1(K)$;*

(iii) *$P_2(K) \subseteq P(K_2)$;*

(iv) *$K_1 \subseteq PN(k_2)$;*

(v) *$P(k_1) \cap K = P(k_1) \cap K_1$;*

(vi) *$P(k_1) \cap P(K) = P(k_1) \cap P(K_1)$;*

(vii) *$PN(s) \cap K_1 = K \setminus D((P(K) \setminus PN(s)) \cup (PN(s) \cap P_2(K)))$.*

Proof of Lemma 80. (i). Consider $q_1, q_2 \in K \cap D(P_2(K))$ and $r \in G$ such that $q_1 \sqsubseteq r \sqsubseteq q_2$. Since K is closed, we have that $r \in K$, and we are left to show that $r \in D(P_2(K))$. That $q_1 \in D(P_2(K))$ implies the existence of some $u \in P_2(K)$ such that $u \sqsubseteq q_1$ and therefore $u \sqsubseteq r$, implying that $r \in P_2(K) \cup D(P_2(K))$. Since $r \in K$, we know that $r \notin P(K)$ and therefore $r \notin P_2(K)$. We infer that indeed $r \in D(P_2(K))$.

(ii). Consider any $q \in P(K_1)$, implying the existence of some $r \in K_1$ such that $q \in P(r)$ and $q \notin K_1$. We are first going to show that $q \notin P_2(K) \cup K_2$. Assume *ex absurdo* that $q \in P_2(K) \cup K_2$, implying that $q \in P_2(K) \cup D(P_2(K))$, which means that there is some $u \in P_2(K)$ for which $u \sqsubseteq q$ and, since $q \in P(r)$, that $u \sqsubseteq r$. From this we infer that $r \in P_2(K) \cup D(P_2(K))$ and therefore that $r \in D(P_2(K))$, since $r \in K_1 \subseteq K$ implies that $r \notin P(K)$, which in turn implies that $r \notin P_2(K)$. We have thus found that $r \in K \cap D(P_2(K)) = K_2$, contradicting $r \in K_1$. Hence indeed $q \notin P_2(K) \cup K_2$, implying $q \notin P_2(K)$ and $q \notin K_2$. Since also $q \notin K_1$, we find that $q \notin K$, which implies that $q \in P(K)$, since $q \in P(r)$ with $r \in K_1 \subseteq K$. Since $P_1(K)$ and $P_2(K)$ form a partition of $P(K)$ and $q \notin P_2(K)$, we conclude that indeed $q \in P_1(K)$.

(iii). Consider any $q \in P_2(K) \subseteq P(K)$, implying the existence of some $r \in K$ such that $q \in P(r)$. From this we infer that $q \sqsubseteq r$ and therefore $r \in P_2(K) \cup D(P_2(K))$. Since $r \in K$, we see that $r \notin P(K)$ and therefore $r \notin P_2(K)$, whence $r \in D(P_2(K))$. Together with $r \in K$, this implies that $r \in K_2$. Since $q \in P(K)$ implies $q \notin K$ and therefore $q \notin K_2$, we can infer from $q \in P(r)$ that $q \in P(K_2)$.

[26] Here too, we allow that one of the sets $P_1(K)$ or $P_2(K)$ may be empty; see Footnote 25_{183}.

(iv). Consider any $q \in K_1$. Assume *ex absurdo* that $q \notin PN(k_2)$, implying that $q \in \{k_2\} \cup D(k_2)$ and therefore that $k_2 \sqsubseteq q$. Since $k_2 \in K_2$, we infer that $k_2 \in D(P_2(K))$, implying the existence of some $r \in P_2(K)$ such that $r \sqsubseteq k_2$ and therefore $r \sqsubseteq q$, which in turn implies that $q \in P_2(K) \cup D(P_2(K))$. Since $q \in K_1 \subseteq K$, we have that $q \notin P(K)$ and therefore that $q \notin P_2(K)$. Hence $q \in D(P_2(K))$ and therefore also $q \in K_2$, since $q \in K$. This contradicts $q \in K_1$, since K_1 and K_2 form a partition of K.

(v)$_{\cap}$. Since it trivially holds that $P(k_1) \cap K \supseteq P(k_1) \cap K_1$, we only need to prove that $P(k_1) \cap K \subseteq P(k_1) \cap K_1$. So consider any $q \in P(k_1) \cap K$. By definition of $P(K_1)$, we derive from $q \in P(k_1)$ that either $q \in P(K_1)$ or $q \in K_1$. Assume *ex absurdo* that $q \in P(K_1)$, then due to (ii)$_{\cap}$, $q \in P_1(K)$. Since $q \in K$ implies $q \notin P(K)$ and therefore $q \notin P_1(K)$, we have a contradiction. We have thus found that $q \in K_1$ and, since $q \in P(k_1)$, that $q \in P(k_1) \cap K_1$.

(vi)$_{\cap}$. Since we know from (ii)$_{\cap}$ that $P(K_1) \subseteq P_1(K)$, we find that $P(K_1) \subseteq P(K)$ and therefore $P(k_1) \cap P(K) \supseteq P(k_1) \cap P(K_1)$. To prove that $P(k_1) \cap P(K) \subseteq P(k_1) \cap P(K_1)$, consider any $q \in P(k_1) \cap P(K)$. By definition of $P(K_1)$, we derive from $q \in P(k_1)$ that either $q \in P(K_1)$ or $q \in K_1$. Since $q \in P(K)$, we have that $q \notin K$ and therefore also that $q \notin K_1$. Hence $q \in P(K_1)$ and therefore, since $q \in P(k_1)$, also $q \in P(k_1) \cap P(K_1)$.

(vii)$_{\cap}$. First, notice that by subtracting both sides of the expression from K, we obtain the equivalent statement

$$K_2 \cup (K_1 \setminus PN(s)) = K \cap D((P(K) \setminus PN(s)) \cup (PN(s) \cap P_2(K))).$$

Since $(P(K) \setminus PN(s)) \cup (PN(s) \cap P_2(K)) = P_2(K) \cup (P(K) \setminus PN(s))$ and also $K_2 \cup (K_1 \setminus PN(s)) = K_2 \cup (K \setminus PN(s))$, this is in turn equivalent to:

$$K_2 \cup (K \setminus PN(s)) = K \cap D(P_2(K) \cup (P(K) \setminus PN(s))).$$

We will prove this statement instead of the original one.

We first prove that $K_2 \cup (K \setminus PN(s)) \subseteq K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. Consider therefore any $q \in K_2 \cup (K \setminus PN(s))$. On the one hand, if $q \in K_2$, then $q \in D(P_2(K))$. Since $q \notin P_2(K) \cup (P(K) \setminus PN(s))$ because K and $P_2(K) \cup (P(K) \setminus PN(s)) \subseteq P(K)$ are disjoint, this allows us to infer that indeed $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. On the other hand, if $q \in K \setminus PN(s)$, then $q \notin PN(s)$ and therefore $q \in \{s\} \cup D(s)$. Since $s \in PN(K)$ and therefore due to Lemma 78(ii)$_{183}$, $s \notin K$, we know from $q \in K$ that $q \neq s$. We can therefore infer that $q \in D(s)$, implying the existence of a directed path $s = r_1, \ldots, r_n = q$, $n > 1$. Now let j be the first index in $\{1, \ldots, n\}$ for which $r_j \in K$. Since $q \in K$, such an index exists, and since $s \notin K$, $j > 1$, and therefore we can consider the node r_{j-1}. Since $r_{j-1} \in P(r_j)$, $r_j \in K$ and $r_{j-1} \notin K$, we infer that $r_{j-1} \in P(K)$. Since $s \sqsubseteq r_{j-1}$, we have $r_{j-1} \in \{s\} \cup D(s)$ and therefore $r_{j-1} \notin PN(s)$, whence $r_{j-1} \in P(K) \setminus PN(s) \subseteq P_2(K) \cup (P(K) \setminus PN(s))$. Since $q \notin P_2(K) \cup (P(K) \setminus PN(s))$ because $q \in K$, and since $r_{j-1} \sqsubseteq q$, we obtain that $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$.

We now prove that $K_2 \cup (K \setminus PN(s)) \supseteq K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. Consider any $q \in K \cap D(P_2(K) \cup (P(K) \setminus PN(s)))$. This implies that there is some $r \in P_2(K) \cup (P(K) \setminus PN(s))$ such that $r \sqsubseteq q$. On the one hand, if $r \in P_2(K)$, then $q \in D(P_2(K))$ because $q \in K$ and therefore $q \notin P(K)$, whence indeed $q \in K \cap D(P_2(K)) = K_2 \subseteq K_2 \cup (K \setminus PN(s))$. On the other hand, if $r \in P(K) \setminus PN(s)$, then since $r \notin PN(s)$, $r \in \{s\} \cup D(s)$ and therefore $s \sqsubseteq r$, implying that $s \sqsubseteq q$. We therefore find that $q \in \{s\} \cup D(s)$, or equivalently, that $q \notin PN(s)$. By combining this with the fact that $q \in K$, we obtain that $q \in K \setminus PN(s) \subseteq K_2 \cup (K \setminus PN(s))$. □

6.B PROOF OF THEOREM 53

Proof of Theorem 53$_{156}$. Fix a closed set $K \subseteq G$, consider any nonzero $f \in \mathscr{G}(\mathscr{X}_K)$, nonzero $h \in \mathscr{G}(\mathscr{X}_{PN(K)})$ and $x_{P(K)} \in \mathscr{X}_{P(K)}$ such that $f \notin \mathscr{D}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}$ and $h \notin \mathscr{D}^{\mathrm{irr}}_{PN(K)}$, and choose any $f_* \in \mathscr{D}^{\mathrm{irr}}_G$ and $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$. We construct a probability mass function p_G on $\mathscr{X}_G$ such that for the corresponding linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$: $P_G(f_*) > 0$, $P_G(h) < 0$ and $P_G(g\mathbb{I}_{\{x_{P(K)}\}}f) < 0$.

First of all, since $f_* \in \mathscr{D}^{\mathrm{irr}}_G$, Proposition 36$_{136}$ implies that

$$f_* = \sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}, \tag{6.16}$$

where every $f_{s,z_{PN(s)}}$ is an element of $\mathscr{D}_{s\rfloor z_{P(s)}} \cup \{0\}$ and at least one of them is non-zero.

As shown in the proof of Proposition 38$_{137}$, it is possible to find, by repeated application of Lemma 52$_{149}$, for all $s \in G$ and all $z_{P(s)} \in \mathscr{X}_{P(s)}$, a local probability mass function $p_s(\cdot\rfloor z_{P(s)})$ on $\mathscr{X}_s$ with corresponding linear prevision $P_s(\cdot|z_{P(s)})$ on $\mathscr{G}(\mathscr{X}_s)$, such that $p_s(z_s\rfloor z_{P(s)}) > 0$ for all $z_s \in \mathscr{X}_s$, and $P_s(f_{s,z_{PN(s)}}\rfloor z_{P(s)}) > 0$ for all $z_{N(s)} \in \mathscr{X}_{N(s)}$ for which $f_{s,z_{PN(s)}} \neq 0$. We will now use these local probability mass functions to create, for specific closed subsets S of G, Bayesian networks that have a graphical structure corresponding to this closed subset S. By an argument similar to the one for local sets of desirable gambles in Section 6.2$_{154}$, we see that in order to do so, all that is needed is for us to instantiate the value of $X_{P(S)}$. Every choice of $y_{P(S)} \in \mathscr{X}_{P(S)}$ then yields, for all $s \in S$ and $z_{P_S(s)} \in \mathscr{X}_{P_S(s)}$, a conditional local probability mass function $p_s(\cdot\rfloor z_{P_S(s)})$ and corresponding linear prevision $P_s(\cdot\rfloor z_{P_S(s)})$, obtained by identifying them with $p_s(\cdot\rfloor z_{P(s)})$ and $P_s(\cdot\rfloor z_{P(s)})$, where we let $z_{P(s)\setminus P_S(s)} = y_{P(s)\setminus P_S(s)}$. We denote the probability mass function of the resulting Bayesian network by $p_S(\cdot\rfloor y_{P(S)})$ and its corresponding linear prevision by $P_S(\cdot\rfloor y_{P(S)})$. In order to explicitly recall the specific choice of $y_{P(S)} \in \mathscr{X}_{P(S)}$ in the notation used for the local models, we will also write $p_s(\cdot\rfloor z_{P_S(s)}, y_{P(S)}) := p_s(\cdot\rfloor z_{P_S(s)})$ and $P_s(\cdot\rfloor z_{P_S(s)}, y_{P(S)}) := P_s(\cdot\rfloor z_{P_S(s)})$. For every fixed $y_{P(S)} \in \mathscr{X}_{P(S)}$, the linear prevision $P_S(\cdot\rfloor y_{P(S)})$ has a number of useful properties.

A first and trivial property is that $P_S(1|y_{P(S)}) = 1$.

Secondly, consider any $s \in S$. S is a closed subset of G and therefore, due to Lemma 78(vi)$_{183}$, $P(S) \subseteq PN(s)$. It then holds for all $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ and $z_{P(S)} = y_{P(S)}$, that $P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} f_{s,z_{PN(s)}} \rfloor y_{P(S)}) > 0$. To see why, first notice that because S is closed, $PN_S(s) = PN(s) \cap S$ due to Lemma 79(iii)$_{185}$. It then follows from the conditional independence properties of Bayesian networks that indeed

$$\begin{aligned}
P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} f_{s,z_{PN(s)}} \rfloor y_{P(S)}) &= P_S(\mathbb{I}_{\{z_{PN_S(s)}\}} f_{s,z_{PN(s)}} \rfloor y_{P(S)}) \\
&= P_S(z_{PN_S(s)} \rfloor y_{P(S)}) P_s(f_{s,z_{PN(s)}} \rfloor z_{PN_S(s)}, y_{P(S)}) \\
&= P_S(z_{PN_S(s)} \rfloor y_{P(S)}) P_s(f_{s,z_{PN(s)}} \rfloor z_{P_S(s)}, y_{P(S)}) \\
&= P_S(z_{PN_S(s)} \rfloor y_{P(S)}) P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)}) > 0,
\end{aligned}$$

where the inequality holds because $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)})$ and $P_S(z_{PN_S(s)} \rfloor y_{P(S)})$ are strictly positive. For $P_s(f_{s,z_{PN(s)}} \rfloor z_{P(s)})$, this is true by construction, and for $P_S(z_{PN_S(s)} \rfloor y_{P(S)})$, this holds because all local probabilities are by construction strictly positive and therefore the global ones are too.

Thirdly, fix $s \in G \setminus S$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $z_{P(S)\cap PN(s)} = y_{P(S)\cap PN(s)}$. By applying the factorisation and conditional independence properties of the resulting Bayesian network, we find that

$$\begin{aligned}
&P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} \rfloor y_{P(S)}) \\
&= \sum_{\substack{w_S \in \mathscr{X}_S \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} p_S(w_S \rfloor y_{P(S)}) \\
&= \sum_{\substack{w_S \in \mathscr{X}_S \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in S} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) \\
&= \prod_{k \in PN(s)\cap S} p_k(z_k \rfloor z_{P_S(k)}, y_{P(S)}) \sum_{\substack{w_S \in \mathscr{X}_S \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in S \setminus PN(s)} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) \\
&= \prod_{k \in PN(s)\cap S} p_k(z_k \rfloor z_{P_S(k)}, y_{P(S)}) \\
&= \prod_{k \in PN(s)\cap S} p_k(z_k \rfloor z_{P(k)}).
\end{aligned}$$

To understand the third equality, notice that since S is closed, Lemma 79(v)$_{185}$ implies, for all $k \in PN(s) \cap S$, that $P_S(k) \subseteq PN(s) \cap S$. For the fifth equality, it suffices to apply Lemma 79(vi)$_{185}$ to find that for all $k \in PN(s) \cap S$: $P(k) \setminus P_S(k) \subseteq PN(s) \cap P(S)$. The fourth equality is a bit more complicated. It is trivial if $S \setminus PN(s) = \emptyset$, so suppose that $S \setminus PN(s) \neq \emptyset$. Pick any leaf ℓ from the sub-DAG that corresponds to the nodes in $S \setminus PN(s)$; this is possible because $S \setminus PN(s) \neq \emptyset$ and a DAG always has at least one leaf. We then find

that

$$
\begin{aligned}
&\sum_{\substack{w_S \in \mathscr{X}_S \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in S \setminus PN(s)} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) \\
&= \sum_{\substack{w_{S\setminus\{\ell\}} \in \mathscr{X}_{S\setminus\{\ell\}} \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \sum_{w_\ell \in \mathscr{X}_\ell} \prod_{k \in S \setminus PN(s)} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) \\
&= \sum_{\substack{w_{S\setminus\{\ell\}} \in \mathscr{X}_{S\setminus\{\ell\}} \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in (S \setminus PN(s)) \setminus \{\ell\}} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) \sum_{w_\ell \in \mathscr{X}_\ell} p_\ell(w_\ell \rfloor w_{P_S(\ell)}, y_{P(S)}) \\
&= \sum_{\substack{w_{S\setminus\{\ell\}} \in \mathscr{X}_{S\setminus\{\ell\}} \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in (S \setminus PN(s)) \setminus \{\ell\}} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}).
\end{aligned}
$$

The first equality holds because $\ell \notin PN(s) \cap S$, and the second one because $\ell \notin P_S(k)$ for all $k \in (S \setminus PN(s)) \setminus \{\ell\}$, since ℓ was assumed to be a leaf of the sub-DAG that corresponds to $S \setminus PN(s)$. By repeating this argument for the sub-DAG that corresponds to the nodes in $(S \setminus PN(s)) \setminus \{\ell\}$, we can remove yet another node, and if we go on in this way until no node remains, we eventually obtain that indeed

$$
\sum_{\substack{w_S \in \mathscr{X}_S \\ w_{PN(s)\cap S} = z_{PN(s)\cap S}}} \prod_{k \in S \setminus PN(s)} p_k(w_k \rfloor w_{P_S(k)}, y_{P(S)}) = 1.
$$

Hence, for any $s \in G \setminus S$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $z_{P(S)\cap PN(s)} = y_{P(S)\cap PN(s)}$:

$$
P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} \rfloor y_{P(S)}) = \prod_{k \in PN(s)\cap S} p_k(z_k \rfloor z_{P(k)}).
$$

We can derive two additional things from this result. First of all, $P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} \rfloor y_{P(S)})$ is strictly positive because all local probabilities are strictly positive by construction. And secondly, $P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} \rfloor y_{P(S)})$ does not depend on the particular value of $y_{P(S)\setminus PN(s)}$ because for all $k \in PN(s) \cap S$, $P(k) \subseteq PN(s)$ due to Lemma 77(i)$_{182}$.

If we now no longer consider a fixed value of $y_{P(S)} \in \mathscr{X}_{P(S)}$, then the results mentioned above have a number of immediate consequences. First of all, the gamble $P_S(1 \rfloor X_{P(S)})$ is constant and equal to 1. Secondly, for all $s \in S$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, $P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} f_{s,z_{PN(s)}} \rfloor z_{P(S)}) > 0$. And thirdly, for all $s \in G \setminus S$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ the gamble

$$
P_S(\mathbb{I}_{\{z_{PN(s)\cap S}\}} \rfloor z_{P(S)\cap PN(s)}, X_{P(S)\setminus PN(s)})
$$

is constant, strictly positive, and equal to $\prod_{k \in PN(s)\cap S} p_k(z_k \rfloor z_{P(k)})$.

We are now ready for the main proof, which consists of four parts. The first three parts are separate. In each of these three parts, we will construct

specific probability mass functions and linear previsions and show that they satisfy a number of properties. In the fourth part of the proof, we will combine these models and their properties with each other to construct p_G and P_G, and to show that $P_G(f_*) > 0$, $P_G(h) < 0$ and $P_G(g\mathbb{I}_{\{x_{P(K)}\}}f) < 0$.

For the first part of the proof, we consider the Bayesian networks that correspond to the subset $S := D(K)$ of G, which is closed because of Lemma 78(iii)$_{183}$. Every $y_{P(D(K))} \in \mathscr{X}_{P(D(K))}$ yields a Bayesian network, and therefore a probability mass function $p_{D(K)}(\cdot\rfloor y_{P(D(K))})$ on $\mathscr{X}_{D(K)}$ and an associated linear prevision $P_{D(K)}(\cdot\rfloor y_{P(D(K))})$ on $\mathscr{G}(\mathscr{X}_{D(K)})$. If we do not fix the value of $y_{P(D(K))} \in \mathscr{X}_{P(D(K))}$, then $P_{D(K)}(\cdot\rfloor X_{P(D(K))})$ satisfies a number of properties, which have already been proved above for general closed subsets S of G. First of all, the gamble $P_{D(K)}(1\rfloor X_{P(D(K))})$ is constant and equal to 1. Secondly, for all $s \in D(K)$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, $P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor z_{P(D(K))}) > 0$. And thirdly, for all $s \in G \setminus D(K)$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ the gamble

$$P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}\rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$$

is constant, strictly positive, and equal to $\prod_{k\in PN(s)\cap D(K)} p_k(z_k\rfloor z_{P(k)})$.

For the second part of the proof, we start by considering the following collection of gambles on $\mathscr{X}_K$:

$$\begin{aligned}\mathscr{A}^*_{K\rfloor x_{P(K)}} := \Big\{\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}f_{s,z_{PN(s)}} : s \in K,\ z_{PN(s)} \in \mathscr{X}_{PN(s)},\\ z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)},\ P(s)\cap K \subseteq K_1 \subseteq K,\\ f_{s,z_{PN(s)}} \neq 0\Big\},\end{aligned}$$

which is a finite subset of $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}} := \text{posi}(\mathscr{A}^{\text{irr}}_{K\rfloor x_{P(K)}})$. To see why, first notice that because $PN_K(s) = PN(s)\cap K$ due to Lemma 79(iii)$_{185}$, $\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}$ is clearly the (finite) sum of all indicators $\mathbb{I}_{\{y_{PN_K(s)}\}}$ such that $y_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}$ and $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$. By definition of the posi operator, we are now left to show that for any $y_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}$ such that $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$, we have $\mathbb{I}_{\{y_{PN_K(s)}\}}f_{s,z_{PN(s)}} \in \mathscr{A}^{\text{irr}}_{K\rfloor x_{P(K)}}$. By construction of $\mathscr{A}^*_{K\rfloor x_{P(K)}}$, we know that $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$, and it therefore suffices to show that $y_{P_K(s)} = z_{P_K(s)}$. To see why this last equality holds, first notice that $P_K(s) = P(s)\cap K$ due to Lemma 76$_{182}$. Also, $P(s)\cap K \subseteq PN(s)\cap K_1$ because $P(s)\cap K \subseteq K_1$ by construction of $\mathscr{A}^*_{K\rfloor x_{P(K)}}$ and $P(s)\cap K \subseteq PN(s)$ by definition of $PN(s)$. Therefore, we find that $P_K(s) \subseteq PN(s)\cap K_1$, implying that $y_{P_K(s)} = z_{P_K(s)}$ is a direct consequence of $y_{PN(s)\cap K_1} = z_{PN(s)\cap K_1}$.

Due to the coherence of $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ and the fact that $f \neq 0$ and $f \notin \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$, we have that $0 \notin \text{posi}(\{-f\} \cup \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}})$. To see why this holds, assume *ex*

absurdo that $0 \in \operatorname{posi}(\{-f\} \cup \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}})$. Hence, since $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ is coherent, we can find $\lambda_1, \lambda_2 > 0$ and $f' \in \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ such that $\lambda_1(-f) + \lambda_2 f' = 0$ and therefore $f = (\lambda_2/\lambda_1)f' \in \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$, contradicting $f \notin \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$.

Since $\mathscr{A}^*_{K\rfloor x_{P(K)}}$ and $\mathscr{G}(\mathscr{X}_K)_{>0}$ are both subsets of $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$, we can infer from $0 \notin \operatorname{posi}(\{-f\} \cup \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}})$ that

$$0 \notin \operatorname{posi}(\{-f\} \cup \mathscr{A}^*_{K\rfloor x_{P(K)}} \cup \mathscr{G}(\mathscr{X}_K)_{>0}) =: \mathscr{E}(\{-f\} \cup \mathscr{A}^*_{K\rfloor x_{P(K)}}).$$

We also know that $\{-f\} \cup \mathscr{A}^*_{K\rfloor x_{P(K)}}$ is a finite subset of $\mathscr{G}(\mathscr{X}_K)$ and therefore, we can apply Lemma 52_{149}. This provides us with a probability mass function $p_{K\rfloor x_{P(K)}}$ on $\mathscr{X}_K$ with an associated linear prevision $P_{K\rfloor x_{P(K)}}$ on $\mathscr{G}(\mathscr{X}_K)$ for which $p_{K\rfloor x_{P(K)}}(z_K) > 0$ for all $z_K \in \mathscr{X}_K$, $P_{K\rfloor x_{P(K)}}(f') > 0$ for all $f' \in \mathscr{A}^*_{K\rfloor x_{P(K)}}$ and $P_{K\rfloor x_{P(K)}}(f) < 0$.

Using this probability mass function $p_{K\rfloor x_{P(K)}}$ on $\mathscr{X}_K$ and the local probability mass functions that were introduced avove, we will now construct, for every instantiation $y_{P(K)} \in \mathscr{X}_{P(K)}$, a conditional probability mass function $p_K(\cdot\rfloor y_{P(K)})$ on $\mathscr{X}_K$. So consider any $y_{P(K)} \in \mathscr{X}_{P(K)}$. If $y_{P(K)} = x_{P(K)}$, we define $p_K(\cdot\rfloor x_{P(K)}) := p_{K\rfloor x_{P(K)}}$. If $y_{P(K)}$ is such that $y_k \neq x_k$ for all $k \in P(K)$, then $p_K(\cdot\rfloor y_{P(K)})$ is constructed in the same way as already discussed above for general closed sets S: we use the local probability mass functions and the instantiation $y_{P(K)}$ of the parent variables $X_{P(K)}$ to construct a Bayesian network that has a graphical structure corresponding to the subset K of G. Unlike the one in the preceding paragraphs, this construction does not take into account the gamble f. In all other cases, we need a more complex construction that includes the previous two as a special case.

Let us denote by $P_1(K)$ the largest subset of $P(K)$ such that $y_k = x_k$ for all $k \in P_1(K)$ and let $P_2(K) := P(K) \setminus P_1(K)$. We also let $K_2 := K \cap D(P_2(K))$ and $K_1 := K \setminus D(P_2(K))$. These sets depend on $y_{P(K)}$, but we have not reflected this in the notation to avoid cluttering up the formulas and because $y_{P(K)}$ is fixed in this part of the proof.

Due to Lemma $80(\text{i})_{186}$, K_2 is a closed subset of G. Therefore, as we already explained for general closed sets S, for any $z_{P(K_2)} \in \mathscr{X}_{P(K_2)}$, we can construct a strictly positive probability mass function $p_{K_2}(\cdot\rfloor z_{P(K_2)})$ on $\mathscr{X}_{K_2}$ and an associated linear prevision $P_{K_2}(\cdot\rfloor z_{P(K_2)})$ on $\mathscr{G}(\mathscr{X}_{K_2})$, by using the local probability mass functions to construct a Bayesian network that has a graphical structure corresponding to the subset K_2 of G. By our results for general closed sets S above, with $S = K_2$, we know for all $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ that $P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}\rfloor z_{P(K_2)}) > 0$ if $s \in K_2$ and that the gamble $P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}\rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ is constant, strictly positive and equal to $\prod_{k \in PN(s)\cap K_2} p_k(z_k\rfloor z_{P(k)})$ if $s \notin K_2$.

Next, we define the mass function $p_{K_1}(\cdot\rfloor x_{P_1(K)})$ on $\mathscr{X}_{K_1}$ as the marginalisation of $p_{K\rfloor x_{P(K)}}$ to $\mathscr{X}_{K_1}$: for all $z_{K_1} \in \mathscr{X}_{K_1}$, we let

$$p_{K_1}(z_{K_1}\rfloor x_{P_1(K)}) := \sum_{\substack{w_K \in \mathscr{X}_K \\ w_{K_1} = z_{K_1}}} p_{K\rfloor x_{P(K)}}(w_K).$$

Since all the terms in this sum are strictly positive by construction, we have that $p_{K_1}(z_{K_1}\rfloor x_{P_1(K)}) > 0$ for all $z_{K_1} \in \mathscr{X}_{K_1}$.

For the corresponding linear prevision $P_{K_1}(\cdot\rfloor x_{P_1(K)})$ on $\mathscr{G}(\mathscr{X}_{K_1})$, we get that $P_{K_1}(f'\rfloor x_{P_1(K)}) = P_{K\rfloor x_{P(K)}}(f')$ for all $f' \in \mathscr{G}(\mathscr{X}_{K_1})$.

We can now construct the probability mass function $p_K(\cdot\rfloor y_{P(K)})$ by defining, for all $z_K \in \mathscr{X}_K$:

$$\begin{aligned} p_K(z_K\rfloor y_{P(K)}) :=& p_{K_1}(z_{K_1}\rfloor y_{P_1(K)}) p_{K_2}(z_{K_2}\rfloor y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}) \\ =& p_{K_1}(z_{K_1}\rfloor x_{P_1(K)}) p_{K_2}(z_{K_2}\rfloor y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}), \end{aligned}$$

which makes sense because $P(K_2)\setminus P(K) \subseteq K_1$. It should be clear that for all $z_K \in \mathscr{X}_K$, we have that $p_K(z_K|y_{P(K)}) > 0$.

For the corresponding linear prevision $P_K(\cdot|y_{P(K)})$, the law of iterated prevision yields for all $f' \in \mathscr{G}(\mathscr{X}_K)$ that

$$P_K(f'\rfloor y_{P(K)}) = P_{K_1}(P_{K_2}(f'\rfloor y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})\rfloor x_{P_1(K)}). \tag{6.17}$$

This linear prevision has two useful properties that we will need further on in this proof.

For the first property of $P_K(\cdot|y_{P(K)})$, consider any $s \in K$, implying that $P(K) \subseteq PN(s)$ due to Lemma 78(vi)$_{183}$. It then holds for all $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$ and $z_{P(K)} = y_{P(K)}$, that $P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}\rfloor z_{P(K)}) > 0$. To see why, consider two distinct cases: $s \in K_2$ and $s \in K_1$.

If $s \in K_2$, then because $K_1 \subseteq PN(s)$ due to Lemma 80(iv)$_{186}$, we infer from Equation (6.17) that

$$\begin{aligned} &P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}\rfloor y_{P(K)}) \\ &\quad = P_{K_1}(P_{K_2}(\mathbb{I}_{\{z_{K_1}\}}\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}\rfloor y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})\rfloor x_{P_1(K)}) \\ &\quad = P_{K_1}(\mathbb{I}_{\{z_{K_1}\}} P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}\rfloor y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)})\rfloor x_{P_1(K)}) \\ &\quad = P_{K_1}(\mathbb{I}_{\{z_{K_1}\}}\rfloor x_{P_1(K)}) P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}\rfloor y_{P(K_2)\cap P(K)}, z_{P(K_2)\setminus P(K)}) \\ &\quad = p_{K_1}(z_{K_1}\rfloor x_{P_1(K)}) P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} f_{s,z_{PN(s)}}\rfloor z_{P(K_2)}) > 0, \end{aligned}$$

where the final expression is strictly positive because both factors have been proved above to be strictly positive.

If $s \in K_1$, then because $P(K) \subseteq PN(s)$, we infer from Equation (6.17)$_{\curvearrowleft}$ that

$$\begin{aligned}
&P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} \rfloor y_{P(K)}) \\
&= P_{K_1}(P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} \mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)}) \rfloor x_{P_1(K)}) \\
&= P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)}) \rfloor x_{P_1(K)}) \\
&= P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} \rfloor x_{P_1(K)}) P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)}) \\
&> 0.
\end{aligned}$$

The third equality holds because $P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ has been shown to be a constant gamble earlier on and the final expression is strictly positive because the two constituting factors are strictly positive. For $P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$, this has already been proved. For $P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} \rfloor x_{P_1(K)})$, this follows from

$$P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} \rfloor x_{P_1(K)}) = P_{K \rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}}) > 0,$$

where the final inequality holds because $\mathbb{I}_{\{z_{PN(s)\cap K_1}\}} f_{s,z_{PN(s)}} \in \mathscr{A}^*_{K \rfloor x_{P(K)}}$. This last inclusion in turn holds because $P(s) \cap K = P(s) \cap K_1 \subseteq K_1 \subseteq K$ due to Lemma 80(v)$_{186}$, and because $P(s) \cap P(K) = P(s) \cap P(K_1) \subseteq P(K_1) \subseteq P_1(K)$ due to Lemma 80(vi)$_{186}$&(ii)$_{186}$ and therefore $z_{P(s)\cap P(K)} = x_{P(s)\cap P(K)}$, implying that $z_{P(s)\setminus P_K(s)} = x_{P(s)\setminus P_K(s)}$ due to Lemma 76$_{182}$.

The second property of $P_K(\cdot \rfloor y_{P(K)})$ is that for all $s \in PN(K)$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $z_{P(K)\cap PN(s)} = y_{P(K)\cap PN(s)}$, $P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} \rfloor y_{P(K)})$ is strictly positive and does not depend on the particular value of $y_{P(K)\setminus PN(s)}$ that was used to construct $P_K(\cdot \rfloor y_{P(K)})$. To prove this, we start by recalling from the discussion above that, because $s \notin K_2$, $P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, X_{P(K_2)\setminus PN(s)})$ is a constant, strictly positive gamble that is furthermore equal to $\prod_{k \in PN(s)\cap K_2} p_k(z_k \rfloor z_{P(k)})$, which implies that this is also the case for

$$P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor z_{P(K_2)\cap PN(s)}, y_{(P(K_2)\cap P(K))\setminus PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)}).$$

Since the fact that $z_{P(K)\cap PN(s)} = y_{P(K)\cap PN(s)}$ allows us to infer that $z_{P(K_2)\cap P(K)\cap PN(s)} = y_{P(K_2)\cap P(K)\cap PN(s)}$, this implies that

$$P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}} \rfloor y_{P(K_2)\cap P(K)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})$$

is a constant, strictly positive gamble that is equal to $\prod_{k \in PN(s)\cap K_2} p_k(z_k \rfloor z_{P(k)})$.

Hence, using Equation (6.17)$_{193}$, we find that

$$\begin{aligned}
&P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}\rfloor y_{P(K)})\\
&= P_{K_1}(P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}\rfloor y_{P(K_2)\cap P(K)}, X_{P(K_2)\setminus P(K)})\rfloor x_{P_1(K)})\\
&= P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\\
&\qquad P_{K_2}(\mathbb{I}_{\{z_{PN(s)\cap K_2}\}}\rfloor y_{P(K_2)\cap P(K)}, z_{(P(K_2)\setminus P(K))\cap PN(s)}, X_{(P(K_2)\setminus P(K))\setminus PN(s)})\\
&\qquad\qquad \rfloor x_{P_1(K)})\\
&= P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\rfloor x_{P_1(K)}) \prod_{k\in PN(s)\cap K_2} p_k(z_k\rfloor z_{P(k)}).
\end{aligned}$$

The property that we are trying to prove will therefore follow if we can show that both $P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\rfloor x_{P_1(K)})$ and $\prod_{k\in PN(s)\cap K_2} p_k(z_k\rfloor z_{P(k)})$ are strictly positive and do not depend on the particular value of $y_{P(K)\setminus PN(s)} \in \mathscr{X}_{P(K)\setminus PN(s)}$.

We start with $P_{K_1}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}}\rfloor x_{P_1(K)})$. It is by definition equal to $P_{K\rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}})$ and therefore strictly positive because $p_{K\rfloor x_{P(K)}}$ is a strictly positive probability mass function. Since $s \in PN(K)$, we can use Lemma 80(vii)$_{186}$ to infer that

$$PN(s)\cap K_1 = K\setminus D((P(K)\setminus PN(s))\cup(PN(s)\cap P_2(K))).$$

We therefore find that $PN(s)\cap K_1$ does not depend on the particular value of $y_{P(K)\setminus PN(s)}$ in $\mathscr{X}_{P(K)\setminus PN(s)}$ because $PN(s)\cap P_2(K)$ is fully determined by $y_{P(K)\cap PN(s)}$. Hence, $P_{K\rfloor x_{P(K)}}(\mathbb{I}_{\{z_{PN(s)\cap K_1}\}})$ does not depend on $y_{P(K)\setminus PN(s)}$ either.

For $\prod_{k\in PN(s)\cap K_2} p_k(z_k\rfloor z_{P(k)})$, it is clearly true that it is strictly positive. To show that it does not depend on $y_{P(K)\setminus PN(s)}$, we start by noticing that $PN(s)\cap K_2$ does not depend on $y_{P(K)\setminus PN(s)}$ because, as we have shown in the previous paragraph, $PN(s)\cap K_1$ does not depend on it, and because $PN(s)\cap K_2 = (PN(s)\cap K)\setminus(PN(s)\cap K_1)$. Next, for all $k\in PN(s)\cap K_2$, the factor $p_k(z_k\rfloor z_{P(k)})$ will not depend on $y_{P(K)\setminus PN(s)}$ because $P(k)\subseteq PN(s)$ due to Lemma 77(i)$_{182}$. Hence, $\prod_{k\in PN(s)\cap K_2} p_k(z_k\rfloor z_{P(k)})$ does not depend on the particular value of $y_{P(K)\setminus PN(s)}$ in $\mathscr{X}_{P(K)\setminus PN(s)}$.

If we now no longer consider a fixed value of $y_{P(K)} \in \mathscr{X}_{P(K)}$, then the results mentioned above have two immediate consequences. Firstly, for all $s\in K$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, we have that $P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}}\rfloor z_{P(K)}) > 0$. Secondly, for all $s\in PN(K)$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, the gamble $P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}}\rfloor z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})$ is constant and strictly positive.

For the third part of the proof, we start by considering the following collection of gambles on $\mathscr{X}_{PN(K)}$:

$$\mathscr{A}^*_{PN(K)} := \Big\{\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}} : s\in PN(K),\ z_{PN(s)} \in \mathscr{X}_{PN(s)},\ f_{s,z_{PN(s)}} \neq 0\Big\}.$$

which is a finite subset of $\mathscr{A}^{\text{irr}}_{PN(K)}$, for the following three reasons. First of all, because $P(PN(K)) = \emptyset$ [see Lemma 78(iv)$_{183}$]. Secondly, because—since $PN(K)$ is closed because of Lemma 78(iii)$_{183}$—Lemma 79(iii)$_{185}$ implies that $PN(s) \cap PN(K) = PN_{PN(K)}(s)$ for all $s \in PN(K)$. Thirdly, because the first point implies that, for all $s \in PN(K)$, $P_{PN(K)}(s) = P(s)$.

Since $\mathscr{A}^{\text{irr}}_{PN(K)} \subseteq \mathscr{D}^{\text{irr}}_{PN(K)}$, we find that $\mathscr{A}^{*}_{PN(K)}$ is a finite subset of $\mathscr{D}^{\text{irr}}_{PN(K)}$. Since $h \neq 0$ and $h \notin \mathscr{D}^{\text{irr}}_{PN(K)}$, we can now use an argument that is completely analogous to the one that we used for $\mathscr{A}^{*}_{K\rfloor x_{P(K)}}$, f and $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ earlier on in this proof to show that $0 \notin \mathscr{E}(\{-h\} \cup \mathscr{A}^{*}_{PN(K)})$. Since we know that $\{-h\} \cup \mathscr{A}^{*}_{PN(K)}$ is a finite subset of $\mathscr{G}(\mathscr{X}_{PN(K)})$, we can now apply Lemma 52$_{149}$. This provides us with a probability mass function $p_{PN(K)}$ on $\mathscr{X}_{PN(K)}$ with an associated linear prevision $P_{PN(K)}$ on $\mathscr{G}(\mathscr{X}_{PN(K)})$ for which $p_{PN(K)}(z_{PN(K)}) > 0$ for all $z_{PN(K)} \in \mathscr{X}_{PN(K)}$, $P_{PN(K)}(f') > 0$ for all $f' \in \mathscr{A}^{*}_{PN(K)}$ and $P_{PN(K)}(h) < 0$. As a direct consequence, we find that $P_{PN(K)}$ has two additional properties as well.

Firstly, for all $s \in PN(K)$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$, we find that $P_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}}) > 0$ because $\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}} \in \mathscr{A}^{*}_{PN(K)}$.

Secondly: $P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}) > 0$. Indeed, since $g\mathbb{I}_{\{x_{P(K)}\}} \in \mathscr{G}(\mathscr{X}_{PN(K)_{>0}})$, we know that $P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})$ is a positive linear combination of probabilities $p_{PN(K)}(z_{PN(K)})$, with $z_{PN(K)} \in \mathscr{X}_{PN(K)}$, of which we already know that they are strictly positive.

We are now ready for the fourth and final part of the proof. We start by defining the mass function p_G on $\mathscr{X}_G$ that we have been after all along. For all $z_G \in \mathscr{X}_G$, we let

$$p_G(z_G) := p_{PN(K)}(z_{PN(K)}) p_K(z_K | z_{P(K)}) p_{D(K)}(z_{D(K)} | z_{P(D(K))}),$$

where all three factors have been defined in earlier parts of this proof. Since $PN(K)$, K and $D(K)$ constitute a partition of G due to Lemma 78(ii)$_{183}$, and since $P(K) \subseteq PN(K)$ and $P(D(K)) \subseteq PN(K) \cup K$, we see that p_G is indeed a mass function on $\mathscr{X}_G$. For the corresponding linear prevision P_G, the law of iterated prevision implies that for all $f' \in \mathscr{G}(\mathscr{X}_G)$:

$$P_G(f') = P_{PN(K)}(P_K(P_{D(K)}(f' \rfloor X_{P(D(K))}) \rfloor X_{P(K)})). \tag{6.18}$$

Hence, since $h \in \mathscr{G}(\mathscr{X}_{PN(K)})$, we find that

$$\begin{aligned} P_G(h) &= P_{PN(K)}(P_K(P_{D(K)}(h \rfloor X_{P(D(K))}) \rfloor X_{P(K)})) \\ &= P_{PN(K)}(hP_K(1P_{D(K)}(1 \rfloor X_{P(D(K))}) \rfloor x_{P(K)})) \\ &= P_{PN(K)}(hP_K(1 \rfloor x_{P(K)})) = P_{PN(K)}(h) < 0. \end{aligned}$$

Similarly, since $g \in \mathscr{G}(\mathscr{X}_{N(K)})_{>0}$ and $f \in \mathscr{G}(\mathscr{X}_K)$, we find that

$$\begin{aligned}
P_G(g\mathbb{I}_{\{x_{P(K)}\}}f) &= P_{PN(K)}(P_K(P_{D(K)}(g\mathbb{I}_{\{x_{P(K)}\}}f\rfloor X_{P(D(K))})\rfloor X_{P(K)})) \\
&= P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}P_K(fP_{D(K)}(1\rfloor X_{P(D(K))})\rfloor x_{P(K)})) \\
&= P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}P_K(f\rfloor x_{P(K)})) \\
&= P_K(f\rfloor x_{P(K)})P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}}) < 0,
\end{aligned}$$

where the final inequality holds because $P_K(f\rfloor x_{P(K)})$ is strictly negative and $P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})$ is strictly positive. For $P_K(f\rfloor x_{P(K)})$, this is true by construction because $P_K(f\rfloor x_{P(K)}) = P_{K\rfloor x_{P(K)}}(f) < 0$ and for $P_{PN(K)}(g\mathbb{I}_{\{x_{P(K)}\}})$, this has been shown earlier on in this proof.

All that is now left to do is to show that $P_G(f_*) > 0$. Since $f_{s,z_{PN(s)}} \neq 0$ for at least one $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$, Equation (6.16)$_{188}$ tells us that it suffices to show that $P_G(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}}) > 0$ for all $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ such that $f_{s,z_{PN(s)}} \neq 0$. So let us fix any such $s \in G$ and $z_{PN(s)} \in \mathscr{X}_{PN(s)}$ and show that $P_G(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}}) > 0$. We consider three cases: $s \in D(K)$, $s \in K$ and $s \in PN(K)$. They are exhaustive and mutually exclusive because, as we already mentioned before, $D(K)$, K and $PN(K)$ constitute a partition of G [see Lemma 78(ii)$_{183}$].

If $s \in D(K)$, then, because $D(K)$ is closed due to Lemma 78(iii)$_{183}$, and because $PN(D(K)) = PN(K) \cup K$ due to Lemma 78(v)$_{183}$, we can use Lemma 78(vii)$_{183}$ to infer that $PN(K) \cup K \subseteq PN(s)$. By combining this with Equation (6.18)$_{\cap}$, we find that

$$\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}}) \\
&\quad = P_{PN(K)}(P_K(P_{D(K)}(\mathbb{I}_{\{z_{PN(s)}\}}f_{s,z_{PN(s)}}\rfloor X_{P(D(K))})\rfloor X_{P(K)})) \\
&\quad = P_{PN(K)}(P_K(P_{D(K)}(\mathbb{I}_{\{z_{PN(K)}\}}\mathbb{I}_{\{z_K\}}\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor X_{P(D(K))})\rfloor X_{P(K)})) \\
&\quad = P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}}P_K(\mathbb{I}_{\{z_K\}}P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor z_{P(D(K))})\rfloor z_{P(K)})) \\
&\quad = P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}}P_K(\mathbb{I}_{\{z_K\}}\rfloor z_{P(K)}))P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor z_{P(D(K))}) \\
&\quad = P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}})P_K(\mathbb{I}_{\{z_K\}}\rfloor z_{P(K)})P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor z_{P(D(K))}) \\
&\quad = p_{PN(K)}(z_{PN(K)})p_K(z_K\rfloor z_{P(K)})P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}}f_{s,z_{PN(s)}}\rfloor z_{P(D(K))}) > 0,
\end{aligned}$$

where the inequality holds because all three constituting factors have been shown to be strictly positive earlier on.

If $s \in K$, then since $PN(K) \subseteq PN(s)$ due to Lemma 78(vii)$_{183}$, Equa-

tion $(6.18)_{196}$ implies that

$$\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= P_{PN(K)}(P_K(P_{D(K)}(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}} \rfloor X_{P(D(K))}) \rfloor X_{P(K)})) \\
&= P_{PN(K)}(P_K(P_{D(K)}(\mathbb{I}_{\{z_{PN(K)}\}} \mathbb{I}_{\{z_{PN(s)\cap K}\}} \mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}} \rfloor X_{P(D(K))}) \rfloor X_{P(K)})) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}} P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) \\
&\qquad\qquad \rfloor z_{P(K)})),
\end{aligned}$$

which, since $P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$ has been shown to be a constant gamble earlier on in this proof (observe that $s \notin D(K)$ because $s \in K$), in turn implies that

$$\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}} P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} \rfloor z_{P(K)})) \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(K)}\}}) P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} \rfloor z_{P(K)}) \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) \\
&= p_{PN(K)}(z_{PN(K)}) P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} f_{s,z_{PN(s)}} \rfloor z_{P(K)}) \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) > 0,
\end{aligned}$$

where the final expression is strictly positive since all three constituting factors have been shown to be strictly positive earlier on in this proof.

If $s \in PN(K)$, Equation $(6.18)_{196}$ implies that

$$\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= P_{PN(K)}(P_K(P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} \mathbb{I}_{\{z_{PN(s)\cap K}\}} \mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} f_{s,z_{PN(s)}} \\
&\qquad\qquad \rfloor X_{P(D(K))}) \rfloor X_{P(K)})) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}} \\
&\qquad P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) \\
&\qquad\qquad \rfloor z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})),
\end{aligned}$$

which, since $P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})$ has been shown to be a constant gamble earlier on in this proof, in turn implies that

$$
\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}} \\
&\qquad P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} \rfloor z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})) \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)})
\end{aligned}
$$

and therefore, since $P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} \rfloor z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)})$ has also already been shown to be a constant gamble, that

$$
\begin{aligned}
&P_G(\mathbb{I}_{\{z_{PN(s)}\}} f_{s,z_{PN(s)}}) \\
&= P_{PN(K)}(\mathbb{I}_{\{z_{PN(s)\cap PN(K)}\}} f_{s,z_{PN(s)}}) \\
&\qquad P_K(\mathbb{I}_{\{z_{PN(s)\cap K}\}} \rfloor z_{P(K)\cap PN(s)}, X_{P(K)\setminus PN(s)}) \\
&\qquad\qquad P_{D(K)}(\mathbb{I}_{\{z_{PN(s)\cap D(K)}\}} \rfloor z_{P(D(K))\cap PN(s)}, X_{P(D(K))\setminus PN(s)}) > 0,
\end{aligned}
$$

where the inequality is again due to the fact that the three constituting factors are strictly positive, as was shown earlier on in this proof. □

6.C PROOF OF PROPOSITION 56

Proof of Proposition 56$_{158}$. First, define $\mathscr{D}_{s\rfloor z_{P(s)}} := \mathscr{D}_{\underline{P}_{s\rfloor z_{P(s)}}}$ for all $s \in G$ and $z_{P(s)} \in \mathscr{X}_{P(s)}$, where the coherent lower previsions $\underline{P}_{s\rfloor z_{P(s)}}$ are the local models that were used to construct $\underline{P}^{\text{irr}}_G$, $\underline{P}^{\text{irr}}_{PN(K)}$ and $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}$. This allows us to consider the corresponding irrelevant natural extensions $\mathscr{D}^{\text{irr}}_G$, $\mathscr{D}^{\text{irr}}_{PN(K)}$ and $\mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$.

Next, define $f_* := f - \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(f)$ and

$$
h_* := h + g\mathbb{I}_{\{x_{P(K)}\}} \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(f) - \underline{P}^{\text{irr}}_{PN(K)}\big(h + g\mathbb{I}_{\{x_{P(K)}\}} \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(f)\big).
$$

Then $\underline{P}^{\text{irr}}_{PN(K)}(h_*) = 0$ and $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(f_*) = 0$ because $\underline{P}^{\text{irr}}_{PN(K)}$ and $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}$ are coherent [C8$_{49}$].

Now fix any $\varepsilon > 0$. Theorem 42$_{139}$ and Equations (2.5)$_{41}$ and (2.6)$_{41}$ then imply that $f_* + \varepsilon \in \mathscr{D}^{\text{irr}}_{K\rfloor x_{P(K)}}$ and $h_* + \varepsilon \in \mathscr{D}^{\text{irr}}_{PN(K)}$. Hence, $g\mathbb{I}_{\{x_{P(K)}\}}[f_* + \varepsilon] \in \mathscr{D}^{\text{irr}}_G$ because of Proposition 55$_{157}$. Furthermore, since $PN(K)$ is closed because of Lemma 78(iii)$_{183}$ and $P(PN(K)) = \emptyset$ because of Lemma 78(iv)$_{183}$, Proposition 55$_{157}$ [with $K' = PN(K)$, $f' = h_* + \varepsilon$ and $g' = 1$] also implies that $h_* + \varepsilon \in \mathscr{D}^{\text{irr}}_G$. Since $\mathscr{D}^{\text{irr}}_G$ is coherent, this allows us to conclude that $g\mathbb{I}_{\{x_{P(K)}\}}[f_* + \varepsilon] + h_* + \varepsilon \in \mathscr{D}^{\text{irr}}_G$ and therefore, due to Theorem 42$_{139}$ and Equation (2.3)$_{40}$, that $\underline{P}^{\text{irr}}_G(g\mathbb{I}_{\{x_{P(K)}\}}[f_* + \varepsilon] + h_* + \varepsilon) \geq 0$.

Hence, since $\underline{P}_G^{\mathrm{irr}}$ is coherent, we find that

$$\begin{aligned}
0 &\leq \underline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}[f_* + \varepsilon] + h_* + \varepsilon) \\
&= \underline{P}_G^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}f + g\mathbb{I}_{\{x_{P(K)}\}}\varepsilon + \varepsilon - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big)\big) \\
&= \underline{P}_G^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}f + g\mathbb{I}_{\{x_{P(K)}\}}\varepsilon\big) + \varepsilon - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big) \\
&\leq \underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f) + \varepsilon\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) + \varepsilon \\
&\qquad\qquad - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big),
\end{aligned}$$

where the second equality follows from $\mathrm{C8}_{49}$ and the last inequality from $\mathrm{C5}_{48}$, $\mathrm{C2}_{48}$ and conjugacy. If we rewrite the final inequality, we find that

$$\underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f) \geq \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big) - \varepsilon\big(\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) + 1\big)$$

and therefore, since this holds for all $\varepsilon > 0$, and because $\mathrm{C1}_{48}$ and $\mathrm{C6}_{48}$ guarantee that $\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) + 1 > 0$, it follows that

$$\underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f) \geq \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big),$$

which is already half of the proof. All that is now left to do is to establish the converse inequality.

Again, fix any $\varepsilon > 0$. Theorem 42_{139} and Equation $(2.3)_{40}$ then imply that $f_* - \varepsilon \notin \mathscr{D}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}$ and $h_* - \varepsilon/2 \notin \mathscr{D}_{PN(K)}^{\mathrm{irr}}$. Hence, by Corollary 54_{156}, we know that $g\mathbb{I}_{\{x_{P(K)}\}}[f_* - \varepsilon] + h_* - \varepsilon/2 \notin \mathscr{D}_G^{\mathrm{irr}}$ and therefore, by Theorem 42_{139} and Equations $(2.5)_{41}$ and $(2.6)_{41}$, that $\underline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}[f_* - \varepsilon] + h_* - \varepsilon) < 0$. Hence, since $\underline{P}_G^{\mathrm{irr}}$ is coherent, we find that

$$\begin{aligned}
0 &> \underline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}[f_* - \varepsilon] + h_* - \varepsilon) \\
&= \underline{P}_G^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}f - g\mathbb{I}_{\{x_{P(K)}\}}\varepsilon - \varepsilon - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big)\big) \\
&= \underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f - g\mathbb{I}_{\{x_{P(K)}\}}\varepsilon) - \varepsilon - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big) \\
&\geq \underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f) - \varepsilon\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) - \varepsilon \\
&\qquad\qquad - \underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big),
\end{aligned}$$

where the second equality follows from $\mathrm{C8}_{49}$ and the final inequality from $\mathrm{C5}_{48}$, $\mathrm{C2}_{48}$ and conjugacy. If we rewrite the final inequality, we find that

$$\underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big) > \underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f) - \varepsilon\big(\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) + 1\big)$$

and therefore, since this holds for all $\varepsilon > 0$, and because $\mathrm{C1}_{48}$ and $\mathrm{C6}_{48}$ guarantee that $\overline{P}_G^{\mathrm{irr}}(g\mathbb{I}_{\{x_{P(K)}\}}) + 1 > 0$, it follows that

$$\underline{P}_{PN(K)}^{\mathrm{irr}}\big(h + g\mathbb{I}_{\{x_{P(K)}\}}\underline{P}_{K\rfloor x_{P(K)}}^{\mathrm{irr}}(f)\big) \geq \underline{P}_G^{\mathrm{irr}}(h + g\mathbb{I}_{\{x_{P(K)}\}}f). \qquad \square$$

6.D PROOF OF THEOREM 62

Proof of Theorem 62. Direct and *reverse redundancy* follow from the consideration that every path from I to S is blocked by I in its first node and by S in its last node.

To prove *direct decomposition*, notice that every path from I to S is also a path from I to $S \cup W$. It is therefore blocked by C due to $\mathrm{AD}(I, S \cup W \mid C)$. *Reverse decomposition* is proved analogously.

To verify *direct weak union*, consider any path from $i \in I$ to $s \in S$. Since this is also a path from I to $S \cup W$, we know from $\mathrm{AD}(I, S \cup W \mid C)$ that it is blocked by C. Now let j be the first node in the path for which that path is blocked by C in j. If the path from i to s is blocked by C in j using condition $\mathrm{B1}_{163}$, $\mathrm{B2}_{163}$ or $\mathrm{B4}_{163}$, then $j \in C \subseteq C \cup W$, implying that the path is blocked by $C \cup W$ in j and concluding the proof.

So suppose that the path is blocked by C in j using condition $\mathrm{B3}_{163}$. We then have that $j \notin C$ and $D(j) \cap C = \emptyset$. If $j \notin W$ and $D(j) \cap W = \emptyset$, then $j \notin C \cup W$ and $D(j) \cap (C \cup W) = \emptyset$, implying that the path is blocked by $C \cup W$ in j and concluding the proof.

So suppose, and this is the only remaining possibility, that there is some node $t \in \{j\} \cup D(j)$ for which $t \in W$. In that case there is a directed path from j to t and one can concatenate the section from i to j with this directed path from j to t, obtaining a path from $i \in I$ to $t \in S \cup W$. This however leads to a contradiction with $\mathrm{AD}(I, S \cup W \mid C)$ because this path from i to t is not blocked by C. To see why, first consider all the nodes in the part from i to j, excluding j. The path from i to t cannot be blocked by C in these nodes, because j was the first node in the original path from i to s for which this path was blocked by C in j. It also cannot be blocked by C in the nodes in the part from j to t because this part is directed, $j \notin C$ and $D(j) \cap C = \emptyset$. This means that this possibility cannot occur, which concludes the proof of direct weak union.

Reverse weak union has a similar proof. Every path from $i \in I$ to $s \in S$ is also a path from $I \cup W$ to S and is thus blocked by C. Let j be the last node in the path for which that path is blocked by C in j. If the path from i to s is blocked by C in j using condition $\mathrm{B1}_{163}$, $\mathrm{B2}_{163}$ or $\mathrm{B4}_{163}$, then $j \in C \subseteq C \cup W$, implying that the path is blocked by $C \cup W$ in j and concluding the proof.

So suppose that the path is blocked by C in j by condition $\mathrm{B3}_{163}$. We then have that $j \notin C$ and $D(j) \cap C = \emptyset$. If $j \notin W$ and $D(j) \cap W = \emptyset$, then $j \notin C \cup W$ and $D(j) \cap (C \cup W) = \emptyset$, implying that the path is blocked by $C \cup W$ in j and concluding the proof.

So suppose, and this is the only remaining possibility, that there is some node $t \in \{j\} \cup D(j)$ for which $t \in W$. In that case there is a directed path from j to t and one can understand it as a reverse directed path from t to j and concatenate it with the section from j to s, obtaining a path from $t \in I \cup W$ to $s \in S$. This however leads to a contradiction with $\mathrm{AD}(I \cup W, S \mid C)$ because this path from t to s is not blocked by C. To see why, first consider all the nodes

in the part from j to s, excluding j. The path from t to s cannot be blocked by C in these nodes, because j was the last node in the original path from i to s for which this path was blocked by C in j. It also cannot be blocked by C in the nodes in the part from t to j because this part is a reverse directed path, $j \notin C$ and $D(j) \cap C = \emptyset$. This means that this possibility cannot occur, which concludes the proof of reverse weak union.

To prove *direct contraction*, consider any path from $i \in I$ to $w \in S \cup W$. We need to show that it is blocked by C. If $w \in S$, this follows directly from $\mathrm{AD}(I, S \mid C)$, so we can assume that $w \in W$, implying that the path from i to w is blocked by $C \cup S$ because of $\mathrm{AD}(I, W \mid C \cup S)$. Let t be one of the nodes for which the path from i to w is blocked by $C \cup S$ in t. If $t \in C$ or $t \notin C \cup S$, then the path from i to w is blocked by C in t, concluding the proof. If $t \in S \setminus C$, and this is the only remaining possibility, then $\mathrm{AD}(I, S \mid C)$ implies that the path from i to t must be blocked by C, from which one can also infer that the path from i to w is blocked by C.

Reverse contraction has a similar proof. Take any path from $w \in I \cup W$ to $s \in S$. We need to show that it is blocked by C. If $w \in I$, this follows directly from $\mathrm{AD}(I, S \mid C)$, so we can assume that $w \in W$, implying that the path from w to s is blocked by $C \cup I$ because of $\mathrm{AD}(W, S \mid C \cup I)$. Let t be one of the nodes for which the path from w to s is blocked by $C \cup I$ in t. If $t \in C$ or $t \notin C \cup I$, then the path from w to s is blocked by C in t, concluding the proof. If $t \in I \setminus C$, and this is the only remaining possibility, then $\mathrm{AD}(I, S \mid C)$ implies that the path from t to s is blocked by C, from which one can also infer that the path from w to s is blocked by C.

For the verification of *direct intersection*, consider any path from $i \in I$ to $w \in S \cup W$. We need to show that it is blocked by C. Due to the symmetry of the problem, we can assume without loss of generality that $w \in W$, implying that the path from i to w is blocked by $C \cup S$ because of $\mathrm{AD}(I, W \mid C \cup S)$. Now let t be the first node in the path from i to w for which this path is blocked by $C \cup S$ in t. If $t \in C$ or $t \notin C \cup S$, then the path from i to w is blocked by C in t, concluding the proof. If $t \in S \setminus C$, and this is the only remaining possibility, then this implies that $t \neq w$, since $S \cap W = \emptyset$ by assumption. It also implies that the path from i to t is blocked by $C \cup W$ because of $\mathrm{AD}(I, S \mid C \cup W)$. If it is blocked by some q for which $q \in C$ or $q \notin C \cup W$, then the path from i to w is blocked by C in q, concluding the proof. If $q \in W \setminus C$, and this is the only remaining possibility, then this implies that $q \neq t$, since $S \cap W = \emptyset$ by assumption. It also implies that the path from i to q is blocked by $C \cup S$ because of $\mathrm{AD}(I, W \mid C \cup S)$. However, this would in turn imply that the path from i to w is blocked by $C \cup S$ in some node of the path from i to q, contradicting the earlier assumption that t is the first node for which this is the case.

Finally, to prove *reverse intersection*, consider any path from $w \in I \cup W$ to $s \in S$. We need to show that it is blocked by C. Due to the symmetry of the problem, we can assume without loss of generality that $w \in W$, implying that the path from w to s is blocked by $C \cup I$ because of $\mathrm{AD}(W, S \mid C \cup I)$. Now

let t be the last node in the path from w to s for which this path is blocked by $C \cup I$ in t. If $t \in C$ or $t \notin C \cup I$, then the path from w to s is blocked by C in t, concluding the proof. If $t \in I \setminus C$, and this is the only remaining possibility, then this implies that $t \neq w$, since $I \cap W = \emptyset$ by assumption. It also implies that the path from t to s is blocked by $C \cup W$ because of $\mathrm{AD}(I, S \mid C \cup W)$. If it is blocked by some q for which $q \in C$ or $q \notin C \cup W$, then the path from w to s is blocked by C in q, concluding the proof. If $q \in W \setminus C$, and this is then only remaining possibility, then this implies that $q \neq t$, since $I \cap W = \emptyset$ by assumption. It also implies that the path from q to s is blocked by $C \cup I$ because of $\mathrm{AD}(W, S \mid C \cup I)$. However, this would in turn imply that the path from w to s is blocked by $C \cup I$ in some node of the path from q to s, contradicting the earlier assumption that t is the last node for which this is the case. □

6.E PROOF OF THEOREM 64

Our proof for Theorem 64_{167} makes use of the following two sets. For all $I, C \subseteq G$, we define

$$S^*_{I,C} := \{s \in G \colon \mathrm{AD}(I, \{s\} \mid C)\}$$

and

$$K^*_{I,C} := \{k \in S^*_{I,C} \colon (\exists s \in S^*_{I,C} \setminus C)\, s \sqsubseteq k\}.$$

The following result establishes some properties of these sets.

Proposition 81. *Consider any $I, C \subseteq G$. Then the following statements hold:*

(i) *$K^*_{I,C}$ is a closed subset of G;*

(ii) *$P(K^*_{I,C}) \subseteq S^*_{I,C} \setminus K^*_{I,C} \subseteq C \subseteq S^*_{I,C}$;*

(iii) *$S^*_{I,C} \setminus C = K^*_{I,C} \setminus C$;*

(iv) *$D(K^*_{I,C}) \cap C = \emptyset$;*

(v) *$I \setminus C \subseteq N(K^*_{I,C}) \setminus C$.*

Proof of Proposition 81. (i). Fix $u, v \in K^*_{I,C}$ and $k \in G$ such that $u \sqsubseteq k \sqsubseteq v$ and assume *ex absurdo* that $k \notin K^*_{I,C}$. Since $u \in K^*_{I,C}$, we can infer the existence of some $s \in S^*_{I,C}$ such that $s \notin C$ and $s \sqsubseteq u \sqsubseteq k$, implying that $k \notin S^*_{I,C}$ because $k \in S^*_{I,C}$ would imply $k \in K^*_{I,C}$, contradicting $k \notin K^*_{I,C}$. So we now know that there is a path from some $i \in I$ to k that is not blocked by C. Since $s \sqsubseteq k$, we also have a directed path from s to k and thus a reverse directed path from k to s. Concatenating the path both from i to k and the reverse directed path from k to s, we obtain a path from $i \in I$ to s, which should be blocked by C since $s \in S^*_{I,C}$. The only way for this to be possible is if $k \notin C$ and $D(k) \cap C = \emptyset$. However,

then the path from $i \in I$ to v, formed by concatenating the path from i to k and a directed path from k to v, is not blocked by C , contradicting $v \in K^*_{I,C} \subseteq S^*_{I,C}$.

(ii)$_{\frown}$. We begin by proving that $P(K^*_{I,C}) \subseteq S^*_{I,C} \setminus K^*_{I,C}$. Consider any node $p \in P(K^*_{I,C})$ and let k be (one of) the child(ren) of p for which $k \in K^*_{I,C} \subseteq S^*_{I,C}$. Assume *ex absurdo* that $p \notin S^*_{I,C}$. This means that there is a path from some $i \in I$ to p that is not blocked by C. If $k \notin C$, then the concatenation of the path from i to p with the node k, yields a path from $i \in I$ to $k \in S^*_{I,C}$ that is not blocked by C, a contradiction. If $k \in C$, then $k \in K^*_{I,C}$ implies the existence of some $s \in S^*_{I,C}$ such that $s \sqsubset k$ and $s \notin C$. Since $s \sqsubset k$, we can now construct a directed path from s to k, yielding a reverse directed path from k to s. If we concatenate the path from i to p with this reverse directed path from k to s, we obtain a path from $i \in I$ to $s \in S^*_{I,C}$ that is not blocked by C, a contradiction. Hence, we may conclude that $P(K^*_{I,C}) \subseteq S^*_{I,C}$. Since $P(K^*_{I,C}) \cap K^*_{I,C} = \emptyset$, this implies that $P(K^*_{I,C}) \subseteq S^*_{I,C} \setminus K^*_{I,C}$.

Next, we prove that $S^*_{I,C} \setminus K^*_{I,C} \subseteq C$. Consider any $k \in S^*_{I,C} \setminus K^*_{I,C}$ and assume *ex absurdo* that $k \notin C$. Then $s = k$ is an element of $S^*_{I,C}$ such that $s \sqsubseteq k$ and $s \notin C$, which implies that $k \in K^*_{I,C}$, a contradiction. Hence, we may conclude that $k \in C$.

The final inclusion—$C \subseteq S^*_{I,C}$—is trivial.

(iii)$_{\frown}$. Since we already know from (ii)$_{\frown}$ that $S^*_{I,C} \setminus K^*_{I,C} \subseteq C$, it follows that $S^*_{I,C} \setminus C \subseteq S^*_{I,C} \setminus (S^*_{I,C} \setminus K^*_{I,C}) = K^*_{I,C}$, where the last equality holds because $K^*_{I,C} \subseteq S^*_{I,C}$. Clearly, this implies that $S^*_{I,C} \setminus C \subseteq K^*_{I,C} \setminus C$. Since $K^*_{I,C} \subseteq S^*_{I,C}$ also implies that $K^*_{I,C} \setminus C \subseteq S^*_{I,C} \setminus C$, we find that $K^*_{I,C} \setminus C = S^*_{I,C} \setminus C$

(iv)$_{\frown}$. Assume *ex absurdo* that $D(K^*_{I,C}) \cap C \neq \emptyset$ and fix any $t \in D(K^*_{I,C}) \cap C$. Since $t \in C$, we know from (ii)$_{\frown}$ that $t \in S^*_{I,C}$. On the other hand, $t \in D(K^*_{I,C})$ implies the existence of some $k \in K^*_{I,C}$ such that $k \sqsubseteq t$. Since $k \in K^*_{I,C}$ in turn implies the existence of some $s \in S^*_{I,C}$ such that $s \sqsubseteq k \sqsubseteq t$ and $s \notin C$, we obtain from $t \in S^*_{I,C}$ that $t \in K^*_{I,C}$, contradicting $t \in D(K^*_{I,C})$. Hence, we may conclude that $D(K^*_{I,C}) \cap C = \emptyset$.

(v). $I \setminus C \subseteq N(K^*_{I,C}) \setminus C$ follows trivially from $I \setminus C \subseteq N(K^*_{I,C})$, so it suffices to prove the latter statement. Consider any $i \in I \setminus C$, implying that $i \notin C$. Since $i \in G = N(K^*_{I,C}) \cup P(K^*_{I,C}) \cup K^*_{I,C} \cup D(K^*_{I,C})$, it suffices to prove that $i \notin P(K^*_{I,C})$, $i \notin K^*_{I,C}$ and $i \notin D(K^*_{I,C})$. First, i cannot be an element of $P(K^*_{I,C})$ because (ii)$_{\frown}$ then implies that $i \in C$, a contradiction. Second, i is not an element of $K^*_{I,C}$ because then $i \in S^*_{I,C}$, which implies that the trivial path from i to i should be blocked by C, again yielding a contradiction with $i \notin C$. Third, assume *ex absurdo* that $i \in D(K^*_{I,C})$. We have shown in the proof of $D(K^*_{I,C}) \cap C = \emptyset$ that this implies the existence of some $s \in S^*_{I,C}$ such that $s \sqsubseteq i$ and $s \notin C$. Since $s \sqsubseteq i$, we can now construct a directed path from s to i, yielding a reverse directed path from i to s that is not blocked by C [because it is a reverse directed path, and because neither i nor s belong to C], which contradicts $s \in S^*_{I,C}$. □

Using these properties, the proof for Theorem 64_{167} is now relatively easy.

Proof of Theorem 64$_{167}$. Consider pairwise disjoint subsets I, S and C of G.

First assume that $\mathrm{AD}(I,S \mid C)$. This implies that $S \subseteq S^*_{I,C}$. Let $K := K^*_{I,C}$. Since I, S and C are pairwise disjoint, it follows from Proposition 81$_{203}$ that K is a closed subset of G for which $P(K) \subseteq C$, $S \subseteq K$, $D(K) \cap C = \emptyset$ and $I \subseteq N(K)$.

Conversely, assume that there is a closed subset K of G such that $S \subseteq K$, $P(K) \subseteq C$, $I \subseteq N(K)$ and $D(K) \cap C = \emptyset$. Since $\mathrm{AD}(N(K),K \mid P(K))$ because of Lemma 82, direct and reverse decomposition [see Theorem 62$_{165}$] imply that

$$\mathrm{AD}(I \cup (N(K) \cap C), S \cup (K \cap C) \mid P(K)).$$

By applying direct and reverse weak union [see Theorem 62$_{165}$] this in turn implies that

$$\mathrm{AD}(I,S \mid P(K) \cup (N(K) \cap C) \cup (K \cap C)). \tag{6.19}$$

Since we know that $P(K) \subseteq C$ and $D(K) \cap C = \emptyset$, Lemma 78(ii)$_{183}$ allows us to infer that $P(K) \cup (N(K) \cap C) \cup (K \cap C) = C$ and therefore, by Equation (6.19), that $\mathrm{AD}(I,S \mid C)$. □

Lemma 82. *For any closed subset K of G:* $\mathrm{AD}(N(K),K \mid P(K))$.

Proof of Lemma 82. Consider any path $t = s_1, \ldots, s_n = k$, $n \geq 1$, from a node $t \in N(K)$ to a node $k \in K$. We will prove that this path is blocked by $P(K)$.

Let i be the last index in $\{1,\ldots,n\}$ for which $s_i \notin K$ [i is well-defined because $s_1 = t \notin K$]. Since $s_n = k \in K$, we know that $i < n$ and we can therefore consider s_{i+1}. Clearly, by definition of i, it holds that $s_{i+1} \in K$. We now consider two mutually exclusive and exhaustive cases: $s_i \rightarrow s_{i+1}$ and $s_i \leftarrow s_{i+1}$.

We first consider the case $s_i \rightarrow s_{i+1}$. Then $s_i \in P(s_{i+1})$. Since $s_i \notin K$ and $s_{i+1} \in K$, this implies that $s_i \in P(K)$. The path $t = s_1, \ldots, s_n = k$ is therefore blocked by $P(K)$ in s_i [using blocking condition B1$_{163}$ or B2$_{163}$].

Next, assume that $s_i \leftarrow s_{i+1}$. Since $s_1 = t \in N(K)$, it follows from Lemma 78(ii)$_{183}$ that $s_1 \notin D(K)$. Since $s_{i+1} \in K$, this implies that there is at least one $j \in \{2,\ldots,i+1\}$ such that $s_{j-1} \rightarrow s_j$. Let j^* be the largest $j \in \{2,\ldots,i+1\}$ for which this is the case. Since $s_i \leftarrow s_{i+1}$ implies that $j^* < i+1$, it follows from the definition of j^* that $s_{j^*-1} \rightarrow s_{j^*} \leftarrow s_{j^*+1}$ and that $s_{j^*} \in D(s_{i+1})$. Consider now any $s \in \{s_{j^*}\} \cup D(s_{j^*})$. Then $s \in D(s_{i+1})$ because $s_{j^*} \in D(s_{i+1})$. Since $s_{i+1} \in K$, this implies that $s \in K \cup D(K)$. Using Lemma 78(ii)$_{183}$, this allows us to infer to $s \notin P(K)$. Since this holds for every $s \in \{s_{j^*}\} \cup D(s_{j^*})$, we find that $s_{j^*} \notin P(K)$ and $D(s_{j^*}) \cap P(K) = \emptyset$. Since we also know that $1 < j^* < i+1 \leq n$, we can combine this with the fact that $s_{j^*-1} \rightarrow s_{j^*} \leftarrow s_{j^*+1}$ to find that the path $t = s_1, \ldots, s_n = k$ is blocked by $P(K)$ in s_{j^*} [using blocking condition B3$_{163}$]. □

6.F PROOF OF PROPOSITION 67

Proof of Proposition 67$_{173}$. Consider a partition $G_1,\dots,G_m$ of G such that $P(G_i)=\emptyset$ for all $i\in\{1,\dots,m\}$. Then, clearly, the sets G_i are disconnected from one another. Indeed, assume *ex absurdo* that there is an arrow from a node $s\in G_i$ to a node $t\in G_j$, with $i,j\in\{1,\dots,m\}$ and $i\neq j$. Then s is an element of $P(G_j)$, which contradicts our assumption that $P(G_j)=\emptyset$. Conversely, and similarly, one can easily see that if the DAG consists of disconnected sets $G_1,\dots,G_m$, then $P(G_i)=\emptyset$ for all $i\in\{1,\dots,m\}$.

In any case, whenever one of these two equivalent conditions holds, then for any $i\in\{1,\dots,m\}$ and $s\in G_i$, we have that $P_{G_i}(s)=P(s)\subseteq G_i$ and $D_{G_i}(s)=D(s)\subseteq G_i$, and therefore also that $PN(s)=PN_{G_i}(s)\cup(G\setminus G_i)$.

Now let $\mathscr{D}_G^{\mathrm{irr}}$ be the irrelevant natural extension of the complete network, as given by Equation (5.2)$_{136}$, and, for all $i\in\{1,\dots,m\}$, let $\mathscr{D}_{G_i}^{\mathrm{irr}}$ be the irrelevant natural extension of the network that has the sub-DAG associated with G_i as its graphical structure, as given by Equation (6.2)$_{156}$. We then find that

$$\begin{aligned}
\otimes_{i=1}^n\mathscr{D}_{G_i}^{\mathrm{irr}} &= \mathrm{posi}(\{\mathbb{I}_{\{z_{G\setminus G_i}\}}f\colon i\in\{1,\dots,m\},\ z_{G\setminus G_i}\in\mathscr{X}_{G\setminus G_i},\ f\in\mathscr{D}_{G_i}^{\mathrm{irr}}\})\\
&= \mathrm{posi}(\{\mathbb{I}_{\{z_{G\setminus G_i}\}}f\colon i\in\{1,\dots,m\},\ z_{G\setminus G_i}\in\mathscr{X}_{G\setminus G_i},\\
&\qquad f\in\mathrm{posi}(\{\mathbb{I}_{\{z_{PN_{G_i}(s)}\}}f'\colon s\in G_i, z_{PN_{G_i}(s)}\in\mathscr{X}_{PN_{G_i}(s)},\\
&\qquad\qquad f'\in\mathscr{D}_{s\rfloor z_{P(s)}}\})\})\\
&= \mathrm{posi}(\{\mathbb{I}_{\{z_{G\setminus G_i}\}}\mathbb{I}_{\{z_{PN_{G_i}(s)}\}}f'\colon i\in\{1,\dots,m\},\ z_{G\setminus G_i}\in\mathscr{X}_{G\setminus G_i},\ s\in G_i,\\
&\qquad z_{PN_{G_i}(s)}\in\mathscr{X}_{PN_{G_i}(s)}, f'\in\mathscr{D}_{s\rfloor z_{P(s)}}\})\\
&= \mathrm{posi}(\{\mathbb{I}_{\{z_{PN(s)}\}}f'\colon s\in G,\ z_{PN(s)}\in\mathscr{X}_{PN(s)},\ f'\in\mathscr{D}_{s\rfloor z_{P(s)}}\}) = \mathscr{D}_G^{\mathrm{irr}}.
\end{aligned}$$

The first equality in this derivation follows from Equation (5.2)$_{136}$. The second one follows from Equation (6.2)$_{156}$ and the earlier proved fact that, for all $i\in\{1,\dots,m\}$ and every $s\in G_i$: $P_{G_i}(s)=P(s)$. The third equality is due to the definition of the posi operator and the fourth equality holds because $G_1\dots,G_m$ is a partition of G and because we have already shown that, for all $i\in\{1,\dots,m\}$ and every $s\in G_i$, $PN(s)=PN_{G_i}(s)\cup(G\setminus G_i)$. The final equality follows from the definition of $\mathscr{D}_G^{\mathrm{irr}}$; see Equation (5.2)$_{136}$. Hence, we already know that $\mathscr{D}_G^{\mathrm{irr}}=\otimes_{i=1}^m\mathscr{D}_{G_i}^{\mathrm{irr}}$.

Without loss of generality, we now assume that $\mathscr{D}_{s\rfloor z_{P(s)}}:=\mathscr{D}_{\underline{P}_{s\rfloor z_{P(s)}}}$ for all $s\in G$ and $z_{P(s)}\in\mathscr{X}_{P(s)}$. For all $i\in\{1,\dots,m\}$, we then infer from Theorem 42$_{139}$ that $\underline{P}_{G_i}^{\mathrm{irr}}(\cdot)=\underline{P}_{\mathscr{D}_{G_i}^{\mathrm{irr}}}$. As explained in the main text, this implies that $\otimes_{i=1}^m\underline{P}_{G_i}^{\mathrm{irr}}(\cdot)=\underline{P}_{\otimes_{i=1}^m\mathscr{D}_{G_i}^{\mathrm{irr}}}$ and therefore, since $\mathscr{D}_G^{\mathrm{irr}}=\otimes_{i=1}^m\mathscr{D}_{G_i}^{\mathrm{irr}}$, that $\otimes_{i=1}^m\underline{P}_{G_i}^{\mathrm{irr}}(\cdot)=\underline{P}_{\mathscr{D}_G^{\mathrm{irr}}}$. Since Theorem 42$_{139}$ also tells us that $\underline{P}_G^{\mathrm{irr}}(\cdot)=\underline{P}_{\mathscr{D}_G^{\mathrm{irr}}}$, we find that $\underline{P}_G^{\mathrm{irr}}(\cdot)=\otimes_{i=1}^m\underline{P}_{G_i}^{\mathrm{irr}}(\cdot)$. □

6.G PROOF OF PROPOSITION 71

Proof of Proposition 71$_{179}$. Due to Lemma 77(v)$_{182}$&(vi)$_{182}$, we know that T and S are closed subsets of G and that $D(S) = \emptyset$.

First consider any $s \in T$. Then, by assumption, $S \subseteq D(s)$, which implies that $PN(s) \subseteq T$. By combining this with Lemma 76$_{182}$ and 79(iii)$_{185}$, we find that $PN_T(s) = PN(s)$ and $P_T(s) = P(s)$.

Next, consider any $s \in S$. Lemma 77(vii)$_{182}$ then implies that $T \subseteq PN(s)$. By combining this with Lemma 79(iii)$_{185}$, we find that $PN_S(s) = PN(s) \setminus T$ and $PN(s) = T \cup PN_S(s)$. Also, due to Lemma 76$_{182}$: $P_S(s) = P(s) \setminus T$.

The following series of equalities establishes that $\mathscr{D}_G^{\text{irr}} = \mathscr{D}_T^{\text{irr}} \oslash \mathscr{D}_{S\rfloor x_{P(S)}}^{\text{irr}}$.

$$
\begin{aligned}
\mathscr{D}_T^{\text{irr}} \oslash \mathscr{D}_{S\rfloor x_{P(S)}}^{\text{irr}} &= \Big\{ f \in \mathscr{G}(\mathscr{X}_G)) \setminus \{0\} \colon f = f_T + \textstyle\sum_{x_T \in \mathscr{X}_T} \mathbb{I}_{\{x_T\}} f_{S\rfloor x_T}, \\
&\qquad f_T \in \mathscr{D}_T^{\text{irr}} \cup \{0\}, (\forall x_T \in \mathscr{X}_T)\, f_{S\rfloor x_T} \in \mathscr{D}_{S\rfloor x_{P(S)}}^{\text{irr}} \cup \{0\} \Big\} \\
&= \Big\{ f \in \mathscr{G}(\mathscr{X}_G)) \setminus \{0\} \colon f = f_T + \textstyle\sum_{x_T \in \mathscr{X}_T} \mathbb{I}_{\{x_T\}} f_{S\rfloor x_T}, \\
&\qquad \Big(f_T = \textstyle\sum_{s \in T} \sum_{z_{PN_T(s)} \in \mathscr{X}_{PN_T(s)}} \mathbb{I}_{\{z_{PN_T(s)}\}} f_{s\rfloor z_{PN_T(s)}}, \\
&\qquad\quad (\forall s \in T)\, (\forall z_{PN_T(s)} \in \mathscr{X}_{PN_T(s)})\, f_{s\rfloor z_{PN_T(s)}} \in \mathscr{D}_{s\rfloor z_{P_T(s)}} \cup \{0\} \Big), \\
&\qquad (\forall x_T \in \mathscr{X}_T) \Big(f_{S\rfloor x_T} = \textstyle\sum_{s \in S} \sum_{z_{PN_S(s)} \in \mathscr{X}_{PN_S(s)}} \mathbb{I}_{\{z_{PN_S(s)}\}} f_{s\rfloor (z_{PN_S(s)}, x_T)}, \\
&\qquad\quad (\forall s \in S)\, (\forall z_{PN_S(s)} \in \mathscr{X}_{PN_S(s)}) \\
&\qquad\quad f_{s\rfloor (z_{PN_S(s)}, x_T)} \in \mathscr{D}_{s\rfloor (z_{P_S(s)}, x_{P(s) \setminus P_S(s)})} \cup \{0\} \Big) \Big\} \\
&= \Big\{ f \in \mathscr{G}(\mathscr{X}_G)) \setminus \{0\} \colon f = f_T + \textstyle\sum_{x_T \in \mathscr{X}_T} \mathbb{I}_{\{x_T\}} f_{S\rfloor x_T}, \\
&\qquad \Big(f_T = \textstyle\sum_{s \in T} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s\rfloor z_{PN(s)}}, \\
&\qquad\quad (\forall s \in T)\, (\forall z_{PN(s)} \in \mathscr{X}_{PN(s)})\, f_{s\rfloor z_{PN(s)}} \in \mathscr{D}_{s\rfloor z_{P(s)}} \cup \{0\} \Big), \\
&\qquad (\forall x_T \in \mathscr{X}_T) \Big(f_{S\rfloor x_T} = \textstyle\sum_{s \in S} \sum_{z_{PN(s) \setminus T} \in \mathscr{X}_{PN(s) \setminus T}} \mathbb{I}_{\{z_{PN(s) \setminus T}\}} f_{s\rfloor (z_{PN(s) \setminus T}, x_T)}, \\
&\qquad\quad (\forall s \in S)\, (\forall z_{PN(s) \setminus T} \in \mathscr{X}_{PN(s) \setminus T}) \\
&\qquad\quad f_{s\rfloor (z_{PN(s) \setminus T}, x_T)} \in \mathscr{D}_{s\rfloor (z_{P(s) \setminus T}, x_{P(s) \cap T})} \cup \{0\} \Big) \Big\} \\
&= \Big\{ f \in \mathscr{G}(\mathscr{X}_G)) \setminus \{0\} \colon f = \textstyle\sum_{s \in G} \sum_{z_{PN(s)} \in \mathscr{X}_{PN(s)}} \mathbb{I}_{\{z_{PN(s)}\}} f_{s\rfloor z_{PN(s)}}, \\
&\qquad (\forall s \in G)\, (\forall z_{PN(s)} \in \mathscr{X}_{PN(s)})\, f_{s\rfloor z_{PN(s)}} \in \mathscr{D}_{s\rfloor z_{P(s)}} \cup \{0\} \Big\} \\
&= \mathscr{D}_G^{\text{irr}}. \qquad (6.20)
\end{aligned}
$$

The second and the last equality follow from Proposition 36$_{136}$. The third and

fourth equality follow from some basic manipulations and the properties that we derived in the beginning of this proof.

The other three equations of this Proposition now follow from Proposition 43_{140}, Theorem 46_{142}, Proposition 49_{143} and the connections that were established in Section 6.7_{175} between the four different expressions for our generalised notion of marginal extension. □

6.H PROOF OF THEOREM 73

Proof of Theorem 73$_{181}$. First assume that $\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) = 0$. For every $B_K \in \mathscr{P}_{\emptyset}(\mathscr{X}_K)$, this implies that $\overline{P}^{\mathrm{irr}}_{G}(B_K \times \{x_{P(K)}\} \times B_{N(K)}) = 0$ and therefore, we infer from Equation $(6.15)_{180}$ that

$$\underline{R}^{\mathrm{irr}}_{K}(\cdot\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K}(\cdot\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}).$$

Since this holds for all $B_K \in \mathscr{P}_{\emptyset}(\mathscr{X}_K)$, we find that

$$\underline{R}^{\mathrm{irr}}_{K}(\cdot\rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K}(\cdot\rfloor \cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot),$$

where the last equality follows from Corollary 60_{160}.

Next, assume that $\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) > 0$. We need to prove for all $B_K \in \mathscr{P}_{\emptyset}(\mathscr{X}_K)$ that $\underline{R}^{\mathrm{irr}}_{K}(\cdot\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor B_K)$. So fix any $B_K \in \mathscr{P}_{\emptyset}(\mathscr{X}_K)$. It follows from Corollary 57_{159} and conjugacy that

$$\overline{P}^{\mathrm{irr}}_{G}(B_K \times \{x_{P(K)}\} \times B_{N(K)}) = \overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)})\overline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(B_K). \quad (6.21)$$

We now consider two cases: $\overline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(B_K) = 0$ and $\overline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(B_K) > 0$.

If $\overline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(B_K) = 0$ then also $\overline{P}^{\mathrm{irr}}_{G}(B_K \times \{x_{P(K)}\} \times B_{N(K)}) = 0$ because of Equation (6.21). Hence, we find that

$$\begin{aligned}\underline{R}^{\mathrm{irr}}_{K}(\cdot\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) &= \underline{P}^{\mathrm{irr}}_{K}(\cdot\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) \\ &= \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor B_K) = \underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor B_K),\end{aligned}$$

where the first and last equality follow from Equation $(6.15)_{180}$ and the second equality follows from Corollary 60_{160}.

If $\overline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(B_K) > 0$ then also $\overline{P}^{\mathrm{irr}}_{G}(B_K \times \{x_{P(K)}\} \times B_{N(K)}) > 0$ because of Equation (6.21) and the assumption that $\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) > 0$. Consider any $f \in \mathscr{G}(B_K)$. It follows from Equation $(3.11)_{91}$ that

$$\begin{aligned}&\underline{R}^{\mathrm{irr}}_{K}(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) \\ &= \inf\big\{P_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)})\colon P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{G}(\cdot\rfloor\cdot)}, \\ &\qquad\qquad P_G(B_K \times \{x_{P(K)}\} \times B_{N(K)}) > 0\big\} \quad (6.22)\end{aligned}$$

and

$$\underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K) = \inf\big\{P_K(f\rfloor B_K)\colon P_K(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}},\, P_K(B_K) > 0\big\}. \quad (6.23)$$

The only thing that we still need to show in order to finish this proof is that $\underline{R}^{\mathrm{irr}}_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K)$.

We first show that $\underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K) \leq \underline{R}^{\mathrm{irr}}_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)})$. Consider any $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)}$ such that $P_G(B_K \times \{x_{P(K)}\} \times B_{N(K)}) > 0$ and let $P'_K(\cdot\rfloor\cdot) := P_K(\cdot\rfloor\cdot \times \{x_{P(K)}\} \times B_{N(K)})$. Since Corollary 60_{160} implies that $\underline{P}^{\mathrm{irr}}_K(\cdot\rfloor\cdot \times \{x_{P(K)}\} \times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$, we infer from $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)}$ that $P'_K(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)}$. Also, since it follows from Bayes's rule that

$$P_G(B_K \times \{x_{P(K)}\} \times B_{N(K)}) = P_K(B_K\rfloor\{x_{P(K)}\} \times B_{N(K)})P_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}), \quad (6.24)$$

we infer from $P_G(B_K \times \{x_{P(K)}\} \times B_{N(K)}) > 0$ that $P'_K(B_K) > 0$. Equation (6.23) therefore implies that

$$\underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K) \leq P'_K(f\rfloor B_K) = P_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}).$$

Hence, it follows from Equation $(6.22)_{\frown}$—and from the fact that this is true for every $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)}$ such that $P_G(B_K \times \{x_{P(K)}\} \times B_{N(K)}) > 0$—that $\underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K) \leq \underline{R}^{\mathrm{irr}}_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)})$.

Next, we show that $\underline{R}^{\mathrm{irr}}_K(f\rfloor B_K \times \{x_{P(K)}\} \times B_{N(K)}) \leq \underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K)$. Consider any $P'_K(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)}$ such that $P'_K(B_K) > 0$. Since it follows from coherence [$C5_{48}$] that

$$\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) \leq \sum_{z_{N(K)} \in \mathscr{X}_{N(K)}} \overline{P}^{\mathrm{irr}}_{PN(K)}\big((x_{P(K)}, z_{N(K)})\big),$$

we infer from $\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times B_{N(K)}) > 0$ that there is some $z^*_{N(K)} \in \mathscr{X}_{N(K)}$ such that $\overline{P}^{\mathrm{irr}}_{PN(K)}\big((x_{P(K)}, z^*_{N(K)})\big) > 0$. Let $x^*_{PN(K)} := (x_{P(K)}, z^*_{N(K)})$. Consider any $s \in PN(K)$. Proposition 58_{159} then implies that $\overline{P}_{s\rfloor x^*_{P(s)}}(x^*_s) > 0$. Hence, by Theorem 7_{52}, there is some linear prevision $P^*_{s\rfloor x^*_{P(K)}} \in \mathscr{M}_{\underline{P}_{s\rfloor x^*_{P(s)}}}$ such that $P^*_{s\rfloor x^*_{P(K)}}(x^*_s) > 0$. We now construct a new credal network that has the same graphical structure and whose local models are given by

$$\underline{P}^*_{s\rfloor z_{P(s)}} = \begin{cases} P^*_{s\rfloor x^*_{P(s)}} & \text{if } s \in PN(K) \text{ and } z_{P(s)} = x^*_{P(s)}; \\ \underline{P}_{s\rfloor z_{P(s)}} & \text{otherwise.} \end{cases}$$

Let $\underline{P}^{\mathrm{irr}*}_G(\cdot\rfloor\cdot)$ be the corresponding 'global' irrelevant natural extension, and similarly for $\underline{P}^{\mathrm{irr}*}_{PN(K)}(\cdot\rfloor\cdot)$ and $\underline{P}^{\mathrm{irr}*}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$. Since $\underline{P}^*_{s\rfloor z_{P(s)}} = \underline{P}_{s\rfloor z_{P(s)}}$ for all

$s \in K$ and $z_{P(s)} \in \mathscr{X}_{P(s)}$, it follows that $\underline{P}^{\mathrm{irr}*}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot) = \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$. Therefore, since Corollary 60_{160} implies that $\underline{P}^{\mathrm{irr}*}_{K}(\cdot\rfloor\cdot\times\{x_{P(K)}\}\times B_{N(K)}) = \underline{P}^{\mathrm{irr}*}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$, we find that $\underline{P}^{\mathrm{irr}*}_{K}(\cdot\rfloor\cdot\times\{x_{P(K)}\}\times B_{N(K)}) = \underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$. Hence, since $P'_K(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)}$, it follows from Theorem 33_{110} that there is some $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}*}_{G}(\cdot\rfloor\cdot)}$ such that $P_K(\cdot\rfloor\cdot\times\{x_{P(K)}\}\times B_{N(K)}) = P'_K(\cdot\rfloor\cdot)$. Since $P'_K(B_K) > 0$, this also implies that $P_K(B_K\rfloor\{x_{P(K)}\}\times B_{N(K)}) > 0$. Due to Proposition 58_{159}, we find that

$$\underline{P}^{\mathrm{irr}*}_{PN(K)}(x^*_{PN(K)}) = \prod_{s\in PN(K)} \underline{P}^*_{s\rfloor x^*_{P(s)}}(x^*_s) = \prod_{s\in PN(K)} P^*_{s\rfloor x^*_{P(s)}}(x^*_s) > 0,$$

which, because of coherence [$C7_{49}$], implies that $\underline{P}^{\mathrm{irr}*}_{PN(K)}(\{x_{P(K)}\}\times B_{N(K)}) > 0$. Hence, since $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}*}_{G}(\cdot\rfloor\cdot)}$, we find that $P_{PN(K)}(\{x_{P(K)}\}\times B_{N(K)}) > 0$. Since we already know that $P_K(B_K\rfloor\{x_{P(K)}\}\times B_{N(K)}) > 0$, Equation $(6.24)_{\curvearrowleft}$ now tells us that $P_G(B_K\times\{x_{P(K)}\}\times B_{N(K)}) > 0$. Since the local models that are used to construct $\underline{P}^{\mathrm{irr}*}_{G}(\cdot\rfloor\cdot)$ dominate the corresponding local models of $\underline{P}^{\mathrm{irr}}_{G}(\cdot\rfloor\cdot)$, we know that $\underline{P}^{\mathrm{irr}*}_{G}(\cdot\rfloor\cdot)$ dominates $\underline{P}^{\mathrm{irr}}_{G}(\cdot\rfloor\cdot)$. Since $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}*}_{G}(\cdot\rfloor\cdot)}$, this implies that $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{G}(\cdot\rfloor\cdot)}$. We conclude from all this that there is some $P_G(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{G}(\cdot\rfloor\cdot)}$ such that $P_K(\cdot\rfloor\cdot\times\{x_{P(K)}\}\times B_{N(K)}) = P'_K(\cdot\rfloor\cdot)$ and $P_G(B_K\times\{x_{P(K)}\}\times B_{N(K)}) > 0$. Equation $(6.22)_{208}$ therefore implies that

$$\underline{R}^{\mathrm{irr}}_{K}(f\rfloor B_K\times\{x_{P(K)}\}\times B_{N(K)}) \leq P_K(f\rfloor B_K\times\{x_{P(K)}\}\times B_{N(K)}) = P'_K(f\rfloor B_K).$$

Since this is true for every $P'_K(\cdot\rfloor\cdot) \in \mathscr{M}_{\underline{P}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)}$ such that $P'_K(B_K) > 0$, Equation $(6.23)_{\curvearrowleft}$ implies that $\underline{R}^{\mathrm{irr}}_{K}(f\rfloor B_K\times\{x_{P(K)}\}\times B_{N(K)}) \leq \underline{R}^{\mathrm{irr}}_{K\rfloor x_{P(K)}}(f\rfloor B_K)$. $\square$

7

INFERENCE ALGORITHMS

"The question of whether a computer can think is no more interesting than the question of whether a submarine can swim."

Edsger W. Dijkstra

If it were not for this final chapter, all the definitions, theoretical properties and philosophical discussions in this dissertation would be nothing more than a—rather masochistic—thought experiment. Although I must admit that I am rather fond of mathematical masochism, it seems to me that this is a rather peculiar personality feature, shared by academics only, and even then, only by a few. Therefore, besides for my personal enjoyment, the main motivation for developing this theory of credal networks under epistemic irrelevance is of course not merely its mathematical beauty. The ultimate goal is to use this theory to solve practical problems.

In principle, credal networks can be applied to solve the same problems as Bayesian networks. The added advantage is that in those cases where it is too costly, time-consuming or simply unrealistic to provide the precise local conditional probabilities that are required for the construction of a Bayesian network, we can use imprecise-probabilistic local models instead, making the resulting global model and the inferences it leads to more reliable. This approach has been successfully applied to various problems. Two important types of applications are knowledge-based expert systems—see for example References [1, 3, 36]—and data-based classifiers—Reference [13] provides a recent overview of the growing literature on credal classification; an early ex-

ample can be found in Reference [119]. Other examples are filtering [6] and state estimation [30] in imprecise hidden Markov models.

With the exception of these last two examples, both of which adopt epistemic irrelevance, nearly all applications of credal networks have been developed using credal networks under strong independence. The main—and often only—reason for this choice is the simple fact that, for credal networks under epistemic irrelevance, there is no generally applicable algorithm that is capable of computing the inferences that are required for these applications—that can evaluate $\underline{P}_G^{\mathrm{irr}}(f\rfloor B)$, for specific choices of f and B. If such an algorithm were available, the applications that are now using credal networks under strong independence could be dealt with equally well using credal networks under epistemic irrelevance.

In fact, we think that in many cases, credal networks under epistemic irrelevance would be better suited for the job. On the one hand, from a philosophical point of view, as has been acknowledged by several authors [16, 46, 81, 106], assessments of epistemic irrelevance—or epistemic independence—naturally have a broader scope because of their more intuitive meaning, which is especially important in the case of knowledge-based expert systems; see Section 4.4_{120} as well. On the other hand, the reason for the—relative—abundance of algorithms for credal networks under strong independence is not that they are somehow inherently more tractable. On the contrary, initial complexity results seem to suggest otherwise.

Indeed, the main conclusion of a very recent paper [67] about the complexity of inference in these two types of credal networks was (a) that there are inferences—even in singly-connected networks—that are NP-hard according to both types of networks [67, Theorem 1] and (b) that each type of network has a particular class of 'simple' inferences that are known to be computable in polynomial time. The type of inference problem that is 'simple' is the computation of the lower or upper prevision of a gamble on a single variable—the query variable—conditional on the value of some set of evidence variables. For credal networks under strong independence, this type of inference is efficient if the network is singly connected—is a polytree—and consists of binary variables only [55]. For credal networks under epistemic irrelevance, this type of inference is efficient if the network has a tree structure, regardless of the cardinality of the variables [42]. Given that tree topologies are very common—Markov chains, hidden Markov models and naive classifiers are popular examples—and that many networks contain non-binary nodes, this suggests that inferences in credal networks under epistemic irrelevance are at least as tractable as those in credal networks under strong independence, especially since inference in trees with non-binary nodes is known to be NP-hard for credal networks under strong independence [67, Theorem 2].

So, given that these complexity results tend to favour credal networks under epistemic irrelevance over credal networks under strong independence, why then is this preference not shared by practitioners? The answer is very sim-

ple: because there are far more algorithms available; see Reference [2, Section 10.5.3] for a recent overview of algorithmic developments for credal networks under strong independence. Although most of these algorithms are either inefficient or not exact, they do allow for approximate inferences to be computed within reasonable time. Of course, as the approximation improves, the required time typically becomes less reasonable. Despite the inherent complexity of inference in credal networks under strong independence, these approximate algorithms do yield results—be it often inner and therefore unsafe approximations—and thereby allow for applications to be tackled—be it inexactly. No such approximate algorithms are available for credal networks under epistemic irrelevance. For this reason, so far, applications of credal networks under epistemic irrelevance have been restricted to networks with a tree topology—such as hidden Markov models [6, 30].

We believe that the availability of approximate algorithms for credal networks under strong independence is not related to an inherent tractability of these networks, but rather to the fact that they can be defined in terms of sets of Bayesian networks, as discussed in Section 5.5_{145}. This has inspired researchers to try and adapt existing algorithms for Bayesian networks and we believe that this is the main reason why there are so many—approximate—algorithms for credal networks under strong independence.

The goal of this chapter is to show that the algorithmic possibilities of credal networks under epistemic irrelevance go far beyond the special cases that have been considered so far. We will explore some of these possibilities and will provide tools and ideas that can be used for further exploration. First of all, as we will see, efficient exact inference in credal networks under epistemic irrelevance is neither limited to inferences with only a single query node—and multiple evidence nodes—nor to credal networks with a tree topology. Credal networks with more complicated topologies allow for efficient inference as well, and it is possible to consider multiple query variables at once. We illustrate this by developing algorithms that can deal with more general topologies, for broad classes of inference problems that significantly extend the case of a single query node with multiple evidence nodes. Furthermore, and perhaps most importantly, we introduce algorithmic tools and ideas that should allow practitioners to move beyond the specific inference problems and topologies that are tackled by our algorithms and to develop their own algorithms for solving the particular inference problems that are relevant for their applications. We illustrate these techniques by means of a number of examples.

7.1 WHICH KINDS OF INFERENCES DO WE CONSIDER?

The most general inference problem that we will consider in this chapter is the numerical evaluation of $\underline{P}^{\mathrm{irr}}_{G}(f\rfloor O)$ or $\underline{R}^{\mathrm{irr}}_{G}(f\rfloor O)$—see Equation $(6.15)_{180}$—for

some arbitrary $O \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$ and $f \in \mathscr{G}(O)$. This corresponds to conditioning by means of natural extension and regular extension, respectively.[1] As argued extensively in Chapter 3_{69}, it makes sense to interpret $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor O)$ and $\underline{R}^{\mathrm{irr}}_G(\cdot\rfloor O)$ as updated uncertainty models that become applicable after O has been observed. If f is the indicator $\mathbb{I}_B$ of some event $B \subseteq O$, we obtain the updated lower and upper probabilities of B as special cases: they are given by $\underline{P}^{\mathrm{irr}}_G(B\rfloor O) := \underline{P}^{\mathrm{irr}}_G(\mathbb{I}_B\rfloor O)$ and $\overline{P}^{\mathrm{irr}}_G(B\rfloor O) := -\underline{P}^{\mathrm{irr}}_G(-\mathbb{I}_B\rfloor O)$ for natural extension, and similarly for regular extension.

The fact that we focus on the framework of lower previsions here is not restrictive. Due to the connections that we have established between the different notions of irrelevant natural extension that were introduced in Chapter 5_{128},[2] the value of $\underline{P}^{\mathrm{irr}}_G(f\rfloor O)$ or $\underline{R}^{\mathrm{irr}}_G(f\rfloor O)$ can also be interpreted in terms of other frameworks. For example, if we start from local credal sets $\mathscr{F}_{s\rfloor x_{P(s)}}$ and let $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ be the irrelevant natural extension of the network that has the lower previsions $\underline{P}_{s\rfloor x_{P(s)}} := \underline{P}_{\mathscr{F}_{s\rfloor x_{P(s)}}}$ as its local models, then

$$\underline{P}^{\mathrm{irr}}_G(f\rfloor O) = \min\left\{\sum_{x_G\in B} f(x_G)p_G(x_G\rfloor O)\colon\, p_G(\cdot\rfloor\cdot) \in \mathscr{F}^{\mathrm{irr}}_G\right\}. \tag{7.1}$$

If $\overline{P}^{\mathrm{irr}}_G(O) = 0$, then $\underline{R}^{\mathrm{irr}}_G(f\rfloor O) = \underline{P}^{\mathrm{irr}}_G(f\rfloor O)$. Otherwise, it follows from Equation $(3.11)_{91}$ and Proposition 49_{143} that

$$\underline{R}^{\mathrm{irr}}_G(f\rfloor O) = \inf\left\{\sum_{x_G\in B} f(x_G)p_G(x_G\rfloor O)\colon\, p_G(\cdot\rfloor\cdot) \in \mathscr{F}^{\mathrm{irr}}_G,\, p_G(O) > 0\right\}. \tag{7.2}$$

Similar interpretations for $\underline{P}^{\mathrm{irr}}_G(f\rfloor O)$ and $\underline{R}^{\mathrm{irr}}_G(f\rfloor O)$ can be obtained in terms of sets of conditional linear previsions as well; it suffices to replace $\mathscr{F}_{s\rfloor x_{P(s)}}$, $\mathscr{F}^{\mathrm{irr}}_G$, $p_G(\cdot\rfloor\cdot)$ and $\sum_{x_G\in B} f(x_G)p_G(x_G\rfloor O)$ by $\mathscr{M}_{s\rfloor x_{P(s)}}$, $\mathscr{M}^{\mathrm{irr}}_G$ and $P_G(\cdot\rfloor\cdot)$ and $P_G(f\rfloor O)$, respectively.

The situation is slightly different for sets of desirable gambles. If we start from arbitrary local sets of desirable gambles $\mathscr{D}_{s\rfloor x_{P(s)}}$ and let $\underline{P}_{s\rfloor x_{P(s)}}$ be the corresponding lower previsions, then, for any $O \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$ such that $\underline{P}^{\mathrm{irr}}_G(O) > 0$, $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor O)$ is equal to $\underline{P}_{\mathscr{D}^{\mathrm{irr}}_G}(\cdot\rfloor O)$.[3] However, in general, $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ is only known to be dominated by $\underline{P}_{\mathscr{D}^{\mathrm{irr}}_G}(\cdot\rfloor\cdot)$[4] and will not always be equal to it. If they differ, then $\underline{P}^{\mathrm{irr}}_G(f\rfloor O)$ serves as a conservative—safe, lower—approximation of $\underline{P}_{\mathscr{D}^{\mathrm{irr}}_G}(f\rfloor O)$. Similarly, $\underline{R}^{\mathrm{irr}}_G(f\rfloor O)$ is only guaranteed to be equal to $\underline{P}_{\mathscr{D}^{\mathrm{r}}_O}(f)$—

[1] Since $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ is coherent, it coincides with its natural extension.

[2] See Theorems 42_{139} and 46_{142} and Proposition 49_{143}.

[3] For $O = \mathscr{X}_G$, this follows from Proposition 43_{140}. The other cases then follow from the GBR; see Section $2.7.2_{61}$.

[4] By the same argument that was used in the beginning of Theorem 42_{139}.

with $\mathscr{D}_O^{\mathrm{r}}$ as in Equation $(3.8)_{88}$, for $\mathscr{D} = \mathscr{D}_G^{\mathrm{irr}}$—if $\overline{P}_G^{\mathrm{irr}}(O) > 0$; otherwise, it might only be a conservative—safe, lower—approximation of $\underline{P}_{\mathscr{D}_O^{\mathrm{r}}}(f)$.

In any case, in practice, the local models that are elicited from experts or data are usually not sets of desirable gambles, but lower previsions or their corresponding credal sets. The role of the framework of sets of desirable gambles is then a theoretical and philosophical one. If we let $\mathscr{D}_{s\rfloor x_{P(s)}}$ be the unique smallest coherent set of desirable gambles that corresponds to $\underline{P}_{s\rfloor x_{P(s)}}$, as given by Equation $(2.4)_{41}$, then $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is the conditional lower prevision that corresponds to $\mathscr{D}_G^{\mathrm{irr}}$ [see Theorem 42_{139}] and, for all $O \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$, $\underline{R}_G^{\mathrm{irr}}(\cdot\rfloor O) = \underline{P}_{\mathscr{D}_O^{\mathrm{r}}}$ [see Equation $(3.9)_{88}$]. The advantage of this connection with the framework of sets of desirable gambles is (a) that it has allowed us to develop and prove the theoretical properties discussed in the previous chapter and (b) that it provides us with a justification for regarding $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor O)$ and $\underline{R}_G^{\mathrm{irr}}(\cdot\rfloor O)$ as updated models [see Chapter 3_{69}]. From a practical and computational point of view, we will mainly focus on the frameworks of lower previsions and credal sets. The choice is a matter of personal preference—we prefer lower previsions—and mathematical convenience—both frameworks have their merits—because, as can be seen from Equations $(7.1)_{\curvearrowleft}$ and $(7.2)_{\curvearrowleft}$, they lead to mathematically equivalent results. For now, we consider the framework of lower previsions.

The choice between $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ and $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$ is also a matter of personal preference; it depends on whether one is willing to make the additional assessments that are required to justify updating by means of regular extension; see Chapter 3_{69}. We consider these additional assessments to be reasonable and therefore prefer regular extension. From a practical point of view, regular extension also has the advantage that it produces more informative inferences, a feature that is of course important in applications.

Although we will be offering methods for computing $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ and $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$ for general events $O \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$ and gambles $f \in \mathscr{G}(O)$ [see for example Section 7.4_{221}], they can only be applied for small networks; for larger networks, these general methods are simply too demanding from a computational point of view. As is the case for Bayesian networks, computational efficiency can only be achieved for specific choices of O and f. We give a brief overview of the specific choices for which we will present efficient algorithms. A first important—and popular—example is the evaluation of $\underline{P}_q^{\mathrm{irr}}(f\rfloor x_E)$ and $\underline{R}_q^{\mathrm{irr}}(f\rfloor x_E)$, for some $q \in G$, $E \subseteq G\setminus\{q\}$, $f \in \mathscr{G}(\mathscr{X}_q)$ and $x_E \in \mathscr{X}_E$. In that case, X_q is referred to as the query variable and the variables that are represented by X_E—the variables X_e, with $e \in E$—are called the evidence variables. The idea is to use information about the evidence variables—in this case their value—to learn something about the query variable. Multiple query nodes can also be considered. This corresponds to considering disjoint subsets Q and E of G and evaluating $\underline{P}_Q^{\mathrm{irr}}(f\rfloor x_E)$ or $\underline{R}_Q^{\mathrm{irr}}(f\rfloor x_E)$, for some $f \in \mathscr{G}(\mathscr{X}_Q)$ and $x_E \in \mathscr{X}_E$. In that case, f will often be structured in some way. It could for example be the indicator $\mathbb{I}_{\{x_Q\}}$ of some value $x_Q \in \mathscr{X}_Q$—in order to compute the lower

probability of x_Q—a sum $\sum_{q \in Q} f_q$ of gambles $f_q \in \mathscr{G}(\mathscr{X}_q)$ or even a product $\prod_{q \in Q} f_q$—of which the indicator $\mathbb{I}_{\{x_Q\}}$ is a special case. We discuss these examples—as well as others—further on in this chapter [see Sections $7.5.3_{249}$–$7.5.6_{264}$]. Although it often makes things easier, it is not necessary for all of the evidence to be specified completely; partially specified evidence can be dealt with as well, in the sense that we condition on an event $B_E \in \mathscr{P}_\emptyset(\mathscr{X}_E)$ that is not of the form $B_E = \{x_E\}$. In order to formalise this, we can consider a third subset L of G, pairwise disjoint from Q and E, and let the evidence consist of a completely specified part $x_E \in \mathscr{X}_E$ and a—possibly—partially specified part $B_L \in \mathscr{P}_\emptyset(\mathscr{X}_L)$. The goal is then to evaluate $\underline{P}^{\mathrm{irr}}_Q(f \rfloor \{x_E\} \times B_L)$ or $\underline{R}^{\mathrm{irr}}_Q(f \rfloor \{x_E\} \times B_L)$, for some $f \in \mathscr{G}(\mathscr{X}_Q)$.[5]

7.2 MAKING THE PROBLEM SMALLER

Before actually trying to compute $\underline{P}^{\mathrm{irr}}_G(f \rfloor O)$ or $\underline{R}^{\mathrm{irr}}_G(f \rfloor O)$, the first step should be to see whether it is possible to simplify the specific inference problem at hand. As we know from Corollary 60_{160} and Theorem 73_{181}, it is sometimes possible to reduce an inference problem in the global network to a similar inference problem in a subnetwork. If the subnetwork is substantially smaller, this preprocessing step can result in significantly decreased computational costs, and can sometimes even turn an intractable problem into a tractable one.

A first important special case is the removal of so-called *barren nodes*. Consider a subset S of G, an event $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and a gamble $f \in \mathscr{G}(B_S)$. Consider a node $\ell_1 \in G \setminus S$ that is a leaf of the network and let $G_1 := G \setminus \{\ell_1\}$. Then G_1 is clearly an ancestral and therefore closed subset of G. Hence, we know from Corollary 60_{160} that $\underline{P}^{\mathrm{irr}}_{G_1}(\cdot \rfloor \cdot)$ is not only the marginalisation of $\underline{P}^{\mathrm{irr}}_G(\cdot \rfloor \cdot)$ to G_1, it is also the irrelevant natural extension of a credal network that has the sub-DAG that corresponds to G_1 as its graphical structure and that has the same local models as the original model—except for the local models of the node ℓ_1, because that node has been removed. In fact, for this reason, our notation does not even distinguish between these two interpretations of $\underline{P}^{\mathrm{irr}}_{G_1}(\cdot \rfloor \cdot)$.[6] Consequently, as far as the computation of $\underline{P}^{\mathrm{irr}}_S(f \rfloor B_S)$ is concerned, we do not need to consider the original network but can simply work with the subnetwork that corresponds to G_1. If there is a node $\ell_2 \in G_1 \setminus S$ that

[5] All these inferences assume that missing evidence is 'Missing At Random' (MAR), which more or less means that it is not missing in some biased way. This requirement is closely related to the honesty that was referred to in Section $3.1.2_{71}$. In this case, in the example of the smoker, the issue is that a smoker might choose not to answer a question about his smoking habits—instead of lying about it. References [49, 117] provide more information about MAR and what can be done if this requirement is not fulfilled.

[6] Our notation does distinguish between them in the other frameworks. For example, in terms of sets of full conditional probability mass functions, we find that $\mathrm{marg}^{\mathrm{c}}_{G_1}(\mathscr{F}^{\mathrm{irr}}_G) = \mathscr{F}^{\mathrm{irr}}_{G_1}$.

Figure 7.1: Example of a Markov chain

is a leaf of the sub-DAG that corresponds to G_1, we can remove ℓ_2 as well and consider the sub-network that corresponds to $G_2 := G_1 \setminus \{\ell_2\}$, which is again an ancestral set, and so on. The nodes that can be removed in this way are called barren nodes. Clearly, a node is barren if it does not belong to S and either has no children—is a leaf—or has only barren children. Any ancestral subset K of G can be obtained by removing barren nodes. If S is a subset of such an ancestral set K, then $\underline{P}^{\mathrm{irr}}_S(f\rfloor B_S)$ can be computed using only the subnetwork that corresponds to K. Similar techniques can be applied to simplify the computation of the regular extension $\underline{R}^{\mathrm{irr}}_S(f\rfloor B_S)$: it follows from Theorem 73_{181} that $\underline{R}^{\mathrm{irr}}_K(\cdot\rfloor\cdot)$ is, besides the marginalisation of $\underline{R}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$, also the regular extension of $\underline{P}^{\mathrm{irr}}_K(\cdot\rfloor\cdot)$.[7]

Example 8. *Consider a simple credal network under epistemic irrelevance that has the DAG in Figure 7.1 as its graphical structure; this is called an imprecise Markov chain [43]. Let $S := \{s_2, s_3, s_4, s_5\}$ and consider an event $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and a gamble $f \in \mathscr{G}(B_S)$. Then as far as the computation of $\underline{P}^{\mathrm{irr}}_S(f\rfloor B_S)$ or $\underline{R}^{\mathrm{irr}}_S(f\rfloor B_S)$ is concerned, the nodes s_6 and s_7 are barren and can therefore be removed from the network prior to performing any computations. Similarly, if the graphical structure of the network is the DAG in Figure 5.1_{130}, then for the same inferences, the nodes s_6, s_7, s_8, s_9 and s_{10} are barren and can therefore be removed beforehand.*

Removing barren nodes can be automated easily. It suffices to look for a leaf that does not belong to S and to remove it from the network. By repeating this process until we obtain a DAG whose leaves all belong to S, all barren nodes will eventually be removed.

Another important special case is the removal of AD-separated evidence. Let I, C and S be three pairwise disjoint subsets of G and let $B_I \in \mathscr{P}_\emptyset(\mathscr{X}_I)$, $x_C \in \mathscr{X}_C$, $B_S \in \mathscr{P}_\emptyset(\mathscr{X}_S)$ and $f \in \mathscr{G}(\mathscr{X}_S)$. Then if $\mathrm{AD}(I, S \mid C)$, we know from Corollary 66_{168} that $\underline{P}^{\mathrm{irr}}_S(f\rfloor B_S \times \{x_C\} \times B_I) = \underline{P}^{\mathrm{irr}}_S(f\rfloor B_S \times \{x_C\})$. Although removing AD-separated evidence in this way might already lead to computational savings, we can go a lot further. By actually removing the AD-separated evidence nodes rather than just the evidence itself, we can reduce the problem to an equivalent problem in a smaller network. Due to Theorem 64_{167}, we know that there is a closed subset K of G such that $S \subseteq K$,

[7] It suffices to apply Theorem 73_{181} with $B_{N(K)} = \mathscr{X}_{N(K)}$. Since $P(K) = \emptyset$, we find that $\overline{P}^{\mathrm{irr}}_{PN(K)}(\{x_{P(K)}\} \times \mathscr{X}_{N(K)}) = \overline{P}^{\mathrm{irr}}_{N(K)}(\mathscr{X}_{N(K)}) = 1 > 0$.

$P(K) \subseteq C$, $I \subseteq N(K)$ and $D(K) \cap C = \emptyset$. It then follows from Corollary 60_{160} that $\underline{P}^{\text{irr}}_K(\cdot\rfloor \cdot \times \{x_{C\setminus K}\} \times B_I) = \underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$ and therefore, we find that

$$\underline{P}^{\text{irr}}_S(f\rfloor B_S \times \{x_C\} \times B_I) = \underline{P}^{\text{irr}}_{S\rfloor x_{P(K)}}(f\rfloor B_S \times \{x_{C\cap K}\}), \tag{7.3}$$

where $\underline{P}^{\text{irr}}_{S\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$ is obtained by applying Equation $(4.13)_{110}$ to the irrelevant natural extension $\underline{P}^{\text{irr}}_{K\rfloor x_{P(K)}}(\cdot\rfloor\cdot)$ of the subnetwork that corresponds to K and $x_{P(K)}$. Since it corresponds to a smaller network, the right-hand side of Equation (7.3) will be easier to compute than the left-hand side. Obviously, if there are multiple closed sets K such that $S \subseteq K$, $P(K) \subseteq C$, $I \subseteq N(K)$ and $D(K) \cap C = \emptyset$ it is best to choose the smallest one. Similar techniques can be used to simplify the computation of the regular extension $\underline{R}^{\text{irr}}_S(f\rfloor B_S \times \{x_C\} \times B_I)$. It follows from Theorem 73_{181} that it is equal to $\underline{R}^{\text{irr}}_{S\rfloor x_{P(K)}}(f\rfloor B_S \times \{x_{C\cap K}\})$ or $\underline{P}^{\text{irr}}_{S\rfloor x_{P(K)}}(f\rfloor B_S \times \{x_{C\cap K}\})$, depending on the sign of the upper probability $\overline{P}^{\text{irr}}_{PN(K)}(\{x_{C\setminus K}\} \times B_{N(K)})$.

In general, it does not seem feasible to have a computer remove AD-separated evidence (nodes) in an automated way, as it would require some method for detecting that the conditioning event O in $\underline{P}^{\text{irr}}_G(f\rfloor O)$ is of the form $B_S \times \{x_C\} \times B_I$, with $\text{AD}(I, S \mid C)$. This seems hard. Nevertheless, of course, for individual inference problems, it might still be easy for a practitioner to detect that O is of this form, in which case he can apply the simplification we suggest.

However, there is one particular—but very popular—case where it does seem feasible to remove AD-separated evidence nodes in an automated way, while at the same time also removing barren nodes. This is the case where f is a gamble on $\mathscr{X}_Q$ and the evidence O is of the form $\{x_E\}$, for some $x_E \in \mathscr{X}_E$, with Q and E disjoint subsets of G, and we wish to compute $\underline{P}^{\text{irr}}_Q(f\rfloor x_E)$ or $\underline{R}^{\text{irr}}_Q(f\rfloor x_E)$. Removing barren nodes and removing AD-separated nodes then amounts to the same thing. In both cases, we end up with a closed subset K of G such that $Q \subseteq K$, $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$. Furthermore, as the following result establishes, there is always a unique smallest such set.

Proposition 83. *Consider two disjoint subsets Q and E of G. Then there will always be a unique smallest closed subset K of G such that $Q \subseteq K$, $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$.*

Proof of Proposition 83. The set G itself is a trivial closed subset of G and, since $P(G) = \emptyset$ and $D(G) = \emptyset$, it also satisfies the other properties. Therefore, since the number of closed subsets of G is finite, it suffices to prove that if two closed sets K_1 and K_2 both satisfy these properties, then their intersection $K = K_1 \cap K_2$ is closed and satisfies them as well.

So consider two closed subsets K_i of G, with $i \in \{1,2\}$, such that $Q \subseteq K_i$, $P(K_i) \subseteq E$ and $D(K_i) \cap E = \emptyset$. Then $K := K_1 \cap K_2$ is clearly a closed subset of G and $Q \subseteq K$. We are left to prove that $P(K) \subseteq E$ and that $D(K) \cap E = \emptyset$.

First assume *ex absurdo* that $P(K)$ is not a subset of E. This means that there is some $s \in P(K)$ such that $s \notin E$, which implies that $s \notin P(K_1)$ and $s \notin P(K_2)$. Since $s \in P(K)$, we know that $s \notin K$ and that there is some $k \in K$ such that $s \in P(k)$. Since $k \in K \subseteq K_1$ and $s \notin P(K_1)$, this implies that $s \in K_1$. In the same way, we also find that $s \in K_2$. This implies that $s \in K$, a contradiction.

Next, assume *ex absurdo* that $D(K) \cap E = \emptyset$. This means that there is some $s \in D(K)$ such that $s \in E$, which implies that $s \notin D(K_1)$ and $s \notin D(K_2)$. Since $s \in D(K)$, there is some $k \in K$ such that $k \sqsubseteq s$ and $s \notin K$. Since $k \in K \subseteq K_1$ and $s \notin D(K_1)$, this implies that $s \in K_1$. In the same way, we find that $s \in K_2$. This implies that $s \in K$, a contradiction. □

Once we know this unique smallest set K—or any other K that satisfies these properties—we can compute $\underline{P}^{\text{irr}}_{Q}(f \rfloor x_E)$ and $\underline{R}^{\text{irr}}_{Q}(f \rfloor x_E)$ using only the subnetwork that corresponds to K and $x_{P(K)}$. $\underline{P}^{\text{irr}}_{Q}(f \rfloor x_E)$ is then equal to $\underline{P}^{\text{irr}}_{Q \rfloor x_{P(K)}}(f \rfloor x_{E \cap K})$ and $\underline{R}^{\text{irr}}_{Q}(f \rfloor x_E)$ is equal to $\underline{R}^{\text{irr}}_{Q \rfloor x_{P(K)}}(f \rfloor x_{E \cap K})$ or $\underline{P}^{\text{irr}}_{Q \rfloor x_{P(K)}}(f \rfloor x_{E \cap K})$, depending on whether or not $\overline{P}^{\text{irr}}_{PN(K)}(x_{E \setminus K})$ is strictly positive. Therefore, in this case, removing barren nodes and AD-separated evidence nodes can be done in a single step: all we have to do is find the unique smallest closed superset K of Q such that $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$. Although we will not pursue this path any further, it seems that it should be possible to construct this set K in an automated manner. Developing an algorithm that is able to do so would be useful in applications where many different inferences need to be computed, for different sets Q and E; we leave this as an interesting line of future research. For individual inferences, this set K can often be found manually, just by looking at the network. The following simple example illustrates this; more involved examples can be found in Section 7.6$_{271}$.

Example 9. *Consider the Markov chain in Figure 7.1$_{217}$ and let $Q = \{s_3, s_4\}$ and $E = \{s_2, s_5\}$. Then the smallest closed set K such that $Q \subseteq K$, $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$ is clearly $\{s_3, s_4, s_5\}$. For the DAG in Figure 5.1$_{130}$, with $Q = \{s_4, s_{10}\}$ and $E = \{s_3, s_6, s_7\}$, the smallest such set is $K = \{s_4, s_5, s_7, s_{10}\}$.*

Before applying any of the algorithms that will be introduced in the remainder of this chapter, these preprocessing steps should always be executed first. Our algorithms can then be used to solve the equivalent—but smaller-sized and therefore usually more tractable—inference problem in the subnetwork that corresponds to K and $x_{P(K)}$ instead of the original—less tractable—inference problem in the original network.

7.3 REDUCTION TO THE UNCONDITIONAL CASE

Another important trick that will help us render the computation of $\underline{P}^{\text{irr}}_{G}(f \rfloor O)$ tractable, even for large networks, is to focus on computing unconditional

lower previsions. As we will see further on in Sections 7.5_{235} and 7.6_{271}, the unconditional part of $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ can sometimes be evaluated by means of recursive techniques that are far more efficient than the brute force linear programming approaches that will be discussed in Section $7.4.3_{225}$. Furthermore, even when this is not the case, it will sometimes be possible to solve small subproblems with brute force methods and to combine the solutions of these subproblems recursively. In any case, for now, the main point is that—as is to be expected—$\underline{P}_G^{\mathrm{irr}}(\cdot)$ is easier to evaluate than $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$.

Once we are able to evaluate $\underline{P}_G^{\mathrm{irr}}(\cdot)$, we can use the techniques in Section $2.7.3_{63}$ to compute the natural and regular extensions of $\underline{P}_G^{\mathrm{irr}}(\cdot)$, given by

$$\underline{E}_G^{\mathrm{irr}*}(f\rfloor O) := \begin{cases} \max\{\mu \in \mathbb{R} \colon \underline{P}_G^{\mathrm{irr}}(\mathbb{I}_O[f-\mu]) \geq 0\} & \text{if } \underline{P}_G^{\mathrm{irr}}(O) > 0 \\ \min f & \text{otherwise} \end{cases}$$

and

$$\underline{R}_G^{\mathrm{irr}*}(f\rfloor O) := \begin{cases} \max\{\mu \in \mathbb{R} \colon \underline{P}_G^{\mathrm{irr}}(\mathbb{I}_O[f-\mu]) \geq 0\} & \text{if } \overline{P}_G^{\mathrm{irr}}(O) > 0 \\ \min f & \text{otherwise,} \end{cases}$$

respectively. In order to be able to apply the techniques in Section $2.7.3_{63}$, all we need is a method for evaluating the real-valued function $\rho_{f,O}^{\mathrm{irr}}$, defined by

$$\rho_{f,O}^{\mathrm{irr}}(\mu) := \underline{P}_G^{\mathrm{irr}}(\mathbb{I}_O[f-\mu]) \text{ for all } \mu \in \mathbb{R}.$$

As explained in Section $2.7.3_{63}$, $\max\{\mu \in \mathbb{R} \colon \underline{P}_G^{\mathrm{irr}}(\mathbb{I}_O[f-\mu]) \geq 0\}$ can then be obtained by a simple root-finding procedure such as the bisection method and $\underline{P}_G^{\mathrm{irr}}(O)$, and $\overline{P}_G^{\mathrm{irr}}(O)$ will be strictly positive if and only if this is the case for $\rho_{f,O}^{\mathrm{irr}}(\mu_0)$ and $\rho_{f,O}^{\mathrm{irr}}(\mu_1)$, respectively, with $\mu_0 < \min f$ and $\mu_1 > \max f$.

It is important to realise that $\underline{E}_G^{\mathrm{irr}*}(f\rfloor O)$ and $\underline{R}_G^{\mathrm{irr}*}(f\rfloor O)$ are not the actual inferences that we are looking for, which are $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ or $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$; see Section 7.1_{213}. In general, $\underline{E}_G^{\mathrm{irr}*}(f\rfloor O)$ and $\underline{R}_G^{\mathrm{irr}*}(f\rfloor O)$ only provide conservative—lower, safe—approximations. However, in many cases, these approximations will be exact. As explained in Section $3.2.3_{75}$, $\underline{E}_G^{\mathrm{irr}*}(f\rfloor O)$ is guaranteed to coincide with $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ whenever $\underline{P}_G^{\mathrm{irr}}(O) > 0$ and otherwise provides a vacuous lower approximation. Similarly, as we know from 6.8_{180}, $\underline{R}_G^{\mathrm{irr}*}(f\rfloor O)$ is guaranteed to coincide with $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$ whenever $\overline{P}_G^{\mathrm{irr}}(O) > 0$; otherwise, $\underline{R}_G^{\mathrm{irr}*}(f\rfloor O) = \min f$ provides a vacuous conservative lower bound for $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$. Since, from a practical point of view, updating on an observation that has upper probability zero will most likely rarely happen anyway, this means that, as far as updating by means of regular extension is concerned, being able to evaluate $\underline{P}_G^{\mathrm{irr}}(\cdot)$ will be sufficient in order to compute $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$ exactly in almost all cases.

7.4 BRUTE FORCE TECHNIQUES

After applying these preprocessing steps and tricks, we are still left with the same type of problem: the numerical evaluation of $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ or $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$, for some $O \in \mathscr{P}_{\emptyset}(\mathscr{X}_G)$ and $f \in \mathscr{G}(O)$. So far, all we have done is provide methods that can—in some cases—make G smaller or that can reduce the problem to the unconditional case, with $O = \mathscr{X}_G$. However, we still do not know how to evaluate $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ or $\underline{R}_G^{\mathrm{irr}}(f\rfloor O)$. In this section, we provide a number of brute force techniques for solving this problem, some of which we already published in Reference [27]. If the problem is small enough for these methods to remain tractable, these techniques can be applied to all networks and all types of inferences. For specific cases, more efficient recursive techniques will be developed in Sections 7.5$_{235}$ and 7.6$_{271}$.

7.4.1 *Regarding the model as a normal natural extension*

One way to compute $\underline{P}_G^{\mathrm{irr}}(f\rfloor O)$ is to regard it as a special case of the problem of computing the natural extension of a partially specified conditional lower prevision. Indeed, as we know from Proposition 45$_{141}$, the irrelevant natural extension $\underline{P}_G^{\mathrm{irr}}(\cdot\rfloor\cdot)$ is the 'normal' natural extension of a conditional lower prevision $\underline{P}_G(\cdot\rfloor\cdot)$ with domain

$$\mathscr{C} = \{(g, x_{PN(s)})\colon s \in G, g \in \mathscr{G}(\mathscr{X}_s), x_{PN(s)} \in \mathscr{X}_{PN(s)}\},$$

defined by

$$\underline{P}_s(g\rfloor x_{PN(s)}) := \underline{P}_{s\rfloor x_{P(s)}}(g) \text{ for all } s \in G,\ g \in \mathscr{G}(\mathscr{X}_s) \text{ and } x_{PN(s)} \in \mathscr{X}_{PN(s)}.$$

Since the domain of $\underline{P}_G(\cdot\rfloor\cdot)$ is infinite, computing its natural extension is generally infeasible. However, in practice, it will often not be necessary to consider an infinite domain because the local models $\underline{P}_{s\rfloor x_{P(s)}}$ are usually finitely generated.

Finitely generated models were introduced in Section 2.6.1$_{52}$ as closed and convex sets of linear previsions that have a finite number of extreme points. Similarly, as defined in Section 2.6.2$_{55}$, a finitely generated credal set is a closed and convex set of probability mass functions that has a finite number of vertices. In the framework of lower previsions, we say that $\underline{P}_{s\rfloor x_{P(s)}}$ is finitely generated if it is fully determined by its restriction to some finite domain $\mathscr{K}_{s\rfloor x_{P(s)}} \subseteq \mathscr{G}(\mathscr{X}_s)$, in the sense that $\underline{P}_{s\rfloor x_{P(s)}}$ is the unconditional part of the natural extension of its restriction to $\mathscr{K}_{s\rfloor x_{P(s)}}$. This terminology is consistent: it is well known that $\underline{P}_{s\rfloor x_{P(s)}}$ is finitely generated if and only if the corresponding credal set $\mathscr{F}_{\underline{P}_{s\rfloor x_{P(s)}}}$ is finitely generated, and similarly for $\mathscr{M}_{\underline{P}_{s\rfloor x_{P(s)}}}$.[8]

[8]This follows from the fact that a bounded subset of Euclidean space is the intersection of

Particular classes of finitely generated models are 2-monotone lower previsions [48],[9] credal sets that correspond to comparative probability assessments [73], ε-contaminated models [7], etcetera. Local models are typically finitely generated because they belong to one of these classes, or because they are based on a finite number of assessments. Furthermore, if the local models are not finitely generated, it is always possible to approximate them arbitrarily closely by a finitely generated one.

More generally, we may have that each of the local models $\underline{P}_{s\rfloor x_{P(s)}}$ is equal to the unconditional part of the natural extension of its restriction to some—not necessarily finite—subdomain $\mathscr{K}_{s\rfloor x_{P(s)}} \subseteq \mathscr{G}(\mathscr{X}_s)$. The irrelevant natural extension $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$ is then equal to the global 'normal' natural extension of a conditional lower prevision $\underline{P}'_G(\cdot\rfloor\cdot)$ with domain

$$\mathscr{C}' = \{(g, x_{PN(s)}) \colon s \in G, x_{PN(s)} \in \mathscr{X}_{PN(s)}, g \in \mathscr{K}_{s\rfloor x_{P(s)}}\},$$

defined by[10]

$$\underline{P}'_s(g\rfloor x_{PN(s)}) := \underline{P}_{s\rfloor x_{P(s)}}(g) \text{ for all } s \in G,\, x_{PN(s)} \in \mathscr{X}_{PN(s)} \text{ and } g \in \mathscr{K}_{s\rfloor x_{P(s)}}. \tag{7.4}$$

If the local models are finitely generated—if $\mathscr{K}_{s\rfloor x_{P(s)}}$ is finite—then $\mathscr{C}'$ will also be finite. In that case, $\underline{P}^{\text{irr}}_G(f\rfloor O)$ can be computed by means of any algorithm that is capable of computing the natural extension of a conditional lower prevision with a finite domain. The most efficient algorithm seems to be the one in Reference [98, Section 17.2.2], which extends the techniques in Reference [111] from probabilities to previsions. Basically, this algorithm consists of a sequence of linear programs; see References [98, 111] for more information.

The advantage of this approach is that it can be applied to compute any $\underline{P}^{\text{irr}}_G(f\rfloor O)$ of interest. However, for large networks, this computation will be intractable because the number of constraints in the linear programs that need to be solved is more or less equal to the cardinality of $\mathscr{C}'$, which is exponential in the size of the network. Also, this approach only works for $\underline{P}^{\text{irr}}_G(f\rfloor O)$. In order to compute $\underline{R}^{\text{irr}}_G(f\rfloor O)$, we need to apply the techniques in Section 7.4.1$_{\frown}$, which, basically, requires us to evaluate $\underline{P}^{\text{irr}}_G(\cdot)$ for a number of different gambles. In the next two sections, we focus on evaluating this unconditional part of $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$.

---

a finite number of closed half spaces if and only if it is the convex hull of a finite number of vertices [60, Theorem 3.1.3].

[9] This class includes, as a special case, the lower previsions that correspond to—are the natural extension of—belief functions [52, 87].

[10] Since $\underline{P}_{s\rfloor x_{P(s)}}$ is—the unconditional part of—the natural extension of its restriction to $\mathscr{K}_{s\rfloor x_{P(s)}}$, every conditional lower prevision that dominates $\underline{P}'_G(\cdot\rfloor\cdot)$ also dominates $\underline{P}_G(\cdot\rfloor\cdot)$. Therefore, since $\underline{P}'_G(\cdot\rfloor\cdot)$ coincides with $\underline{P}_G(\cdot\rfloor\cdot)$ on its domain $\mathscr{C}'$, they both have the same natural extension, which, due to Proposition 45_{141}, is equal to $\underline{P}^{\text{irr}}_G(\cdot\rfloor\cdot)$.

7.4.2 *Alternative characterisations for the unconditional part*

We already know from the previous section that $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ is the natural extension of a *conditional* lower prevision. Our next result establishes that its unconditional part $\underline{P}^{\mathrm{irr}}_G(\cdot)$ is the natural extension of an *unconditional* lower prevision.

Proposition 84. *For all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, consider some subdomain $\mathscr{K}_{s\rfloor x_{P(s)}} \subseteq \mathscr{G}(\mathscr{X}_s)$ such that the local lower prevision $\underline{P}_{s\rfloor x_{P(s)}}$ on $\mathscr{G}(\mathscr{X}_s)$ is equal to the unconditional part of the natural extension of its restriction to $\mathscr{K}_{s\rfloor x_{P(s)}}$.*[11] *Consider a lower prevision $\underline{P}_G$ that is equal to zero on its domain*

$$\mathscr{K} = \left\{\mathbb{I}_{\{x_{PN(s)}\}}[g - \underline{P}_{s\rfloor x_{P(s)}}(g)] \colon s \in G, x_{PN(s)} \in \mathscr{X}_{PN(s)}, g \in \mathscr{K}_{s\rfloor x_{P(s)}}\right\}$$

and let $\underline{E}_G$ be the unconditional part of the natural extension of $\underline{P}_G$. Then $\underline{P}^{\mathrm{irr}}_G(f) = \underline{E}_G(f)$ for all $f \in \mathscr{G}(\mathscr{X}_G)$.

Proof of Proposition 84. Fix any $f \in \mathscr{G}(\mathscr{X}_G)$. For all $s \in G$, $x_{PN(s)} \in \mathscr{X}_{PN(s)}$ and $g \in \mathscr{K}_{s\rfloor x_{P(s)}}$, we have that

$$0 = \underline{P}^{\mathrm{irr}}_G(\mathbb{I}_{\{x_{PN(s)}\}}[g - \underline{P}^{\mathrm{irr}}_s(g\rfloor x_{PN(s)})]) = \underline{P}^{\mathrm{irr}}_G(\mathbb{I}_{\{x_{PN(s)}\}}[g - \underline{P}_{s\rfloor x_{P(s)}}(g)]),$$

where the first equality follows from the coherence of $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ [C4$_{48}$] and the second one from Corollary 44$_{141}$. Hence, $\underline{P}^{\mathrm{irr}}_G(\cdot)$ is equal to zero on $\mathscr{K}$. Since $\underline{E}_G$ is the pointwise smallest coherent lower prevision on $\mathscr{G}(\mathscr{X}_G)$ for which this is the case, it follows that $\underline{P}^{\mathrm{irr}}_G(f) \geq \underline{E}_G(f)$.

As explained in Section 7.4.1$_{221}$, $\underline{P}^{\mathrm{irr}}_G(\cdot\rfloor\cdot)$ is the natural extension of the conditional lower prevision $\underline{P}'_G(\cdot\rfloor\cdot)$ on $\mathscr{C}'$ that is defined by Equation (7.4)$_{\frown}$. Fix any $\varepsilon \in \mathbb{R}_{>0}$. Then $f - \underline{P}^{\mathrm{irr}}_G(f) + \varepsilon \in \mathscr{E}_{\underline{P}'_G(\cdot\rfloor\cdot)}$ because of Equation (2.11)$_{44}$, coherence [C8$_{49}$] and Equations (2.5)$_{41}$ and (2.6)$_{41}$. By the definition of $\mathscr{E}_{\underline{P}'_G(\cdot\rfloor\cdot)}$, this implies that there are $m \in \mathbb{N}$ and, for all $i \in \{1,\dots,m\}$, $(g_i, B_i) \in \mathscr{C}'$ and $\lambda_i \in \mathbb{R}_{\geq 0}$ such that

$$f - \underline{P}^{\mathrm{irr}}_G(f) + \varepsilon \geq \sum_{i=1}^{m} \lambda_i \mathbb{I}_{B_i}[g_i - \underline{P}'_G(g_i\rfloor B_i)] = \sum_{i=1}^{m} \lambda_i f_i, \tag{7.5}$$

where, for all $i \in \{1,\dots,m\}$, due to Equation (7.4)$_{\frown}$ and the definition of $\mathscr{C}'$, $f_i := \mathbb{I}_{B_i}[g_i - \underline{P}'_G(g_i\rfloor B_i)]$ is an element of $\mathscr{K}$. This implies that

$$\underline{E}_G(f) \geq \underline{E}_G\Big(\underline{P}^{\mathrm{irr}}_G(f) - \varepsilon + \sum_{i=1}^{m} \lambda_i f_i\Big) \geq \underline{P}^{\mathrm{irr}}_G(f) - \varepsilon + \sum_{i=1}^{m} \lambda_i \underline{E}_G(f_i) \geq \underline{P}^{\mathrm{irr}}_G(f) - \varepsilon,$$

[11] This is always possible; $\mathscr{K}_{s\rfloor x_{P(s)}} = \mathscr{G}(\mathscr{X}_s)$ is a trivial choice.

where the first inequality follows from coherence [C7$_{49}$] and Equation (7.5)$_{\frown}$ and the second one from coherence alone [C2$_{48}$, C3$_{48}$ and C8$_{49}$]. The third inequality holds because $\underline{E}_G$ dominates $\underline{P}_G$ on $\mathscr{K}$ [Equation (2.12)$_{44}$] and because $\underline{P}_G$ is equal to zero on $\mathscr{K}$. Since this holds for all $\varepsilon \in \mathbb{R}_{>0}$, we find that $\underline{E}_G(f) \geq \underline{P}_G^{\mathrm{irr}}(f)$. □

This result allows us to evaluate $\underline{P}_G^{\mathrm{irr}}(\cdot)$ by means of a general purpose algorithm for calculating the natural extension of an unconditional lower prevision, which typically involves solving a single linear program; we will construct the dual form of this linear program explicitly in the next section. Of course, from a practical point of view, this approach is only useful if the local models are finitely generated—if $\mathscr{K}_{s\rfloor x_{P(s)}}$ consists of a finite number of gambles—and even then, it will only be tractable for networks that are sufficiently small.

The following rather immediate consequence of Proposition 84$_{\frown}$ provides a simple characterisation of the linear previsions that dominate $\underline{P}_G^{\mathrm{irr}}(\cdot)$.

Corollary 85. *For all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, consider some subdomain $\mathscr{K}_{s\rfloor x_{P(s)}} \subseteq \mathscr{G}(\mathscr{X}_s)$ such that the local lower prevision $\underline{P}_{s\rfloor x_{P(s)}}$ on $\mathscr{G}(\mathscr{X}_s)$ is equal to the unconditional part of the natural extension of its restriction to $\mathscr{K}_{s\rfloor x_{P(s)}}$. A linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$ then belongs to $\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot)}$ if and only if for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$:*

$$P_G(\mathbb{I}_{\{x_{PN(s)}\}}[g - \underline{P}_{s\rfloor x_{P(s)}}(g)]) \geq 0 \text{ for all } g \in \mathscr{K}_{s\rfloor x_{P(s)}} \tag{7.6}$$

Proof of Corollary 85. We know from Proposition 84$_{\frown}$ that $\underline{P}_G^{\mathrm{irr}}(\cdot)$ is the pointwise smallest coherent lower prevision on $\mathscr{G}(\mathscr{X}_G)$ that satisfies Equation (7.6)—that dominates zero on $\mathscr{K}$, with $\mathscr{K}$ as in Proposition 84$_{\frown}$. This implies that a linear prevision P_G on $\mathscr{G}(\mathscr{X}_G)$ satisfies Equation (7.6) if and only if it is an element of $\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}(\cdot)}$—if and only if it dominates $\underline{P}_G^{\mathrm{irr}}(\cdot)$. □

Due to the one-to-one correspondence between linear previsions and probability mass functions, this result leads to the following intuitive characterisation of $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$—the unconditional part of $\mathscr{F}_G^{\mathrm{irr}}$.

Corollary 86. *A probability mass function p_G on $\mathscr{X}_G$ belongs to $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ if and only if, for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$:*

$$p_{PN(s)}(x_{PN(s)}) = 0 \text{ or } p_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{F}_{s\rfloor x_{P(s)}}, \tag{7.7}$$

where, when $p_{PN(s)}(x_{PN(s)}) \neq 0$,

$$p_s(z_s\rfloor x_{PN(s)}) := \frac{p_{\{s\}\cup PN(s)}(z_s, x_{PN(s)})}{p_{PN(s)}(x_{PN(s)})} \text{ for all } z_s \in \mathscr{X}_s.$$

Proof of Corollary 86$_{\curvearrowleft}$. For all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, let $\underline{P}_{s\rfloor x_{P(s)}}$ be the unique coherent lower prevision that corresponds to $\mathscr{F}_{s\rfloor x_{P(s)}}$. Consider any probability mass function p_G on $\mathscr{X}_G$. It then follows from Theorem 46$_{142}$ and Proposition 49$_{143}$ that p_G belongs to $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ if and only if the corresponding linear prevision P_G is an element of $\mathscr{M}_{\underline{P}_G^{\mathrm{irr}}()}$. By combining this with Corollary 85$_{\curvearrowleft}$, with $\mathscr{K}_{s\rfloor x_{P(s)}} = \mathscr{G}(\mathscr{X}_s)$, we find that p_G belongs to $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ if and only if, for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$:

$$\sum_{z_G \in \mathscr{X}_G} \mathbb{I}_{\{x_{PN(s)}\}}(z_{PN(s)})[g(z_s) - \underline{P}_{s\rfloor x_{P(s)}}(g)]p_G(z_G) \geq 0 \text{ for all } g \in \mathscr{G}(\mathscr{X}_s). \tag{7.8}$$

If $p_{PN(s)}(x_{PN(s)}) = 0$, then $p_G(z_G) = 0$ for all $z_G \in \mathscr{X}_G$ such that $z_{PN(s)} = x_{PN(s)}$ [because $p_{PN(s)}(x_{PN(s)}) \geq p_G(z_G) \geq 0$], which implies that Equation (7.8) is trivially satisfied. If $p_{PN(s)}(x_{PN(s)}) > 0$, then Equation (7.8) is equivalent to

$$\sum_{z_s \in \mathscr{X}_s} g(z_s) \frac{p_{\{s\}\cup PN(s)}(z_s, x_{PN(s)})}{p_{PN(s)}(x_{PN(s)})} \geq \underline{P}_{s\rfloor x_{P(s)}}(g) \text{ for all } g \in \mathscr{G}(\mathscr{X}_s),$$

which, since $p_s(\cdot\rfloor x_{PN(s)})$ is a probability mass function on $\mathscr{X}_s$, is true if and only if $p_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{F}_{s\rfloor x_{P(s)}}$. □

As we will see in the next section, this characterisation can be used to obtain a description of $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ in terms of linear constraints.

7.4.3 *Reducing the problem to linear programming*

For readers who are not familiar with the notion of natural extension, it will probably not be clear why the fact that the irrelevant natural extension is a specific kind of 'normal' natural extension implies that computing $\underline{P}_G^{\mathrm{irr}}(\cdot)$ basically comes down to solving a linear program. Therefore, and because we think that this link with linear programming is important, we will now explicitly construct the linear program that needs to be solved.

The starting point are the local models. For every $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, we have a coherent lower prevision $\underline{P}_{s\rfloor x_{P(s)}}$ on $\mathscr{G}(\mathscr{X}_s)$ and a corresponding credal set $\mathscr{F}_{s\rfloor x_{P(s)}}$. It then follows from Equation (7.1)$_{214}$ that the unconditional parts $\underline{P}_G^{\mathrm{irr}}(\cdot)$ and $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ of the respective irrelevant natural extensions are related in the following way. For all $f \in \mathscr{G}(\mathscr{X}_G)$, we have that

$$\underline{P}_G^{\mathrm{irr}}(f) = \min\left\{\sum_{x_G \in \mathscr{X}_G} f(x_G)p_G(x_G) \colon p_G \in \mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G\right\}. \tag{7.9}$$

Computing $\underline{P}_G^{\mathrm{irr}}(f)$ is therefore a matter of minimising the linear function $\sum_{x_G \in \mathscr{X}_G} f(x_G)p_G(x_G)$ over the set of vectors $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$. Hence, it follows that if we can characterise $\mathscr{F}_G^{\mathrm{irr}}\rfloor\mathscr{X}_G$ as the solution set of some collection of linear

constraints, then computing $\underline{P}_G^{\text{irr}}(f)$ comes down to solving a linear program. This means that, provided that the number of variables—$|G|$—and constraints is small enough, this problem can be tackled by one of the many available algorithms that have been specifically designed for this task; see for example Reference [9]. In the remainder of this section, we show that it is indeed possible to construct a set of linear constraints that is able to fully characterise the elements of $\mathscr{F}_G^{\text{irr}} \rfloor \mathscr{X}_G$.

In order to derive such a representation for the joint model $\mathscr{F}_G^{\text{irr}} \rfloor \mathscr{X}_G$, we start from similar representations for the local models. For all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, we characterise the local credal set $\mathscr{F}_{s \rfloor x_{P(s)}}$ as the set of all real-valued functions $p_{s \rfloor x_{P(s)}}$ on $\mathscr{X}_s$ that satisfy the unitary constraint

$$\sum_{z_s \in \mathscr{X}_s} p_{s \rfloor x_{P(s)}}(z_s) = 1 \tag{7.10}$$

and a—possibly infinite—set of linear *homogeneous* inequalities

$$\sum_{z_s \in \mathscr{X}_s} p_{s \rfloor x_{P(s)}}(z_s)\gamma(z_s) \geq 0, \tag{7.11}$$

where γ takes values in a—possibly infinite—set $\Gamma_{s \rfloor x_{P(s)}}$ of gambles on $\mathscr{X}_s$. Such a description for $\mathscr{F}_{s \rfloor x_{P(s)}}$ always exists because it can be derived from the corresponding coherent lower prevision $\underline{P}_{s \rfloor x_{P(s)}}$ by letting

$$\Gamma_{s, x_{P(s)}} = \left\{ g - \underline{P}_{s \rfloor x_{P(s)}}(g) : g \in \mathscr{G}(\mathscr{X}_s) \right\}. \tag{7.12}$$

Indeed, for this particular choice of $\Gamma_{s \rfloor x_{P(s)}}$, the combination of Equations (7.10) and (7.11) will always be equivalent to requiring that $p_{s \rfloor x_{P(s)}}$ should be a probability mass function on $\mathscr{X}_s$ for which the corresponding linear prevision dominates $\underline{P}_{s \rfloor x_{P(s)}}$ or, equivalently, that $p_{s \rfloor x_{P(s)}}$ should be an element of $\mathscr{F}_{s \rfloor x_{P(s)}}$. To understand why this is so, the starting point is to notice that if $\gamma = g - \underline{P}_{s \rfloor x_{P(s)}}(g)$, with $g \in \mathscr{G}(\mathscr{X}_s)$, then due to Equation (7.10), Equation (7.11) becomes equivalent to

$$\sum_{z_s \in \mathscr{X}_s} p_{s \rfloor x_{P(s)}}(z_s) g(z_s) \geq \underline{P}_{s \rfloor x_{P(s)}}(g). \tag{7.13}$$

Due to the coherence [C1$_{48}$] of $\underline{P}_{s \rfloor x_{P(s)}}$, this implies, for all $z_s \in \mathscr{X}_s$, that $p_{s \rfloor x_{P(s)}}(z_s) = \underline{P}_{s \rfloor x_{P(s)}}(\mathbb{I}_{\{z_s\}}) \geq 0$. By combining this with Equation (7.10), it follows that $p_{s \rfloor x_{P(s)}}$ is a probability mass function on $\mathscr{X}_s$. Since Equation (7.13) imposes that the corresponding linear prevision dominates $\underline{P}_{s \rfloor x_{P(s)}}$, this establishes the equivalence.

Equation (7.12) produces an infinite set of constraints that is guaranteed to characterise $\mathscr{F}_{s \rfloor x_{P(s)}}$. However, in practice, most of these constraints will

often be redundant. This is especially so for finitely generated local models [see Section 7.4.1$_{221}$]. Since these are fully determined by their value on some finite domain $\mathscr{K}_{s\rfloor x_{P(s)}}$, one can easily construct a set $\Gamma_{s,x_{P(s)}}$ that contains only a finite number of gambles and yet fully characterises $\mathscr{F}_{s\rfloor x_{P(s)}}$ by means of Equations (7.10)$_{\frown}$ and (7.11)$_{\frown}$. For example, by an argument that is similar to the one above, it follows that

$$\Gamma_{s,x_{P(s)}} = \left\{ g - \underline{P}_{s\rfloor x_{P(s)}}(g) : g \in \mathscr{K}_{s\rfloor x_{P(s)}} \text{ or } (\exists x_s \in \mathscr{X}_s)\, g = \mathbb{I}_{\{x_{x_s}\}} \right\}$$

satisfies this property. It might also be possible to consider even smaller sets, as some of the indicators $\mathbb{I}_{\{x_s\}}$ may also be redundant.

The importance of these local representations in terms of linear constraints—regardless of whether $\Gamma_{s,x_{P(s)}}$ is finite or not—is that we can use the local constraints to derive global ones. This results in the following characterisation of $\mathscr{F}_G^{\text{irr}}\rfloor\mathscr{X}_G$ in terms of linear constraints. Due to the homogeneity of the local constraints—except for the unitary constraint—the global constraints in Equation (7.14) are also homogeneous, thereby making them especially elegant.

Corollary 87. *For every $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, consider some subset $\Gamma_{s\rfloor x_{P(s)}}$ of $\mathscr{G}(\mathscr{X}_s)$ such that the local credal set $\mathscr{F}_{s\rfloor x_{P(s)}}$ is fully characterised by Equations* (7.10)$_{\frown}$ *and* (7.11)$_{\frown}$. *A probability mass function p_G on $\mathscr{X}_G$ then belongs to $\mathscr{F}_G^{\text{irr}}\rfloor\mathscr{X}_G$ if and only if*

$$\left(\forall \gamma \in \Gamma_{s\rfloor x_{P(s)}}\right) \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p_G(x_{PN(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0 \tag{7.14}$$

for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$.

Proof of Corollary 87. Consider any probability mass function p_G on $\mathscr{X}_G$ and fix some $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$. Due to Corollary 86$_{224}$, it suffices to prove that Equations (7.7)$_{224}$ and (7.14) are equivalent.

First assume that Equation (7.7)$_{224}$ holds. If $p_{PN(s)}(x_{PN(s)}) = 0$, Equation (7.14) is trivially true. If $p_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{F}_{s\rfloor x_{P(s)}}$, Equation (7.14) follows from the fact that $\mathscr{F}_{s\rfloor x_{P(s)}}$ satisfies Equation (7.11)$_{\frown}$.

Next, assume that Equation (7.14) holds. If $p_{PN(s)}(x_{PN(s)}) \neq 0$, dividing both sides of the inequality in Equation (7.14) by $p_{PN(s)}(x_{PN(s)})$ tells us that $p_s(\cdot\rfloor x_{PN(s)})$ satisfies Equation (7.11)$_{\frown}$. Since $p_s(\cdot\rfloor x_{PN(s)})$ is a probability mass function—because p_G is one—we also know that it satisfies Equation (7.10). Since $\mathscr{F}_{s\rfloor x_{P(s)}}$ is fully characterised by Equations (7.10)$_{\frown}$ and (7.11)$_{\frown}$, this allows us to infer that $p_s(\cdot\rfloor x_{PN(s)}) \in \mathscr{F}_{s\rfloor x_{P(s)}}$. □

When all lower probabilities are strictly positive, this result is fairly straightforward. The global inequalities can then be obtained by imposing all irrelevancies through element-wise Bayes's rule and clearing the denominators, as

is done in Reference [16, Section 8.3]. The importance of Corollary $87_{\frown}$ is that it shows that these inequalities remain valid *even if* lower (and upper) probabilities are allowed to be zero.

Reference [16] does not explicitly impose that p_G should be a probability mass function. The author seems to assume that it suffices to impose only the unitary constraint $\sum_{z_G \in \mathscr{X}_G} p(z_G) = 1$; the requirement that $p(z_G) \geq 0$ for all $z_G \in \mathscr{X}_G$ is regarded as redundant. Although we agree with this statement, we do not consider it to be trivial. The following theorem formalises this property. The proof can be found in Appendix $7.\mathrm{A}_{278}$.

Theorem 88. *For every $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, consider some subset $\Gamma_{s\rfloor x_{P(s)}}$ of $\mathscr{G}(\mathscr{X}_s)$ such that the local credal set $\mathscr{F}_{s\rfloor x_{P(s)}}$ is fully characterised by Equations $(7.10)_{226}$ and $(7.11)_{226}$. $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ then consists of those real-valued functions p_G on $\mathscr{X}_G$ for which $\sum_{z_G \in \mathscr{X}_G} p_G(z_G) = 1$ and, for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$:*

$$\left(\forall \gamma \in \Gamma_{s\rfloor x_{P(s)}}\right) \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p_G(x_{PN(s)}, z_s, z_{D(s)}) \gamma(z_s) \geq 0.$$

Corollary $87_{\frown}$ and Theorem 88 are valid regardless of whether or not $\Gamma_{s\rfloor x_{P(s)}}$ is finite. However, in the infinite case, their value is mainly of a theoretical nature. They can only be used in practice—at least in an exact way—if $\Gamma_{s\rfloor x_{P(s)}}$ is finite for all $s \in G$ and $x_{P(s)} \in \mathscr{X}_{P(s)}$, or equivalently, if all local credal sets are finitely generated.[12] In that case, Corollary $87_{\frown}$ and Theorem 88 characterise $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ as the solution set of a finite number of constraints and therefore, as explained in the beginning of this section, allow us to reformulate the computation of $\underline{P}_G^{\mathrm{irr}}(f)$ as a linear programming problem. Although the size of this linear program is exponential in the number of variables of the network, this approach allows for inference problems in small networks to be solved exactly. For large networks, this brute force approach will not be tractable. However, as we will show in Section 7.5_{235}, there are classes of networks for which specific types of inferences can be computed more intelligently, resulting in algorithms that remain tractable even for large networks.

By considering a network that consists of disconnected nodes only, the results in this—and the previous—section trivially lead to analogous statements for the independent natural extension, because, as we know from Section 6.6_{171}, the independent natural extension is equal to the unconditional part of the irrelevant natural extension of such a network. For example, for all

[12] If we allow for non-linear constraints, then local credal sets that are not finitely generated could be practical as well, as they can often be described by means of a finite set of non-linear constraints. We believe that Corollary $87_{\frown}$ and Theorem 88 could easily be adapted to allow for such non-linear (homogeneous) constraints, thereby expanding their practical use if one combines them with non-linear solvers.

$f \in \mathscr{G}(\mathscr{X}_G)$, Equation $(7.9)_{225}$ implies that

$$\big(\otimes_{s\in G}\underline{P}_s\big)(f) = \min\bigg\{\sum_{x_G\in\mathscr{X}_G} f(x_G)p_G(x_G)\colon\, p_G \in \otimes_{s\in G}\mathscr{F}_s\bigg\}. \tag{7.15}$$

The following immediate consequence of Theorem $88_{\frown}$ implies that this optimisation problem can be solved by means of standard linear programming techniques.

Corollary 89. *Consider a finite number of local credal sets $\mathscr{F}_s$, $s \in G$, each of which is fully characterised means of Equations* $(7.10)_{226}$ *and* $(7.11)_{226}$*. Then $\otimes_{s\in G}\mathscr{F}_s$ consists of those real-valued functions p_G on $\mathscr{X}_G$ for which $\sum_{z_G\in\mathscr{X}_G} p_G(z_G) = 1$ and, for all $s \in G$ and $x_{G\setminus\{s\}} \in \mathscr{X}_{G\setminus\{s\}}$:*

$$\big(\forall\,\gamma\in\Gamma_s\big)\quad \sum_{z_s\in\mathscr{X}_s} p(x_{G\setminus\{s\}}, z_s)\gamma(z_s) \geq 0.$$

Proof of Corollary 89. This is an immediate consequence of Theorem $88_{\frown}$ and the fact that—in this special case of the independent natural extension—$P(s) = \emptyset$ and $N(s) = G\setminus\{s\}$ for all $s \in G$. □

7.4.4 *Enumerating the extreme points explicitly*

A final brute force method for computing $\underline{P}^{\mathrm{irr}}_G(\cdot)$ is by enumerating the extreme points of $\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G$. Since we know from Section $2.6.2_{55}$ that $\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G$ is the convex hull of its extreme points, it follows from Equation $(7.9)_{225}$ that, for all $f \in \mathscr{G}(\mathscr{X}_G)$:

$$\underline{P}^{\mathrm{irr}}_G(f) = \min\bigg\{\sum_{x_G\in\mathscr{X}_G} f(x_G)p_G(x_G)\colon\, p_G \in \mathrm{ext}(\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G)\bigg\}. \tag{7.16}$$

Computing $\underline{P}^{\mathrm{irr}}_G(f)$ is therefore a matter of evaluating $\sum_{x_G\in\mathscr{X}_G} f(x_G)p_G(x_G)$ for all $p_G \in \mathrm{ext}(\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G)$. If the extreme points of $\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G$ are available, and if their number is finite and sufficiently small, this approach allows us to compute $\underline{P}^{\mathrm{irr}}_G(f)$ at blazing speeds.

Unfortunately, these extreme points are usually not available, and if they are, their number is often huge. One example where the extreme points are known, are imprecise Markov chains, which are credal networks under epistemic irrelevance of which the underlying DAG is a directed chain; see for example Figure 7.1_{217}. In that specific case, the extreme points of $\mathscr{F}^{\mathrm{irr}}_G \rfloor \mathscr{X}_G$ correspond to probability trees whose local probability mass functions are the extreme points of the local models $\mathscr{F}_{s\rfloor x_{P(s)}}$; see for example Reference [43]. However, since the number of such probability trees is exponential in the length of the Markov chain, a direct application of Equation (7.16) becomes intractable for long Markov chains.

If the extreme points are not directly available, they can be derived from the characterisations in terms of linear constraints that were discussed in the previous section, by applying standard vertex enumeration algorithms. However, since the number of linear constraints that characterise $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ is exponential in the size of the network, constructing the extreme points of $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ in this way will be intractable for large networks, and their number will usually be exponential in the size of the network as well. Furthermore, even if it is tractable, constructing these extreme points will typically be less efficient than computing $\underline{P}_G^{\mathrm{irr}}(f)$ directly by solving a linear program. Nevertheless, if we need to compute $\underline{P}_G^{\mathrm{irr}}(f)$ for many different $f \in \mathscr{G}(\mathscr{X}_G)$, it might sometimes still be a good idea to construct the extreme points of $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ explicitly. Once we have constructed these extreme points, we can use them to compute $\underline{P}_G^{\mathrm{irr}}(f)$ for as many gambles $f \in \mathscr{G}(\mathscr{X}_G)$ as we want, whereas a direct linear programming approach would require us to solve a new linear program for every new gamble.

In the remainder of this section, we focus on a simple special case: a credal network with two disconnected binary variables. For all $i \in G = \{1,2\}$, the variable X_i assumes values in its binary state space $\mathscr{X}_i = \{h_i, t_i\}$ and has a given local uncertainty model in the form of a credal set $\mathscr{F}_i$. In this case, $\mathscr{F}_G^{\mathrm{irr}} \rfloor \mathscr{X}_G$ is equal to the independent natural extension $\mathscr{F}_1 \otimes \mathscr{F}_2$ of the local models $\mathscr{F}_i$, $i \in \{1,2\}$. We intend to show that it is possible to obtain elegant closed-form expressions for the extreme points of this independent natural extension. The starting point is to describe $\mathscr{F}_1 \otimes \mathscr{F}_2$ in terms of linear constraints, by applying Corollary $89_{\cap}$. As we will see, the corresponding vertex enumeration problem can then be solved symbolically.

Since X_i, $i \in \{1,2\}$, is a binary variable, the credal set $\mathscr{F}_i$ is completely characterised by the lower and upper probabilities of one of its elements. For example: the lower and upper probability of h_i, which we denote as $\underline{p}(h_i)$ and $\overline{p}(h_i)$, respectively. Each of these two probabilities defines a unique probability mass function on $\mathscr{X}_i$ and $\mathscr{F}_i$ is equal to their convex hull. In other words: $\mathscr{F}_i$ consists of the probability mass functions p on $\mathscr{X}_i$ for which $p(h_i) \in [\underline{p}(h_i), \overline{p}(h_i)]$. The corresponding lower and upper probability of t_i is equal to $\underline{p}(t_i) := 1 - \overline{p}(h_i)$ and $\overline{p}(t_i) := 1 - \underline{p}(h_i)$, respectively.

Before we can apply Corollary $89_{\cap}$, we first need to characterise $\mathscr{F}_i$, $i \in \{1,2\}$, by means of the unitary constraint and a finite number of linear *homogeneous* inequalities. In this binary case, we can use the following two inequalities:

$$\begin{aligned} \overline{p}(t_i)p(h_i) - \underline{p}(h_i)p(t_i) &\geq 0 \\ -\underline{p}(t_i)p(h_i) + \overline{p}(h_i)p(t_i) &\geq 0. \end{aligned}$$

By applying Corollary $89_{\cap}$, these local constraints can be turned into global ones. We find that $\mathscr{F}_1 \otimes \mathscr{F}_2$ consists of the real-valued functions p on $\mathscr{X}_1 \times \mathscr{X}_2$ for which

$$p(h_1,h_2)+p(h_1,t_2)+p(t_1,h_2)+p(t_1,t_2)=1 \tag{7.17}$$

and

$$\overline{p}(t_1)p(h_1,h_2)-\underline{p}(h_1)p(t_1,h_2)\geq 0 \tag{I1}$$
$$-\underline{p}(t_1)p(h_1,h_2)+\overline{p}(h_1)p(t_1,h_2)\geq 0 \tag{I2}$$
$$\overline{p}(t_1)p(h_1,t_2)-\underline{p}(h_1)p(t_1,t_2)\geq 0 \tag{I3}$$
$$-\underline{p}(t_1)p(h_1,t_2)+\overline{p}(h_1)p(t_1,t_2)\geq 0 \tag{I4}$$
$$\overline{p}(t_2)p(h_1,h_2)-\underline{p}(h_2)p(h_1,t_2)\geq 0 \tag{I5}$$
$$-\underline{p}(t_2)p(h_1,h_2)+\overline{p}(h_2)p(h_1,t_2)\geq 0 \tag{I6}$$
$$\overline{p}(t_2)p(t_1,h_2)-\underline{p}(h_2)p(t_1,t_2)\geq 0 \tag{I7}$$
$$-\underline{p}(t_2)p(t_1,h_2)+\overline{p}(h_2)p(t_1,t_2)\geq 0. \tag{I8}$$

If the inequalities in equations (I1)–(I8) are replaced by equalities, we will refer to the resulting equations as (E1)–(E8):

$$\overline{p}(t_1)p(h_1,h_2)-\underline{p}(h_1)p(t_1,h_2)=0 \tag{E1}$$
$$-\underline{p}(t_1)p(h_1,h_2)+\overline{p}(h_1)p(t_1,h_2)=0 \tag{E2}$$
$$\overline{p}(t_1)p(h_1,t_2)-\underline{p}(h_1)p(t_1,t_2)=0 \tag{E3}$$
$$-\underline{p}(t_1)p(h_1,t_2)+\overline{p}(h_1)p(t_1,t_2)=0 \tag{E4}$$
$$\overline{p}(t_2)p(h_1,h_2)-\underline{p}(h_2)p(h_1,t_2)=0 \tag{E5}$$
$$-\underline{p}(t_2)p(h_1,h_2)+\overline{p}(h_2)p(h_1,t_2)=0 \tag{E6}$$
$$\overline{p}(t_2)p(t_1,h_2)-\underline{p}(h_2)p(t_1,t_2)=0 \tag{E7}$$
$$-\underline{p}(t_2)p(t_1,h_2)+\overline{p}(h_2)p(t_1,t_2)=0. \tag{E8}$$

These equations can be used to state the following necessary and sufficient conditions for being an extreme point of $\mathscr{F}_1\otimes\mathscr{F}_2$.

Proposition 90. *Every $p\in\operatorname{ext}(\mathscr{F}_1\otimes\mathscr{F}_2)$ is the unique solution to the unitary constraint* (7.17) *and three of the equations* (E1)–(E8).

Proof of Proposition 90. Consider an arbitrary extreme point p_{ext} of $\mathscr{F}_1\otimes\mathscr{F}_2$. Denote by $\mathfrak{E}^{=}$ the set of all equalities in (E1)–(E8) that are satisfied by p_{ext} and use $\mathfrak{E}_1^{=}$ to denote the union of $\mathfrak{E}^{=}$ and the unitary constraint (7.17). Let C be the intersection of the hyperplanes that are defined by the equalities in $\mathfrak{E}_1^{=}$. Clearly, $p_{\text{ext}}\in C$. Since all the equalities in $\mathfrak{E}^{=}$ are linear and homogeneous, C will either be a singleton (and therefore be equal to $\{p_{\text{ext}}\}$) or contain a line (that necessarily contains p_{ext}).

Assume *ex absurdo* that $C \neq \{p_{\text{ext}}\}$, which implies that there is some line $L \subseteq C$ such that $p_{\text{ext}} \in L$. Since p_{ext} is an element of $\mathscr{F}_1 \otimes \mathscr{F}_2$, we know that it satisfies the unitary constraint and the inequalities $(\text{I1})_{\cap}$–$(\text{I8})_{\cap}$. If we denote by $\mathfrak{E}^{>}$ the set of inequalities in $(\text{I1})_{\cap}$–$(\text{I8})_{\cap}$ that are strictly satisfied by p_{ext} (and therefore not with equality), we can construct a closed ball B around p_{ext}, with radius $\varepsilon > 0$, such that all the elements of B satisfy the inequalities in $\mathfrak{E}^{>}$. It should now be clear that the elements of $B \cap C$ satisfy $(\text{I1})_{\cap}$–$(\text{I8})_{\cap}$, because these inequalities are either an element of $\mathfrak{E}^{>}$ or their corresponding equality is an element of $\mathfrak{E}^{=}$. Since $B \cap C$ also satisfies the unitary constraint (because it is a subset of C), it follows that $B \cap C$ is a subset of $\mathscr{F}_1 \otimes \mathscr{F}_2$. This implies that the closed line segment $B \cap L$, of which the midpoint is equal to p_{ext}, is a subset of $\mathscr{F}_1 \otimes \mathscr{F}_2$. Therefore, if we denote the endpoints of $B \cap L$ by p_1 and p_2, we find that $p_1, p_2 \in \mathscr{F}_1 \otimes \mathscr{F}_2$ and $p_{\text{ext}} = (p_1 + p_2)/2$. Since this contradicts the fact that p_{ext} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$, we may conclude that $C = \{p_{\text{ext}}\}$.

As a direct consequence of the fact that $C = \{p_{\text{ext}}\}$, it follows that p_{ext} is the unique solution of the equations in $\mathfrak{E}_1^{=}$. Since we are working in $\mathbb{R}^4$ and because all the equations in $\mathfrak{E}_1^{=}$ are linear, this implies that $\mathfrak{E}_1^{=}$ contains a subset of four equalities for which it also holds that p_{ext} is their unique solution. This subset must contain the unitary constraint $(7.17)_{\cap}$ because it would otherwise contain only homogeneous linear equalities, which would imply that their solution set can only be a singleton if it is equal to the origin, which clearly is not the case here. Therefore we conclude that p_{ext} is the unique solution to the unitary constraint and three equations in $\mathfrak{E}^{=}$, which by construction is a subset of the equalities $(\text{E1})_{\cap}$–$(\text{E8})_{\cap}$. □

Proposition 91. *Consider any ternary subset of* $(\text{E1})_{\cap}$–$(\text{E8})_{\cap}$ *that, in combination with the unitary constraint* $(7.17)_{\cap}$*, has a unique solution* p*. Then* $p \in \mathscr{F}_1 \otimes \mathscr{F}_2$ *if and only if* $p \in \text{ext}(\mathscr{F}_1 \otimes \mathscr{F}_2)$.

Proof of Proposition 91. One of the directions is trivial: $p \in \text{ext}(\mathscr{F}_1 \otimes \mathscr{F}_2)$ clearly implies that $p \in \mathscr{F}_1 \otimes \mathscr{F}_2$. Now assume *ex absurdo* that $p \in \mathscr{F}_1 \otimes \mathscr{F}_2$ and $p \notin \text{ext}(\mathscr{F}_1 \otimes \mathscr{F}_2)$, which implies that there are $p_a, p_b \in (\mathscr{F}_1 \otimes \mathscr{F}_2) \setminus \{p\}$ and $\lambda \in (0,1)$ such that $p = \lambda p_a + (1-\lambda) p_b$. Without loss of generality, we may assume that p is the unique solution of the unitary constraint $(7.17)_{\cap}$ and Equations $(\text{E1})_{\cap}$, $(\text{E2})_{\cap}$ and $(\text{E3})_{\cap}$; the proof for other ternary subsets of $(\text{E1})_{\cap}$–$(\text{E8})_{\cap}$ is analogous. Since p_a is a probability mass function that differs from p, it follows that p_a does not satisfy Equation $(\text{E1})_{\cap}$, $(\text{E2})_{\cap}$ or $(\text{E3})_{\cap}$. Without loss of generality, we may assume that Equation $(\text{E1})_{\cap}$ is not satisfied; the proof for the other cases is analogous. Since p_a and p_b belong to $\mathscr{F}_1 \otimes \mathscr{F}_2$, we know that they satisfy Equation $(\text{I1})_{\cap}$. Since p_a does not satisfy Equation $(\text{E1})_{\cap}$, this implies that $p = \lambda p_a + (1-\lambda) p_b$ satisfies Equation $(\text{I1})_{\cap}$ strictly, which implies that p does not satisfy Equation $(\text{E1})_{\cap}$, a contradiction. □

	$p(h_1,h_2)\Sigma$	$p(h_1,t_2)\Sigma$	$p(t_1,h_2)\Sigma$	$p(t_1,t_2)\Sigma$
p_{S1}	$\underline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\underline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(t_2)$
p_{S2}	$\underline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\underline{p}(t_2)$
p_{S3}	$\overline{p}(h_1)\underline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_2)$
p_{S4}	$\overline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(t_2)$
p_{A1}	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(t_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(h_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(t_2)$
p_{A2}	$\underline{p}(h_1)\overline{p}(h_1)\overline{p}(h_2)$	$\underline{p}(h_1)\overline{p}(h_1)\underline{p}(t_2)$	$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)$	$\overline{p}(t_1)\overline{p}(h_1)\underline{p}(t_2)$
p_{A3}	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(t_2)$
p_{A4}	$\underline{p}(t_1)\underline{p}(h_1)\overline{p}(h_2)$	$\overline{p}(h_1)\overline{p}(t_1)\underline{p}(t_2)$	$\underline{p}(t_1)\overline{p}(t_1)\overline{p}(h_2)$	$\underline{p}(t_1)\overline{p}(t_1)\underline{p}(t_2)$
p_{B1}	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(h_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(t_1)$
p_{B2}	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(h_1)$	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(t_1)$
p_{B3}	$\underline{p}(h_2)\overline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(h_2)\underline{p}(t_2)\overline{p}(h_1)$	$\underline{p}(h_2)\overline{p}(h_2)\underline{p}(t_1)$	$\overline{p}(t_2)\overline{p}(h_2)\underline{p}(t_1)$
p_{B4}	$\underline{p}(t_2)\underline{p}(h_2)\overline{p}(h_1)$	$\underline{p}(t_2)\overline{p}(t_2)\overline{p}(h_1)$	$\overline{p}(h_2)\overline{p}(t_2)\underline{p}(t_1)$	$\underline{p}(t_2)\overline{p}(t_2)\underline{p}(t_1)$

Table 7.1: Numerators of the candidates for the extreme points of the independent natural extension of two binary variables

The extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ can therefore be found in the following way. First, we need to consider every possible subset of three equalities from $(\mathrm{E1})_{231}$–$(\mathrm{E8})_{231}$. Then, for every such combination of three equalities, we need to combine them with the unitary constraint and check whether this results in a unique solution. If this is the case, we need to check whether this unique solution satisfies the inequalities in $(\mathrm{I1})_{231}$–$(\mathrm{I8})_{231}$. If yes, then that unique solution is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

Since there are 56 possible ways of choosing three equalities out of eight, solving this problem manually might seem like a daunting task. However, due to the extreme symmetry—switching X_1 and X_2, h_1 and t_1 or h_2 and t_2 yields an equivalent set of inequalities—only 7 of those 56 cases need to be considered, as the others can be derived from these 7 by an argument of symmetry. By exploiting this symmetry, we have managed to obtain elegant closed-form expressions for the extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

	Σ		Σ		Σ
p_{S1}	1	p_{A1}	$\underline{p}(h_1)\overline{p}(t_2)+\overline{p}(h_1)\underline{p}(h_2)$	p_{B1}	$\underline{p}(h_2)\overline{p}(t_1)+\overline{p}(h_2)\underline{p}(h_1)$
p_{S2}	1	p_{A2}	$\underline{p}(h_1)\overline{p}(h_2)+\overline{p}(h_1)\underline{p}(t_2)$	p_{B2}	$\underline{p}(t_2)\overline{p}(t_1)+\overline{p}(t_2)\underline{p}(h_1)$
p_{S3}	1	p_{A3}	$\underline{p}(t_1)\overline{p}(t_2)+\overline{p}(t_1)\underline{p}(h_2)$	p_{B3}	$\underline{p}(h_2)\overline{p}(h_1)+\overline{p}(h_2)\underline{p}(t_1)$
p_{S4}	1	p_{A4}	$\underline{p}(t_1)\overline{p}(h_2)+\overline{p}(t_1)\underline{p}(t_2)$	p_{B4}	$\underline{p}(t_2)\overline{p}(h_1)+\overline{p}(t_2)\underline{p}(t_1)$

Table 7.2: Denominators of the candidates for the extreme points of the independent natural extension of two binary variables

Theorem 92. *Tables 7.1$_{\frown}$ and 7.2 contain expressions for up to 12 probability mass functions, which can be obtained by dividing the rows of Table 7.1$_{\frown}$ by the corresponding denominator in Table 7.2. Based on the particular values of $\underline{p}(h_1)$, $\overline{p}(h_1)$, $\underline{p}(t_1)$, $\overline{p}(t_1)$, $\underline{p}(h_2)$, $\overline{p}(h_2)$, $\underline{p}(t_2)$ and $\overline{p}(t_2)$, the diagram in Figure 7.2$_{\frown}$—which is continued in Figures 7.3$_{236}$ and 7.4$_{237}$—indicates which subset of these probability mass functions are the extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$. In these diagrams, we use the shorthand notation $p_{\mathrm{S1=S2}}$ to denote that p_{S1} and p_{S2} are two coinciding extreme points, and similarly for other combinations.*[13]

The diagrams in Figures 7.2$_{\frown}$–7.4$_{237}$ might seem rather complicated, but this is only because they treat a number of special cases with probability zero. The main result can be summarised quite easily. If one of the local models is precise or vacuous, then the independent natural extension has the same extreme points (p_{S1}, p_{S2}, p_{S3} and p_{S4}) as—and therefore coincides with—the so-called strong product of $\mathscr{F}_1$ and $\mathscr{F}_2$,[14,15] defined by

$$\mathscr{F}_1 \times \mathscr{F}_2 := \mathrm{CH}\big(\{p_1 \cdot p_2 \colon p_1 \in \mathscr{F}_1, p_2 \in \mathscr{F}_2\}\big), \tag{7.18}$$

where CH is the convex hull operator. In all other cases, $\mathscr{F}_1 \otimes \mathscr{F}_2$ has up to four additional extreme points. If we ignore the cases with lower probability zero, then these extreme points are determined by the relative value of $\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2)$ and $\overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2)$. Depending on which one of these two parameters is higher, the additional extreme points are either p_{A1},

[13] We only indicate this in the instances where we consider this information to be of interest. Otherwise, we simply depict one representative element of the set of coinciding extreme points.

[14] This is the strong extension of a credal network with local models $\mathscr{F}_1$ and $\mathscr{F}_2$; see Section 5.5$_{145}$ as well.

[15] This partial result was already proved in Reference [46, Section 5.5].

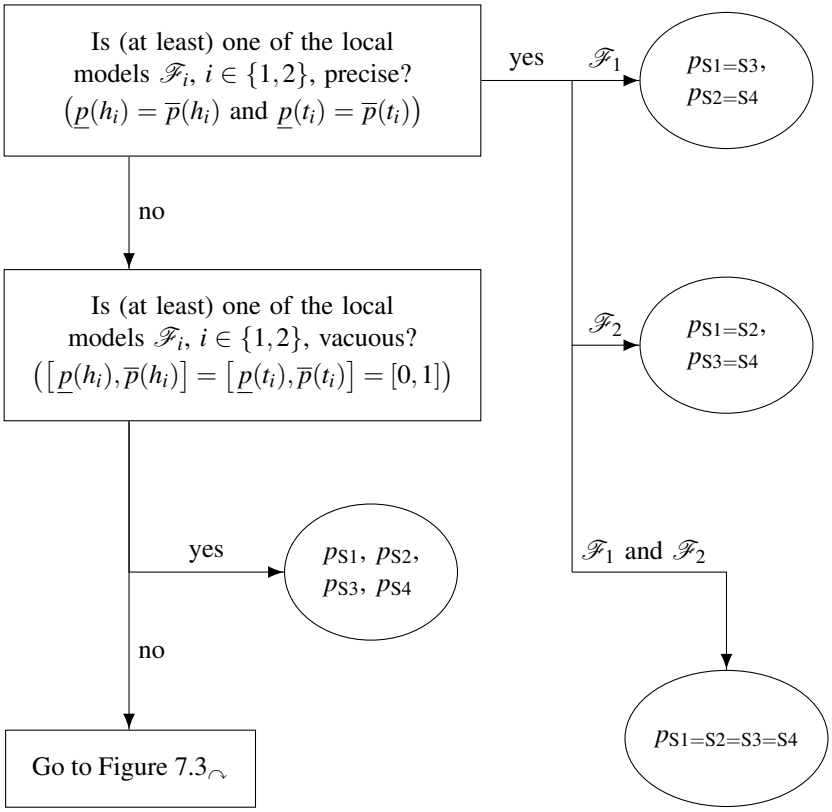


Figure 7.2: Diagram to obtain the extreme points of the independent natural extension of two binary variables (PART 1)

p_{A2}, p_{A3} and p_{A4} or p_{B1}, p_{B2}, p_{B3} and p_{B4}. If the parameters are equal, these two sets of additional extreme points are identical. Walley's well-known numerical example [106, Section 9.3.4] has six extreme points; it corresponds to the case $\mathscr{F}_1 = \mathscr{F}_2$ in Figure 7.4$_{237}$.

7.5 RECURSIVELY DECOMPOSABLE NETWORKS

So far, the most promising algorithms for credal networks under epistemic irrelevance have been developed for networks with a tree topology [6, 30, 42]. The underlying reason is that for trees, the global model can be constructed recursively, by combining the local models using marginal extension and independent natural extension [42, Section 4]. This allows for the development of efficient recursive algorithms. The most well-known example is the algorithm

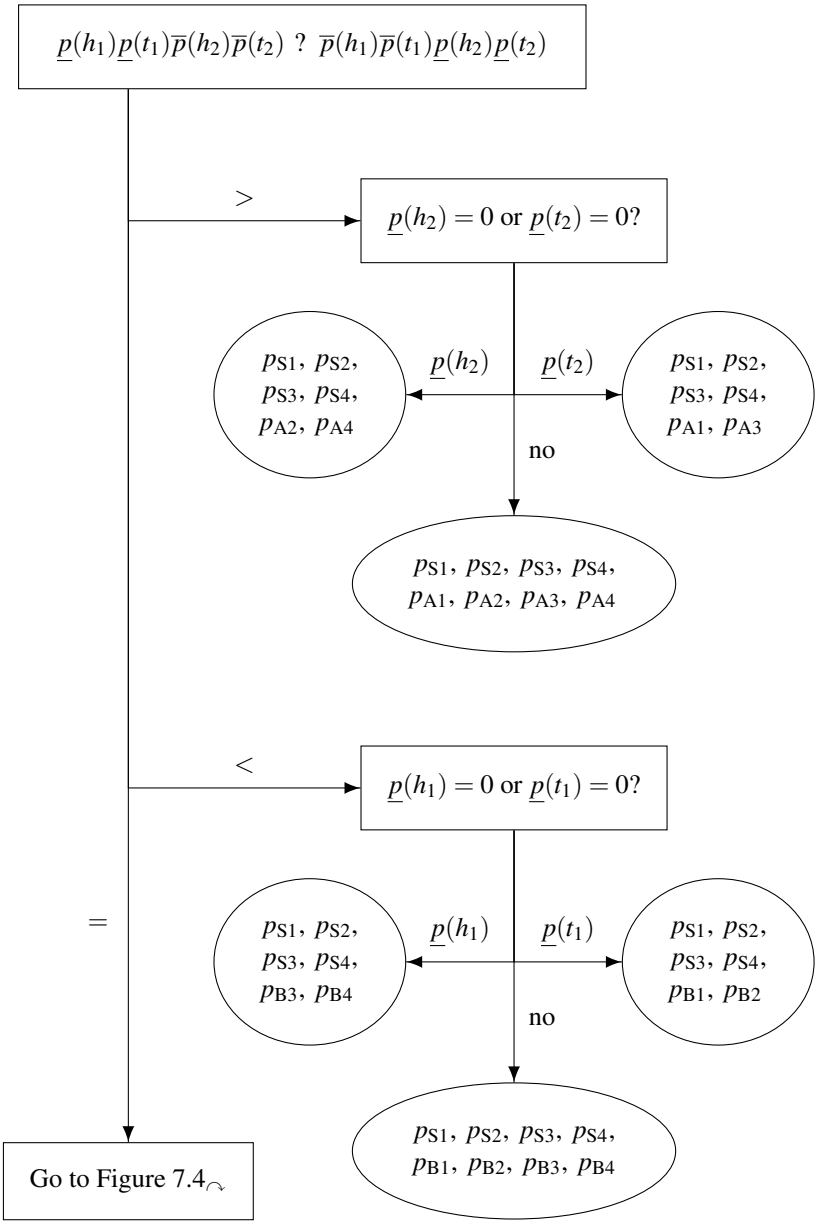


Figure 7.3: Diagram to obtain the extreme points of the independent natural extension of two binary variables (PART 2)

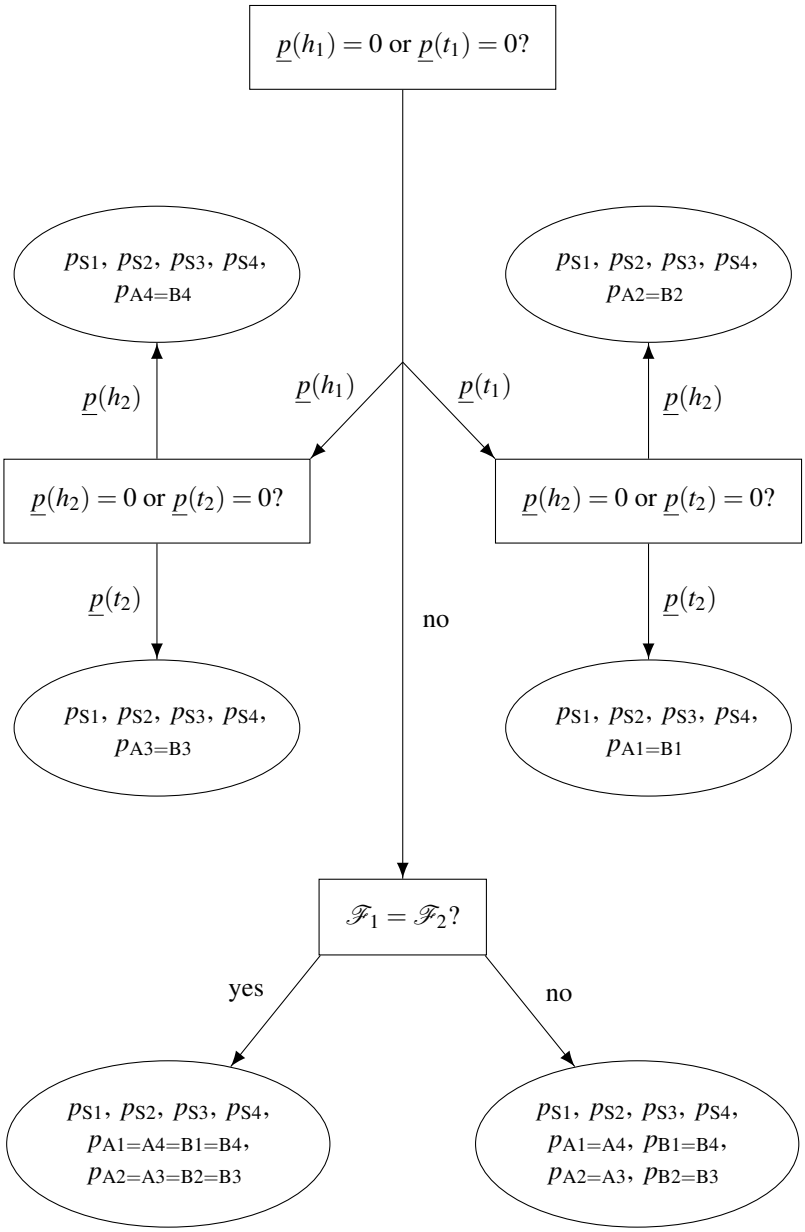


Figure 7.4: Diagram to obtain the extreme points of the independent natural extension of two binary variables (PART 3)

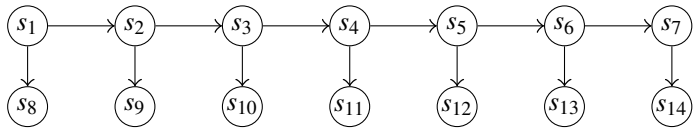


Figure 7.5: Example of a hidden Markov model (HMM)

by De Cooman et al. [42], which is able to compute $\underline{P}_q(f\rfloor x_E)$ efficiently, for any $q \in G$, $E \subseteq G \setminus \{q\}$, $f \in \mathscr{G}(\mathscr{X}_q)$ and $x_E \in \mathscr{X}_E$. As we are about to show, this recursive approach can be extended in two ways. First of all, this approach is not limited to networks with a tree topology; it can be applied to a more general class of networks, which we call recursively decomposable. Secondly, it can be applied to more general inference problems, with multiple query nodes.

7.5.1 *Recursively decomposable DAGs*

We start by introducing some additional notation and terminology. Two nodes $s,t \in G$ are said to be *incomparable*, denoted by $s \parallel t$, if they are different and if there is no directed path that connects them:

$$s \parallel t \;\Leftrightarrow\; \big(s \neq t, s \notin D(t) \text{ and } t \notin D(s)\big)$$

We then call a DAG *recursively decomposable* if every pair of incomparable nodes has no common descendants:

$$(\forall s,t \in G)\;\big(s \parallel t \;\Rightarrow\; D(s) \cap D(t) = \emptyset\big).$$

To the best of our knowledge, this type of DAG has never been considered before. Why we call these DAGs recursively decomposable will become clear in the next section, where we show that the irrelevant natural extension of credal networks with such a DAG can be decomposed in a natural, recursive way. For now, we focus on the topological properties of recursively decomposable DAGs.

First of all: the class of recursively decomposable DAGs includes trees as a special case. In a tree, two nodes s and t are incomparable if and only if they belong to a different branch, where a branch is taken to be a directed path from the root of the tree to one of its leaves. Clearly, this implies that s and t have no common descendants. Therefore, recursively decomposable DAGs also include Markov chains [see Figure 7.1$_{217}$] and hidden Markov models (HMMs) [see Figure 7.5] as special cases.

However, not every recursively decomposable DAG is a tree. A first—rather trivial—example are forests, which are unions of pairwise disjoint trees. By an argument similar to that for trees, it follows that every forest is a recursively decomposable DAG. A second class of recursively decomposable

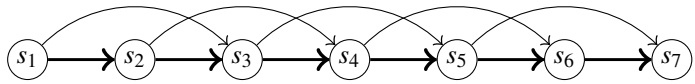


Figure 7.6: Example of a Markov chain of order 2

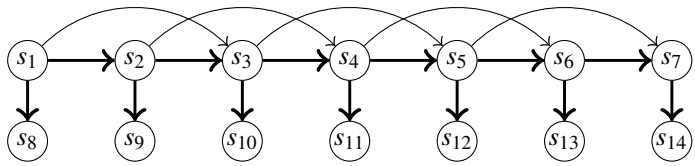


Figure 7.7: Example of a hidden Markov model (HMM) of order 2

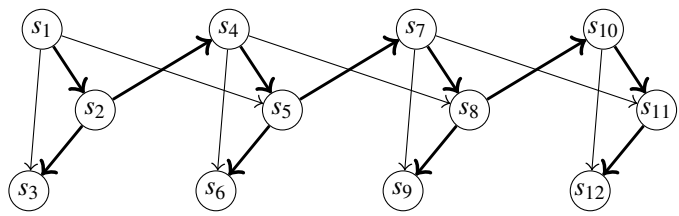


Figure 7.8: Example of a dynamic network that is recursively decomposable

DAGs that are not trees are Markov chains of order m, with $m \geq 2$. They have no incomparable nodes and are therefore trivially recursively decomposable. Figure 7.6 depicts a simple Markov chain of order 2. Hidden Markov models of order m, with $m \geq 2$ are also recursively decomposable; Figure 7.7 depicts an example of a hidden Markov model of order 2. Figures 7.8 and 7.9$_{\curvearrowright}$ provide two additional examples of recursively decomposable DAGs that do not fit into any of the above-mentioned general classes. Figure 7.8 can be regarded as a simple dynamic network; Figure 7.9$_{\curvearrowright}$ is just some random example.

For any DAG with nodes G and any node $s \in G$, we now define two important subsets of $C(s)$ and $P(s)$, respectively. The first set consists of the roots of the sub-DAG that corresponds to $D(s)$:

$$\tilde{C}(s) := Ro(D(s)) = \left\{ t \in D(s) \colon P_{D(s)}(t) = \emptyset \right\} \tag{7.19}$$

The following result shows that $\tilde{C}(s)$ is a subset of the children of s and provides a convenient alternative characterisation for this set.

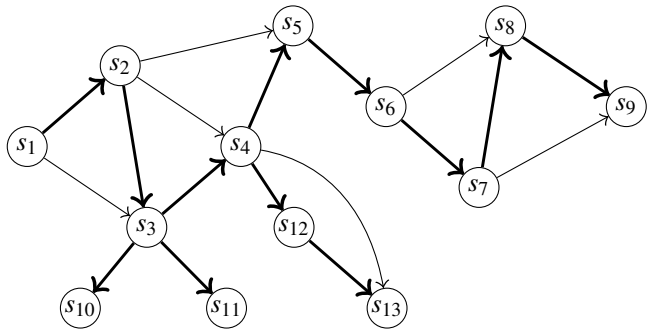


Figure 7.9: Example of a recursively decomposable DAG

Proposition 93. *Consider any $s \in G$. Then $\tilde{C}(s) \subseteq C(s)$ and, for all $t \in G$, $t \in \tilde{C}(s)$ if and only if the only directed path from s to t is a single edge $s \to t$.*

Proof of Proposition 93$_{\curvearrowleft}$. Consider any $t \in \tilde{C}(s)$, which implies that $t \in D(s)$ and $P_{D(s)}(t) = \emptyset$. Since $t \in D(s)$, there is at least one directed path from s to t. Consider any such directed path $s = s_1, \ldots, s_{n-1}, s_n = t$ from s to t. The node s_{n-1} is then clearly an element of $D(s) \cup \{s\}$. Hence, since $P_{D(s)}(t) = \emptyset$, it must be that $s_{n-1} = s$. Since $s_1 = s$, and because a DAG has by definition no cycles, this implies that $n - 1 = 1$, which implies that the path $s = s_1, \ldots, s_{n-1}, s_n = t$ consists of a single edge $s \to t$.

Next, consider any $t \in G$ for which the only directed path from s to t is a single edge $s \to t$. It then follows from $s \to t$ that $t \in D(s)$. Now assume *ex absurdo* that $P_{D(s)}(t) \neq \emptyset$. This implies that there is some $u \in P(t)$ such that $u \in D(s)$. Since $u \in D(s)$, there is a directed path $s = s_1, \ldots, s_n = u$ from s to u, with $n \geq 2$. Since, $u \in P(t)$, this implies that $s = s_1, \ldots, s_n = u, t$ is a directed path from s to t that consists of more than one edge, a contradiction. Therefore, we may conclude that $P_{D(s)}(t) = \emptyset$, which, since $t \in D(s)$, implies that $t \in \tilde{C}(s)$. □

We will call the elements of $\tilde{C}(s)$ the *induced children* of s. The *induced parents* of s are then defined as

$$\tilde{P}(s) := \{t \in P(s) \colon s \in \tilde{C}(t)\}.$$

Table 7.3$_{\curvearrowright}$ illustrates these concepts by means of examples, for nodes in various recursively decomposable DAGs.

For any DAG, regardless of whether it is recursively decomposable, we can use these concepts to select a subset of its edges: the edges $s \to t$ for which $s \in G$ and $t \in \tilde{C}(s)$. By removing all the other edges, we obtain a new DAG, which we call the *induced DAG*. For any node $s \in G$, $\tilde{C}(s)$ and $\tilde{P}(s)$ are the set

	s	$D(s)$	$C(s)$	$\tilde{C}(s)$	$P(s)$	$\tilde{P}(s)$
Figure 7.5_{238}	s_5	$s_6, s_7, s_{12}, s_{13}, s_{14}$	s_6, s_{12}	s_6, s_{12}	s_4	s_4
Figure 7.1_{217}	s_3	s_4, s_5, s_6, s_7	s_4	s_4	s_2	s_2
Figure 7.6_{239}	s_4	s_5, s_6, s_7	s_5, s_6	s_5	s_2, s_3	s_3
Figure 7.7_{239}	s_5	$s_6, s_7, s_{12}, s_{13}, s_{14}$	s_6, s_7, s_{12}	s_6, s_{12}	s_3, s_4	s_4
Figure 7.8_{239}	s_8	$s_9, s_{10}, s_{11}, s_{12}$	s_9, s_{10}	s_9, s_{10}	s_4, s_7	s_7
Figure $7.9_{\frown}$	s_3	$s_4, \ldots, s_{13}$	s_4, s_{10}, s_{11}	s_4, s_{10}, s_{11}	s_1, s_2	s_2

Table 7.3: Examples of $\tilde{C}(s)$ and $\tilde{P}(s)$, for various recursively decomposable DAGs

of children and parents of s in this induced DAG, respectively. We use $\tilde{D}(s)$ to denote the descendants of s in this induced DAG. In Figures 7.1_{217} and 7.5_{238}, the induced DAG coincides with the original DAG. In Figures 7.6_{239}, 7.7_{239}, 7.8_{239} and $7.9_{\frown}$, the edges of the induced DAG have been thickened. The induced DAG is closely related to the original one. Intuitively speaking, the induced DAG is obtained by removing shortcuts from the original DAG: if there is a direct edge from s to t as well as a longer directed path from s to t, the induced DAG removes the direct edge; this can be regarded as an informal statement of Proposition 93_{239}. It is therefore not surprising that both DAGs assign the same descendants to every node.

Proposition 94. *For any $s \in G$, we have that $D(s) = \tilde{D}(s)$.*

Proof of Proposition 94. Since the edges of the induced DAG are a subset of the edges of the original DAG, $\tilde{D}(s)$ is clearly a subset of $D(s)$. So consider any $t \in D(s)$ and assume *ex absurdo* that $t \notin \tilde{D}(s)$.

Consider the sub-DAG that corresponds to $D(s)$. Since $t \in D(s)$, it follows from Lemma $95_{\frown}$ that this sub-DAG has some root node $s_1 \in Ro(D(s))$ such that $s_1 \sqsubseteq t$. It follows from Equation $(7.19)_{239}$ and Proposition 93_{239} that $s_1 \in \tilde{C}(s) \subseteq C(s)$. If $s_1 = t$, then $t \in \tilde{C}(s) \subseteq \tilde{D}(s)$, a contradiction. Therefore, it must be that $s_1 \neq t$. Since $s_1 \sqsubseteq t$, this implies that $t \in D(s_1)$.

Consider now the sub-DAG that corresponds to $D(s_1)$. Since $t \in D(s_1)$, it follows from Lemma $95_{\frown}$ that this sub-DAG has some root node $s_2 \in Ro(D(s_1))$ such that $s_2 \sqsubseteq t$. It follows from Equation $(7.19)_{239}$ and Proposition 93_{239} that $s_2 \in \tilde{C}(s_1) \subseteq C(s_1)$. If $s_2 = t$, then $t \in \tilde{C}(s_1)$. Since we already know that $s_1 \in \tilde{C}(s)$, this would imply that $t \in \tilde{D}(s)$, a contradiction. Therefore, it must be that $s_2 \neq t$. Since $s_2 \sqsubseteq t$, this implies that $t \in D(s_2)$.

By continuing in this way, we obtain an infinite sequence of nodes $s, s_1, s_2, \ldots, s_n, \ldots$ in G such that $s_1 \in C(s)$ and, for all $i \in \mathbb{N}$, $s_{i+1} \in C(s_i)$.

Since a DAG contains no cycles, all nodes in this sequence must be different. Since $|G|$ is finite, this is a contradiction. □

Lemma 95. *Consider a DAG with nodes G. For every node $t \in G$, there is some root node $s \in Ro(G)$ such that $s \sqsubseteq t$.*

Proof of Lemma 95. If $P(t) = \emptyset$, then $t \in Ro(G)$, which concludes the proof [choose $s = t$]. Otherwise, let t_1 be an arbitrary element of $P(t)$. This implies that $t_1 \sqsubseteq t$. If $P(t_1) = \emptyset$, then $t_1 \in Ro(G)$, which concludes the proof [choose $s = t_1$]. Otherwise, let t_2 be an arbitrary element of $P(t_1)$. Since $t_1 \sqsubseteq t$, this implies that $t_2 \sqsubseteq t$. If $P(t_2) = \emptyset$, then $t_2 \in Ro(G)$, which concludes the proof [choose $s = t_2$]. If we continue in this way, this process will eventually end, thereby concluding the proof. Indeed, assume *ex absurdo* that this process does not end. Then we obtain an infinite sequence of nodes $t, t_1, t_2, \ldots, t_n, \ldots$ in G such that $t_1 \in P(t)$ and, for all $i \in \mathbb{N}$, $t_{i+1} \in P(t_i)$. Since a DAG contains no cycles, all nodes in this sequence most be different. Since $|G|$ is finite, this is a contradiction. □

Proposition $94_{\frown}$ implies that the root nodes and leaves of the induced DAG are identical to those of the original DAG.

Corollary 96. *For any node $s \in G$, we have that $C(s) = \emptyset \Leftrightarrow \tilde{C}(s) = \emptyset$ and $P(s) = \emptyset \Leftrightarrow \tilde{P}(s) = \emptyset$.*

Proof of Corollary 96. For any $s \in G$, we have that

$$C(s) = \emptyset \Leftrightarrow D(s) = \emptyset \Leftrightarrow \tilde{D}(s) = \emptyset \Leftrightarrow \tilde{C}(s) = \emptyset$$

and

$$P(s) \neq \emptyset \Leftrightarrow (\exists t \in G)\, s \in D(t) \Leftrightarrow (\exists t \in G)\, s \in \tilde{D}(t) \Leftrightarrow \tilde{P}(s) \neq \emptyset,$$

where, in each case, the first and last equivalences are trivial and the second one follows from Proposition $94_{\frown}$. □

Consequently, for forests, and for trees in particular, the induced DAG is identical to the original DAG.

Corollary 97. *For a forest, the induced DAG is identical to the original one.*

Proof of Corollary 97. By its definition, the edges of the induced DAG are a subset of those of the original DAG. So consider any edge $s \to t$ in the original DAG and assume *ex absurdo* that it is not an edge of the induced DAG. This implies that $s \in P(t)$ and $s \notin \tilde{P}(t)$. Since in a forest, a node has at most one parent, we know that $P(t) = \{s\}$. Therefore, because $\tilde{P}(t) \subseteq P(t)$ and $s \notin \tilde{P}(t)$, we find that $\tilde{P}(t) = \emptyset$. Due to Corollary 96, this implies that $P(t) = \emptyset$, a contradiction. □

Not every DAG is a forest; see for example Figure 7.9_{240}. However, interestingly, if a DAG is recursively decomposable, the corresponding induced DAG will always be a forest.

Proposition 98. *Consider a recursively decomposable DAG. In the corresponding induced DAG, every node $s \in G$ then has at most one parent: $|\tilde{P}(s)| \leq 1$.*

Proof of Proposition 98. Assume *ex absurdo* that there is some node $s \in G$ such that $|\tilde{P}(s)| \geq 2$. This means that we can consider $t_1, t_2 \in \tilde{P}(s)$ such that $t_1 \neq t_2$. Since $\tilde{P}(s) \subseteq P(s)$, we find that $t_1, t_2 \in P(s)$ and it therefore follows that $s \in D(t_1)$ and $s \in D(t_2)$, which implies that $D(t_1) \cap D(t_2) \neq \emptyset$. Since we are considering a recursively decomposable DAG, this implies that t_1 and t_2 are not incomparable. Therefore, because $t_1 \neq t_2$, we can assume without loss of generality that $t_2 \in D(t_1)$. Since we also know that $t_2 \in P(s)$, this implies that $t_2 \in P_{D(t_1)}(s)$. However, since it follows from $t_1 \in \tilde{P}(s)$ that $s \in \tilde{C}(t_1)$, we also find that $P_{D(t_1)}(s) = \emptyset$, a contradiction. □

This works in the other direction as well, leading to the following alternative characterisation of recursively decomposable DAGs.

Proposition 99. *A DAG is recursively decomposable if and only if the corresponding induced DAG is a forest.*

Proof of Proposition 99. If a DAG is recursively decomposable, it follows from Proposition 98 that the corresponding induced DAG is a forest.

So, assume that the induced DAG is a forest and consider any $s, t \in G$ that are incomparable with respect to the original DAG. Since the edges of the induced DAG are a subset of the edges of the original DAG, it follows that s and t are also incomparable with respect to the induced DAG. Since the induced DAG is a forest, this implies that $\tilde{D}(s) \cap \tilde{D}(t) = \emptyset$. It then follows from Proposition 94_{241} that $D(s) \cap D(t) = \emptyset$. □

By combining this result with Corollary $96_{\frown}$, it follows that a DAG with a single root node is recursively decomposable if and only if the induced DAG is a tree. The examples in Figure 7.1_{217} and Figures 7.5_{238}–7.9_{240} are all of this type.

Next, for any $s \in G$, we define $K_s := D(s) \cup \{s\}$ as the union of s and its descendants $D(s)$. Table $7.4_{\frown}$ provides a number of examples, for nodes in various decomposable DAGs. For any DAG, $D(s)$ and K_s are always closed subsets of G.

Proposition 100. *For any $s \in G$, $D(s)$ and K_s are closed subsets of G.*

	s	$D(s)$	K_s	$P(K_s)$
Figure 7.5_{238}	s_5	$s_6, s_7, s_{12}, s_{13}, s_{14}$	$s_5, s_6, s_7, s_{12}, s_{13}, s_{14}$	s_4
Figure 7.1_{217}	s_3	s_4, s_5, s_6, s_7	$s_3 s_4, s_5, s_6, s_7$	s_2
Figure 7.6_{239}	s_4	s_5, s_6, s_7	s_4, s_5, s_6, s_7	s_2, s_3
Figure 7.7_{239}	s_5	$s_6, s_7, s_{12}, s_{13}, s_{14}$	$s_5, s_6, s_7, s_{12}, s_{13}, s_{14}$	s_3, s_4
Figure 7.8_{239}	s_8	$s_9, s_{10}, s_{11}, s_{12}$	$s_8, s_9, s_{10}, s_{11}, s_{12}$	s_4, s_7
Figure 7.9_{240}	s_3	$s_4, \ldots, s_{13}$	$s_3, \ldots, s_{13}$	s_1, s_2

Table 7.4: Examples of K_s and $P(K_s)$, for various recursively decomposable DAGs

Proof of Proposition 100$_{\frown}$. First consider any $u, v \in D(s)$ and $k \in G$ such that $u \sqsubseteq k \sqsubseteq v$. Since $u \in D(s)$, we know that $s \sqsubset u$ and therefore, we find that $s \sqsubset k$, which implies that $k \in D(s)$.

Next, consider any $u, v \in K_s$ and $k \in G$ such that $u \sqsubseteq k \sqsubseteq v$. Since $u \in K_s$, we know that $s \sqsubseteq u$ and therefore, we find that $s \sqsubseteq k$, which implies that $k \in K_s$. □

For recursively decomposable DAGs, the sets $D(s)$ and K_s satisfy a number of additional properties. The following result establishes that every recursively decomposable DAG consists of disconnected sub-DAGs, each of which corresponds to a set K_s, with s a root node of the original DAG.

Proposition 101. *Consider a recursively decomposable DAG. Then the sets K_s, $s \in Ro(G)$, constitute a partition of G and, for all $s \in Ro(G)$, K_s is ancestral—$P(K_s) = \emptyset$.*

Proof of Proposition 101. Since the induced DAG is a forest [see Proposition 99$_{\frown}$] that has the same roots as the original DAG [see Proposition 98$_{\frown}$] we know that the sets $\tilde{D}(s) \cup \{s\}$, $s \in Ro(G)$, constitute a partition of G. Due to Proposition 94_{241}, this implies that the sets K_s, $s \in Ro(G)$, constitute a partition of G.

Consider now any $s \in Ro(G)$ and assume *ex absurdo* that $P(K_s) \neq \emptyset$. This implies that there is some $u \in K_s$ and $v \in P(u)$ such that $v \notin K_s$. Therefore, it follows from the first part of this proof that there is some $s' \in Ro(G)$ such that $s' \neq s$ and $v \in K_{s'}$, since $v \in P(u)$, this implies that $u \in K_{s'}$. Since we already know that $u \in K_s$, the first part of this proof leads us to a contradiction. □

For any $s \in G$, the sub-DAG that correspond to $D(s)$ can be decomposed similarly.

Corollary 102. *Consider a recursively decomposable DAG with nodes G. Then for any node $s \in G$, the sets K_c, $c \in \tilde{C}(s)$, constitute a partition of $D(s)$ and, for all $c \in \tilde{C}(s)$, $P_{D(s)}(K_c) = \emptyset$.*

Proof of Corollary 102. Since the induced DAG is a forest [see Proposition 99_{243}] we know that the sets $\tilde{D}(c) \cup \{c\}$, $c \in \tilde{C}(s)$ constitute a partition of $\tilde{D}(s)$. Due to Proposition 94_{241}, this implies that the sets K_c, $c \in \tilde{C}(s)$, constitute a partition of $D(s)$.

Consider now any $c \in \tilde{C}(s)$ and assume *ex absurdo* that $P_{D(s)}(K_c) \neq \emptyset$. This implies that there is some $u \in K_c$ and $v \in D(s)$ such that $v \in P(u)$ and $v \notin K_c$. Since $v \in D(s)$ and $v \notin K_c$, it follows from the first part of this proof that there is some $c' \in \tilde{C}(s)$ such that $c' \neq c$ and $v \in K_{c'}$, since $v \in P(u)$, this implies that $u \in K_{c'}$. Since we already know that $u \in K_c$, the first part of this proof leads us to a contradiction. □

The inference algorithms for recursively decomposable networks that we will develop further on in this chapter make extensive use of the concepts we have just introduced. In particular, for every node $s \in G$, they require the sets $\tilde{C}(s)$, $\tilde{P}(s)$ and $P(K_s)$. If these sets are not available, as will usually be the case, they first need to be constructed, preferably in some automated way. Therefore, before moving on, we briefly explain how to construct these sets efficiently.

For any $s \in G$, we know from Proposition 93_{239} that $\tilde{C}(s)$ consists of those nodes $t \in C(s)$ for which the only directed path from s to t is a single edge $s \to t$. Therefore, in order to find those nodes t efficiently, all we need to do is to start from s and move along the paths of the DAG following the direction of the edges—for example by means of a depth-first search—until we reach the leaves of the DAG and, for every node, keep track of how often we have visited it by raising a counter. Every $t \in C(s)$ will clearly be visited at least once. If a node $t \in C(s)$ is visited more than once, it means that there is more than one directed path from s to t, and t is therefore not an element of $\tilde{C}(s)$. If a node $t \in C(s)$ is visited only once, it must be through the trivial path $s \to t$ and it therefore follows that $t \in \tilde{C}(s)$.

After constructing $\tilde{C}(s)$, its elements need to be stored as attributes of the node s and, for all $t \in \tilde{C}(s)$, s needs to be added as an element of the set $\tilde{P}(t)$, which is initialised as empty at the beginning of this procedure, before processing any of the nodes s. After doing this for every $s \in G$, all the elements of $\tilde{P}(t)$ will have been added, for all $t \in G$. The sets $\tilde{C}(s)$ and $\tilde{P}(t)$ are now available for further use by inference algorithms.

This procedure is not restricted to recursively decomposable networks; it works for any DAG. In fact, it even allows us to find out whether a DAG is recursively decomposable. All we need to check is whether the induced DAG is a forest, or equivalently, whether $|\tilde{P}(t)| \leq 1$ for all $t \in G$. As we know from Proposition 99_{243}, this will be so if and only if the original DAG is recursively decomposable.

Finally, for any $s \in G$, the set $P(K_s)$ can also be constructed by starting from s and moving along the paths of the DAG following the direction of the edges until we reach its leaves. We need to do this twice. The first time, we simply mark the nodes we have visited; these are the elements of K_s. The second time, for every node we visit—every element of K_s—we check each of its parent nodes and store the ones that do not belong to K_s—that have not been marked during the first pass. The nodes that are stored during this second pass are the elements of $P(K_s)$.

7.5.2 *Recursively decomposable credal networks*

By adding local imprecise-probabilistic models to the nodes of a recursively decomposable DAG, we obtain a *recursively decomposable credal network*. We call them—and the corresponding DAGs—recursively decomposable because, as we are about to show, for such a credal network (the unconditional part of) its irrelevant natural extension can be decomposed into (the unconditional parts of) the irrelevant natural extensions of its subnetworks, recursively, and eventually even into its local models.

The first step in this recursive decomposition is to decompose the credal network into its disconnected subnetworks. Since we know from Proposition 101_{244} that the sets K_s, $s \in Ro(G)$, form a partition of G such that, for all $s \in Ro(G)$, $P(K_s) = \emptyset$, it follows directly from Proposition 67_{173} that

$$\underline{P}^{\mathrm{irr}}_{G}(\cdot) = \otimes_{s \in Ro(G)} \underline{P}^{\mathrm{irr}}_{K_s}(\cdot). \tag{7.20}$$

If the DAG has only a single root node s—if $Ro(G) = \{s\}$—this step is trivial; we then find that $\underline{P}^{\mathrm{irr}}_{G}(\cdot) = \underline{P}^{\mathrm{irr}}_{K_s}(\cdot)$. In fact, in this case, we even have that $\underline{P}^{\mathrm{irr}}_{G}(\cdot \rfloor \cdot) = \underline{P}^{\mathrm{irr}}_{K_s}(\cdot \rfloor \cdot)$.

The next step consists in further decomposing $\underline{P}^{\mathrm{irr}}_{K_s}(\cdot)$, for $s \in Ro(G)$. Since $P(K_s) = \emptyset$, this is a special case of the more general task of decomposing the irrelevant natural extension $\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot)$, for $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$. We therefore solve that problem first.

The more general task is trivial if s is a leaf of the network—if $D(s) = \emptyset$. In that case, it follows from Corollary 44_{141} that

$$\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot) = \underline{P}_{s \rfloor x_{P(s)}}(\cdot) \tag{7.21}$$

because $K_s = \{s\}$ and $P(K_s) = P(s)$, where $\underline{P}_{s \rfloor x_{P(s)}}(\cdot)$ is the local model that is attached to the node s. In all other cases, the following two important results allow us to decompose $\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot)$ recursively.

Proposition 103. *Consider a—not necessarily recursively decomposable—credal network and a node $s \in G$ such that $D(s) \neq \emptyset$. Then $P(s)$ and*

$P(D(s)) \setminus \{s\}$ are subsets of $P(K_s)$, s is an element of $P(D(s))$ and, for all $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$:

$$\underline{P}^{\text{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot) = \underline{P}_{s \rfloor x_{P(s)}}(\cdot) \oslash \underline{P}^{\text{irr}}_{D(s) \rfloor (x_{P(D(s)) \setminus \{s\}}, X_s)}(\cdot)$$

Proof of Proposition 103$_{\curvearrowleft\cap}$. The inclusions that are mentioned in this proposition are trivial. Now let $K := K_s = D(s) \cup \{s\}$, $T := \{s\}$ and $S := D(s)$. Then $P(T) = P(s)$, $P_K(T) = P_K(s) = P(s) \cap K = P(s) \cap K_s = \emptyset$ and $P_K(S) = \{s\}$. Furthermore, for all $t \in T$, we have that $t = s$ and therefore, since $D(s) = D_{K_s}(s)$, that $S \subseteq D_K(t)$. Since K is also a closed set [see Proposition 100$_{243}$] the result now follows from Corollaries 72$_{180}$ and 44$_{141}$. □

Proposition 104. *Consider a recursively decomposable credal network and a node $s \in G$ such that $D(s) \neq \emptyset$. Then, for all $c \in \tilde{C}(s)$, $P(K_c)$ is a subset of $P(D(s))$ that contains s and, for all $x_{P(D(s))} \in \mathscr{X}_{P(D(s))}$:*

$$\underline{P}^{\text{irr}}_{D(s) \rfloor x_{P(D(s))}}(\cdot) = \otimes_{c \in \tilde{C}(s)} \underline{P}^{\text{irr}}_{K_c \rfloor x_{P(K_c)}}(\cdot)$$

Proof of Proposition 104. Let $K := D(s)$. Due to Corollary 102$_{245}$, we know that the sets K_c, $c \in \tilde{C}(s)$, form a partition of K and that, for all $c \in \tilde{C}(s)$, $P_K(K_c) = \emptyset$. Since we know from Proposition 100$_{243}$ that K is a closed set, the result now follows directly from Corollary 68$_{174}$. □

Equipped with these new tools, we can now go back to the original problem: decomposing $\underline{P}^{\text{irr}}_{K_s}(\cdot)$, for some $s \in Ro(G)$. If s is a leaf of the network, then $\underline{P}^{\text{irr}}_{K_s}(\cdot) = \underline{P}_s(\cdot)$. Otherwise, since $P(K_s) = \emptyset$, it follows from Proposition 103$_{\curvearrowleft\cap}$ that

$$\underline{P}^{\text{irr}}_{K_s}(\cdot) = \underline{P}_s(\cdot) \oslash \underline{P}^{\text{irr}}_{D(s) \rfloor X_s}(\cdot) \tag{7.22}$$

The local model $\underline{P}_s(\cdot)$ is given. $\underline{P}^{\text{irr}}_{D(s) \rfloor X_s}(\cdot)$ can be decomposed further by applying Proposition 104. Since $P(K_c) = \{s\}$ for all $c \in \tilde{C}(s)$,[16] we find that for all $x_s \in \mathscr{X}_s$:

$$\underline{P}^{\text{irr}}_{D(s) \rfloor x_s}(\cdot) = \otimes_{c \in \tilde{C}(s)} \underline{P}^{\text{irr}}_{K_c \rfloor x_s}(\cdot).$$

We are now again faced with the very same general problem: decomposing $\underline{P}^{\text{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot)$ into smaller models—in this case, with $s = c$ and $P(K_s) = \{s\}$. The solution is identical. If s is a leaf, then $\underline{P}^{\text{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot) = \underline{P}_{s \rfloor x_{P(s)}}(\cdot)$ because $K_s = s$ and $P(K_s) = P(s)$. Otherwise, $\underline{P}^{\text{irr}}_{K_s \rfloor x_{P(K_s)}}(\cdot)$ can be decomposed into

[16] We know from Proposition 104 that $s \in P(K_c) \subseteq P(D(s))$. Furthermore, since $P(K_s) = \emptyset$, Proposition 103$_{\curvearrowleft\cap}$ tells us that $P(D(s)) \setminus \{s\} = \emptyset$. Hence, indeed: $P(K_c) = \{s\}$.

smaller models of the form $\underline{P}^{\text{irr}}_{\underline{K_c}\rfloor x_{P(K_c)}}(\cdot)$, with $c \in \tilde{C}(s)$, by applying Propositions 103_{246} and $104_{\curvearrowleft}$.

By continuing in this way, we eventually reach all the leaves of the network, at which point $\underline{P}^{\text{irr}}_{G}(\cdot)$ has been fully decomposed into the local models of the network. During this recursive decomposition, we have followed the edges of the induced forest: for every node s, we have considered its children $\tilde{C}(s)$ in the induced forest and used these to decompose the model. Furthermore, since the root nodes and leaves of the induced DAG are identical to those of the original DAG [see Corollary 96_{242}], what we have effectively done is start from the roots of the trees that make up the induced forest and follow their edges up to their leaves.

By reversing this process, this recursive decomposition of $\underline{P}^{\text{irr}}_{G}(\cdot)$ turns into a construction of $\underline{P}^{\text{irr}}_{G}(\cdot)$. Starting from the local models at the leaves of the induced forest, we work our way down to the roots. For every leaf s, $\underline{P}^{\text{irr}}_{\underline{K_s}\rfloor x_{P(K_s)}}(\cdot) = \underline{P}^{\text{irr}}_{\underline{s}\rfloor x_{P(s)}}(\cdot)$ is equal to the local model $\underline{P}_{\underline{s}\rfloor x_{P(s)}}(\cdot)$. For every node $s \in G$ that is not a leaf, $\underline{P}^{\text{irr}}_{\underline{K_s}\rfloor x_{P(K_s)}}(\cdot)$ is constructed from the models $\underline{P}^{\text{irr}}_{\underline{K_c}\rfloor x_{P(K_c)}}(\cdot)$, with $c \in \tilde{C}(s)$, using Propositions 103_{246} and $104_{\curvearrowleft}$. Finally, we use Equation 7.20_{246} to construct $\underline{P}^{\text{irr}}_{G}(\cdot)$ from the models $\underline{P}^{\text{irr}}_{K_s}(\cdot)$, with $s \in Ro(G)$.

Of course, this construction of $\underline{P}^{\text{irr}}_{G}(\cdot)$ is only symbolic. So far, we have not computed any of the marginal extensions and independent natural extensions that appear in these recursive expressions. However, as we will illustrate in the next sections, for specific types inferences, we can turn these symbolic recursive expressions into an actual recursive algorithm that is able to compute our inference of choice efficiently.

A similar recursive construction of $\underline{P}^{\text{irr}}_{G}(\cdot)$ has already been described by De Cooman et al. [42], for credal networks under epistemic irrelevance whose graphical structure is a tree, and under the simplifying assumption that the local upper probabilities are strictly positive. In our case, for trees, the induced DAG coincides with the original one [see Corollary 97_{242}] and, for all $s \in G$: $P(K_s) = P(s)$—with $|P(s)| = 1$—$\tilde{C}(s) = C(s)$, $P(D(s)) = \{s\}$ and, for all $c \in \tilde{C}(s) = C(s)$, $P(K_c) = \{s\}$. Therefore, if s is not a leaf, we find that

$$(\forall x_{P(s)} \in \mathscr{X}_{P(s)})\ \ \underline{P}^{\text{irr}}_{\underline{K_s}\rfloor x_{P(s)}}(\cdot) = \underline{P}_{\underline{s}\rfloor x_{P(s)}}(\cdot) \oslash \underline{P}^{\text{irr}}_{\underline{D(s)}\rfloor X_s}(\cdot)$$

and

$$(\forall x_s \in \mathscr{X}_s)\ \ \underline{P}^{\text{irr}}_{\underline{D(s)}\rfloor x_s}(\cdot) = \otimes_{c \in C(s)} \underline{P}^{\text{irr}}_{\underline{K_c}\rfloor x_s}(\cdot),$$

which are exactly the recursive expressions used in Reference [42]. Our construction can therefore be regarded as a generalisation of the one in Reference [42]; we extend it from trees to recursively decomposable networks and drop the local positivity assumptions.

7.5.3 *Sums of univariate gambles*

The reason why recursively decomposable networks are interesting is because the symbolic recursive construction of $\underline{P}^{\mathrm{irr}}_G(\cdot)$ can be used to compute $\underline{P}^{\mathrm{irr}}_G(f)$ recursively—and therefore efficiently—for various types of gambles $f \in \mathscr{G}(\mathscr{X}_G)$. A first class of gambles for which this is the case are gambles of the form $f = \sum_{i \in G} f_i$ where, for all $i \in G$, $f_i \in \mathscr{G}(\mathscr{X}_i)$ is a univariate gamble that only depends on the value of the variable X_i.

Computing $\underline{P}^{\mathrm{irr}}_G(\sum_{i \in G} f_i)$ is a matter of applying the symbolic recursive decomposition introduced in the previous section and combining it with the properties of marginal extension and independent natural extension discussed in Sections 6.6$_{171}$ and 6.7$_{175}$. In order to simplify the notation, we let $\psi_s := \sum_{i \in K_s} f_i$ for all $s \in G$.

The first step consists in combining Equations (7.20)$_{246}$ and (6.8)$_{173}$. Since it follows from Proposition 101$_{244}$ that $\sum_{i \in G} f_i = \sum_{s \in Ro(G)} \psi_s$, we find that

$$\underline{P}^{\mathrm{irr}}_G\left(\sum_{i \in G} f_i\right) = \underline{P}^{\mathrm{irr}}_G\left(\sum_{s \in Ro(G)} \psi_s\right) = \sum_{s \in Ro(G)} \underline{P}^{\mathrm{irr}}_{K_s}(\psi_s). \tag{7.23}$$

Our global optimisation problem is therefore already reduced to the smaller-sized problem of computing $\underline{P}^{\mathrm{irr}}_{K_s}(\psi_s)$, with $s \in Ro(G)$. Since $P(K_s) = \emptyset$ for these root nodes [see Proposition 101$_{244}$] this is a special case of a more general problem that, as we are about to show, can be solved recursively. For any $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, we will compute $\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\psi_s)$. We consider two cases.

If s is a leaf of the network—or, equivalently, if it is a leaf of the induced forest—then $K_s = \{s\}$ and $P(K_s) = P(s)$ and the problem becomes trivial. By applying Equation (7.21)$_{246}$, we find that

$$\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\psi_s) = \underline{P}_{s \rfloor x_{P(s)}}(f_s). \tag{7.24}$$

If s is not a leaf of the induced forest, then $D(s) \neq \emptyset$ and we can therefore apply Proposition 103$_{246}$. In combination with Equation (6.14)$_{178}$, this proposition implies that

$$\begin{aligned}
\underline{P}^{\mathrm{irr}}_{K_s \rfloor x_{P(K_s)}}(\psi_s) &= \underline{P}_{s \rfloor x_{P(s)}}\left(\underline{P}^{\mathrm{irr}}_{D(s) \rfloor (x_{P(D(s))\setminus\{s\}}, X_s)}(\psi_s)\right) \\
&= \underline{P}_{s \rfloor x_{P(s)}}\left(f_s + \underline{P}^{\mathrm{irr}}_{D(s) \rfloor (x_{P(D(s))\setminus\{s\}}, X_s)}\left(\sum_{i \in D(s)} f_i\right)\right) \\
&= \underline{P}_{s \rfloor x_{P(s)}}\left(f_s + \sum_{x_s \in \mathscr{X}_s} \mathbb{I}_{\{x_s\}} \underline{P}^{\mathrm{irr}}_{D(s) \rfloor x_{P(D(s))}}\left(\sum_{i \in D(s)} f_i\right)\right),
\end{aligned} \tag{7.25}$$

where, since Corollary 102$_{245}$ implies that $\sum_{i \in D(s)} f_i = \sum_{c \in \tilde{C}(s)} \psi_s$,

$$\underline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\sum_{i\in D(s)} f_i\right) = \underline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\sum_{c\in\tilde{C}(s)} \psi_c\right) = \sum_{c\in\tilde{C}(s)} \underline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\psi_c) \tag{7.26}$$

because of Proposition 104_{247} and Equation $(6.8)_{173}$. By combining Equations $(7.25)_{\curvearrowleft}$ and (7.26), it follows that, for any $s \in G$ that is not a leaf, and for any $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$:

$$\underline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\psi_s) = \underline{P}_{s\rfloor x_{P(s)}}\left(f_s + \sum_{x_s\in\mathscr{X}_s} \mathbb{I}_{\{x_s\}} \sum_{c\in\tilde{C}(s)} \underline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\psi_c)\right). \tag{7.27}$$

For all $s \in Ro(G)$, the above equations allow us to compute $\underline{P}^{\mathrm{irr}}_{K_s}(\psi_s)$ recursively, in the following simple manner: we start by applying Equation $(7.24)_{\curvearrowleft}$ in the leaves of the induced forest and then work our way down towards its roots by means of Equation (7.27). Once we have computed $\underline{P}^{\mathrm{irr}}_{K_s}(\psi_s)$ for all $s \in Ro(G)$, a final application of Equation $(7.23)_{\curvearrowleft}$ provides us with the value of $\underline{P}^{\mathrm{irr}}_{G}(\sum_{i\in G} f_i)$ that we are after.

The value of $\overline{P}^{\mathrm{irr}}_{G}(\sum_{i\in G} f_i)$ can be computed similarly. It suffices to replace the lower previsions in the recursion equations above by the corresponding upper previsions. However, developing such an approach is unnecessary because the task of computing $\overline{P}^{\mathrm{irr}}_{G}(\sum_{i\in G} f_i)$ can also be regarded as a special case of computing $\underline{P}^{\mathrm{irr}}_{G}(\sum_{i\in G} f_i)$. All we need to do is let $f'_i := -f_i$ for all $i \in G$ and use the above procedure to compute $\underline{P}^{\mathrm{irr}}_{G}(\sum_{i\in G} f'_i)$. It then follows from conjugacy that

$$\overline{P}^{\mathrm{irr}}_{G}\left(\sum_{i\in G} f_i\right) = -\underline{P}^{\mathrm{irr}}_{G}\left(-\sum_{i\in G} f_i\right) = -\underline{P}^{\mathrm{irr}}_{G}\left(\sum_{i\in G} f'_i\right).$$

The computational complexity of these recursive algorithms depends on the topology of the network. For every $s \in G$, we need to compute $\underline{P}^{\mathrm{irr}}_{K_s\rfloor X_{P(K_s)}}(\psi_s)$, which is a function on $\mathscr{X}_{P(K_s)}$ that is defined by

$$\left(\underline{P}^{\mathrm{irr}}_{K_s\rfloor X_{P(K_s)}}(\psi_s)\right)(x_{P(K_s)}) := \underline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\psi_s) \text{ for all } x_{P(K_s)} \in \mathscr{X}_{P(K_s)}.$$

Provided that the local models can be evaluated in constant time[17] and that $|\tilde{C}(s)|$ is small enough[18] it follows from Equations (7.27) and $(7.24)_{\curvearrowleft}$ that for a fixed $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, the value of this function can be computed in constant time because the recursive character of our algorithm implies that for every $c \in \tilde{C}(s)$—for every child of s in the induced forest—the function $\underline{P}^{\mathrm{irr}}_{K_c\rfloor X_{P(K_c)}}(\psi_c)$ has already been computed earlier on in the algorithm.

[17] If not, there is no point in devising a global algorithm anyway.

[18] Since $\tilde{C}(s) \subseteq C(s)$, this assumption is easily satisfied. Table $7.5_{\curvearrowleft}$ provides maximum values of $|\tilde{C}(s)|$ for different types of recursively decomposable networks.

	$\max_{s\in G}\lvert P(K_s)\rvert$	$\max_{s\in G}\lvert\tilde{C}(s)\rvert$
forest	1	$\max_{s\in G}\lvert C(s)\rvert$
trees	1	$\max_{s\in G}\lvert C(s)\rvert$
Markov chains	1	1
hidden Markov models	1	2
Markov chains of order m	m	1
hidden Markov models of order m	m	2
Figure 7.8$_{239}$	2	2
Figure 7.9$_{240}$	2	3

Table 7.5: Parameters of different types of recursively decomposable DAGs

Therefore, for a single node $s \in G$, evaluating this function for every $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$ has a computational complexity that is exponential in the number of nodes in $P(K_s)$. Since evaluating Equation (7.23) takes a constant amount of time, this allows us to conclude that the computational complexity of the complete recursive algorithm is linear in the number of nodes and exponential in $\max_{s\in G}|P(K_s)|$. This last parameter is the determining factor; if it is small enough, the procedure that we have just described can compute $\underline{P}^{\mathrm{irr}}_G(\sum_{i\in G} f_i)$ efficiently. Table 7.5 provides examples of $\max_{s\in G}|P(K_s)|$ for different types of recursively decomposable networks; in our examples—Figure 7.1$_{217}$ and Figures 7.5$_{238}$–7.9$_{240}$—this parameter never exceeds two. If the network is a tree or a forest, $\max_{s\in G}|P(K_s)|$ will always be at most one[19] and therefore, in that case, the algorithm is linear in the number of nodes.

Although the class of gambles that are sums of univariate functions is of course fairly limited, it does already allow us to compute various inferences that are of practical interest. For example, in a decision making context, additive utility functions are of this form. It is also important to realise that the sum does not need to run over all the nodes of the network. We can easily consider functions of the form $\sum_{i\in S} f_i$, for some $S \subseteq G$, simply by choosing $f_i = 0$ for all $i \in G \setminus S$. In this way, for any $q \in G$ and $f \in \mathscr{G}(\mathscr{X}_q)$, the algorithm in this section allows us to compute the marginal lower prevision $\underline{P}^{\mathrm{irr}}_q(f) := \underline{P}^{\mathrm{irr}}_G(f)$; it suffices to choose $S = \{q\}$ and $f_q = f$. Other interesting examples occur in networks where the variables X_i, $i \in S \subseteq G$, have the same

[19] Because in a forest, for every $s \in G$, we have that $P(K_s) = P(s)$ and $|P(s)| \leq 1$.

state space $\mathscr{X}_i = \mathscr{X}$ and represent the same type of object. We can then let $f_i = f$, for $i \in S$, where f is some function on $\mathscr{X}$. For $f = \mathbb{I}_x$, with $x \in \mathscr{X}$, the resulting sum $\sum_{i \in S} f_i$ represents the number of occurrences of x. For example, in case of a stationary queue of length n (possibly of order $m \geq 2$), where each variable X_i, $i \in G = \{1, \ldots, n\}$, represents the number of customers in the queue at time i, with $\mathscr{X} = \{0, \ldots, k\}$, we can choose $S = G$ and let f be the indicator of zero to compute the lower—and by conjugacy the upper—prevision—expected value—of the number of time points in which the queue is empty. In that same example, letting f be equal to the identity function divided by n—$f(x) := x/n$ for all $x \in \mathscr{X} = \{0, \ldots, k\}$—allows us to compute the lower and upper prevision—expected value—of the average number of customers over time.

7.5.4 *Products of univariate gambles*

A second class of gambles $f \in \mathscr{G}(\mathscr{X}_G)$ for which $\underline{P}^{\mathrm{irr}}_G(f)$ and $\overline{P}^{\mathrm{irr}}_G(f)$ can be computed efficiently in a recursively decomposable credal network are products of univariate functions: gambles of the form $f = \prod_{i \in G} f_i$, with $(\forall i \in G)\ f_i \in \mathscr{G}(\mathscr{X}_i)$. However, only for specific types of such products.

The simplest case is when each of the univariate gambles is non-negative: $f_i \geq 0$ for all $i \in G$. Despite its simplicity, it already covers a number of important cases. For example, many multiplicative utility functions are of this form. Perhaps the most important special case is the one where, for all $i \in G$, $f_i = \mathbb{I}_{B_i}$, with $B_i \in \mathscr{P}_{\emptyset}(\mathscr{X}_i)$ an event that represents partial information about the value of X_i. The resulting product $\prod_{i \in G} \mathbb{I}_{B_i}$ is then the indicator of the event $\times_{i \in G} B_i$ and the lower and upper prevision of this indicator are the lower and upper probability of $\times_{i \in G} B_i$, respectively. This also includes events of the form x_E as a special case: it suffices to let $B_i = \{x_i\}$ if $i \in E$ and $B_i = \mathscr{X}_i$ otherwise. Since the case of non-negative functions f_i includes these important special cases, we think we are justified in treating it separately, especially since—as we will see—it leads to simplified expressions. More complicated cases will be discussed later.

So, for now, we assume that $f_i \geq 0$ for all $i \in G$. In this case, the value of $\underline{P}^{\mathrm{irr}}_G(\prod_{i \in G} f_i)$ can be computed by means of a recursive scheme that is very similar to the one we developed in the previous section. The main difference is that we now exploit the factorisation property of the independent natural extension rather than its external additivity. In order to simplify the notation, we let $\phi_s := \prod_{i \in K_s} f_i$ for all $s \in G$.

First of all, for all $s \in Ro(G)$, we have that $\phi_s \geq 0$ and therefore, because of coherence [C1$_{48}$], that $\underline{P}^{\mathrm{irr}}_{K_s}(\phi_s) \geq 0$. Therefore, and since it follows from Proposition 101$_{244}$ that $\prod_{i \in G} f_i = \prod_{s \in Ro(G)} \phi_s$, we can combine Equations (7.20)$_{246}$ and (6.9)$_{173}$ to find that

$$\underline{P}_G^{\mathrm{irr}}\left(\prod_{i\in G} f_i\right) = \underline{P}_G^{\mathrm{irr}}\left(\prod_{s\in Ro(G)} \phi_s\right) = \prod_{s\in Ro(G)} \underline{P}_{K_s}^{\mathrm{irr}}(\phi_s). \tag{7.28}$$

Hence, solving the problem is now reduced to computing $\underline{P}_{K_s}^{\mathrm{irr}}(\phi_s)$, for $s \in Ro(G)$. As before, this is a special case—with $P(K_s) = \emptyset$—of a more general problem that can be solved recursively. For any $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, we will compute the value of $\underline{P}_{K_s\rfloor x_{P(K_s)}}^{\mathrm{irr}}(\phi_s)$. We consider two cases.

If s is a leaf of the induced forest, the problem becomes trivial. Since $K_s = \{s\}$ and $P(K_s) = P(s)$, it follows from Equation (7.21)$_{246}$ that

$$\underline{P}_{K_s\rfloor x_{P(K_s)}}^{\mathrm{irr}}(\phi_s) = \underline{P}_{s\rfloor x_{P(s)}}(f_s). \tag{7.29}$$

If s is not a leaf of the induced forest, we can apply Proposition 103$_{246}$ and Equation (6.14)$_{178}$ to find that

$$\begin{aligned}\underline{P}_{K_s\rfloor x_{P(K_s)}}^{\mathrm{irr}}(\phi_s) &= \underline{P}_{s\rfloor x_{P(s)}}\left(\underline{P}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}^{\mathrm{irr}}(\phi_s)\right)\\ &= \underline{P}_{s\rfloor x_{P(s)}}\left(\sum_{x_s\in\mathscr{X}_s} \mathbb{I}_{\{x_s\}} f_s(x_s)\underline{P}_{D(s)\rfloor x_{P(D(s))}}^{\mathrm{irr}}\left(\prod_{i\in D(s)} f_i\right)\right),\end{aligned} \tag{7.30}$$

where, since Corollary 102$_{245}$ implies that $\prod_{i\in D(s)} f_i = \prod_{c\in\tilde{C}(s)} \phi_s$, it follows from Proposition 104$_{247}$ and Equation (6.9)$_{173}$ that

$$\underline{P}_{D(s)\rfloor x_{P(D(s))}}^{\mathrm{irr}}\left(\prod_{i\in D(s)} f_i\right) = \underline{P}_{D(s)\rfloor x_{P(D(s))}}^{\mathrm{irr}}\left(\prod_{c\in\tilde{C}(s)} \phi_c\right) = \prod_{c\in\tilde{C}(s)} \underline{P}_{K_c\rfloor x_{P(K_c)}}^{\mathrm{irr}}(\phi_c) \tag{7.31}$$

because, for all $c \in \tilde{C}(s)$, $\phi_c \geq 0$ and therefore, due to coherence [C1$_{48}$], also $\underline{P}_{K_c\rfloor z_{P(K_c)}}^{\mathrm{irr}}(\phi_c) \geq 0$. By combining Equations (7.30) and (7.31), we find that

$$\underline{P}_{K_s\rfloor x_{P(K_s)}}^{\mathrm{irr}}(\phi_s) = \underline{P}_{s\rfloor x_{P(s)}}\left(\sum_{x_s\in\mathscr{X}_s} \mathbb{I}_{\{x_s\}} f_s(x_s) \prod_{c\in\tilde{C}(s)} \underline{P}_{K_c\rfloor x_{P(K_c)}}^{\mathrm{irr}}(\phi_c)\right) \tag{7.32}$$

for any $s \in G$ that is not a leaf and any $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$.

Using these equations, $\underline{P}_G^{\mathrm{irr}}(\prod_{i\in G} f_i)$ can easily be computed recursively. We start by applying Equation (7.29) at the leaves of the induced forest and then move along its branches towards its roots by means of Equation (7.32). A final application of Equation (7.28) provides us with the value of $\underline{P}_G^{\mathrm{irr}}(\prod_{i\in G} f_i)$.

The value of $\overline{P}_G^{\mathrm{irr}}(\prod_{i\in G} f_i)$ can be computed similarly. It suffices to replace the lower previsions in the above expressions by the corresponding upper ones. However, unlike in the case of sums of univariate gambles [see Section 7.5.3$_{249}$] this problem cannot be regarded as a special case of computing the lower prevision. Although conjugacy implies that $\overline{P}_G^{\mathrm{irr}}(\prod_{i\in G} f_i) = -\underline{P}_G^{\mathrm{irr}}(-\prod_{i\in G} f_i)$, we cannot apply the above procedure to

compute $\underline{P}^{\mathrm{irr}}_G(-\prod_{i\in G} f_i)$ because the assumption that all the univariate gambles in the product should be non-negative fails. Therefore, let us briefly illustrate how it can be proved that computing $\overline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$ indeed amounts to replacing the lower previsions in the above expressions by the corresponding upper ones.

In order to prove the upper prevision version of Equation $(7.28)_{\frown}$, the starting point is the conjugacy relation $\overline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i) = -\underline{P}^{\mathrm{irr}}_G(-\prod_{i\in G} f_i)$. The minus sign within the argument of $\underline{P}^{\mathrm{irr}}_G$ can then be dealt with by considering some arbitrary $t \in Ro(G)$ and assigning the minus sign to the partial product $\phi_t = \prod_{i\in K_t} f_i$. Because coherence $[\mathrm{C6}_{48}]$ guarantees that $\underline{P}^{\mathrm{irr}}_{K_t}(-\phi_t) \leq 0$, the combination of Equations $(7.20)_{246}$ and $(6.9)_{173}$ then allows us to infer that

$$\begin{aligned}\overline{P}^{\mathrm{irr}}_G\left(\prod_{i\in G} f_i\right) &= -\underline{P}^{\mathrm{irr}}_G\left(-\prod_{s\in Ro(G)} \phi_s\right)\\ &= -\underline{P}^{\mathrm{irr}}_{K_t}(-\phi_t)\prod_{s\in Ro(G)\setminus\{t\}} \overline{P}^{\mathrm{irr}}_{K_s}(\phi_s) = \prod_{s\in Ro(G)} \overline{P}^{\mathrm{irr}}_{K_s}(\phi_s),\end{aligned}$$

where the final equality is again due to an application of conjugacy. We find that, indeed, as mentioned before, the only difference with Equation $(7.28)_{\frown}$ is that the lower previsions are replaced by the corresponding upper previsions. The upper prevision version of Equation $(7.31)_{\frown}$ can be proved similarly; we find that:

$$\overline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) = \prod_{c\in\tilde{C}(s)} \overline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c). \tag{7.33}$$

The remaining equations follow trivially from conjugacy and the argumentation for the original equation. For every $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, we find that

$$\overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \overline{P}_{s\rfloor x_{P(s)}}(f_s). \tag{7.34}$$

if s is a leaf. If s is not a leaf, then

$$\overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \overline{P}_{s\rfloor x_{P(s)}}\left(\sum_{x_s\in\mathscr{X}_s} \mathbb{I}_{\{x_s\}} f_s(x_s) \prod_{c\in\tilde{C}(s)} \overline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c)\right). \tag{7.35}$$

These upper prevision versions of our recursion equations can also be regarded as special cases of—and can therefore be derived from—the more general expressions that we will discuss later in this section.[20]

Let us now try and weaken the requirement that each of the gambles f_i, $i \in G$, should be non-negative. We have introduced this assumption because it leads to simplified expressions. However, it is not necessary. As we will now show, some of the gambles f_i can be allowed to have negative values. The only

[20] See Equations $(7.37)_{256}$, $(7.40)_{257}$, $(7.41)_{257}$, $(7.43)_{257}$ and $(7.44)_{258}$.

requirement that we need to impose is that the nodes of the gambles for which this is the case should all be mutually comparable. In other words, if we call a subset T of G *comparable* if there are no two nodes $t, t' \in T$ such that $t \parallel t'$,[21] then what we need to impose is that $f_i \geq 0$ for those $i \in G$ that do not belong to T, for some comparable subset T of G. Before we go on to show that this condition is sufficient in order to develop a recursive algorithm, we first take a closer look at this condition itself and characterise it alternatively.

Consider the partial order that is imposed on the elements of G by the binary operator $\sqsubseteq$. The roots and leaves of the DAG are then the minimal and maximal elements of this partially ordered set, respectively. Using the standard terminology from partial orders, the condition above simply requires that T should be a chain: a totally ordered subset of G. Equivalently, since $|G|$ is finite, T is comparable if and only if it is a subset of a directed path. For any T that satisfies this condition, the following two properties hold.

Proposition 105. *Consider a recursively decomposable DAG and a non-empty comparable set T. Then there is a unique $t \in Ro(G)$ such that $K_t \cap T \neq \emptyset$.*

Proof of Proposition 105. Since the DAG is recursively decomposable, we know from Proposition 101_{244} that the sets K_s, $s \in Ro(G)$, form a partition of G. Therefore, since $T \neq \emptyset$, it follows that there is some $t \in Ro(G)$ such that $K_t \cap T \neq \emptyset$. Assume *ex absurdo* that there is some $t' \in Ro(G)$ such that $t' \neq t$ and $K_{t'} \cap T \neq \emptyset$. Since $K_t \cap T \neq \emptyset$, we can consider some $u \in K_t \cap T$. Since $K_{t'} \cap T \neq \emptyset$, we can consider some $v \in K_{t'} \cap T$. Since T is comparable, and because u and v are elements of T, it follows that u and v are comparable. Therefore, we may assume—without loss of generality—that $u \sqsubseteq v$. Since it follows from $u \in K_t$ that $t \sqsubseteq u$, this implies that $t \sqsubseteq v$, which in turn implies that $v \in K_t$. Since K_t and $K_{t'}$ are elements of a partition, this is a contradiction. □

Proposition 106. *Consider a recursively decomposable DAG and a comparable set T. Then for all $s \in G$ such that $D(s) \cap T \neq \emptyset$, there is a unique $t \in \tilde{C}(s)$ such that $K_t \cap T \neq \emptyset$.*

Proof of Proposition 106. Consider any $s \in G$ such that $D(s) \cap T \neq \emptyset$. Since the DAG is recursively decomposable, we know from Corollary 102_{245} that the sets K_c, $c \in \tilde{C}(s)$, form a partition of $D(s)$. Therefore, it follows that there is some $t \in \tilde{C}(s)$ such that $K_t \cap T \neq \emptyset$. Assume *ex absurdo* that there is some $t' \in \tilde{C}(s)$ such that $t' \neq t$ and $K_{t'} \cap T \neq \emptyset$. The same argument that we already used in the proof of Proposition 105 then leads to a contradiction. □

[21] This terminology is compatible with our definition for (in)comparability of nodes at the beginning of Section $7.5.1_{238}$: if T consists of two elements, then T is (in)comparable if and only if its two elements are.

We now consider a comparable set T and a product $\prod_{i\in G} f_i$ of univariate gambles $f_i \in \mathscr{G}(\mathscr{X}_i)$, $i \in G$, such that $f_i \geq 0$ for all $i \in G \setminus T$. As we will see, it is still possible to compute $\underline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$ and $\overline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$ recursively. The expressions that allow us to do so just become a bit more complicated.

We start by computing $\underline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$. If $T = \emptyset$, we have the special case discussed in the beginning of this section, and the recursive expressions for that special case can then be used. If $T \neq \emptyset$, we need some new techniques. In that case, the first step is to apply Proposition $105_{\frown}$, from which we can infer that there is a unique $t \in Ro(G)$ such that $K_t \cap T \neq \emptyset$. Therefore, for all $s \in Ro(G) \setminus \{t\}$, we have that $K_s \cap T = \emptyset$, which implies that $\phi_s \geq 0$. This allows us to combine Equations $(7.20)_{246}$ and $(6.9)_{173}$ to find that

$$\underline{P}^{\mathrm{irr}}_G\left(\prod_{i\in G} f_i\right) = \begin{cases} \underline{P}^{\mathrm{irr}}_{K_t}(\phi_t) \prod\limits_{s\in Ro(G)\setminus\{t\}} \underline{P}^{\mathrm{irr}}_{K_s}(\phi_s) & \text{if } \underline{P}^{\mathrm{irr}}_{K_t}(\phi_t) \geq 0 \\ \underline{P}^{\mathrm{irr}}_{K_t}(\phi_t) \prod\limits_{s\in Ro(G)\setminus\{t\}} \overline{P}^{\mathrm{irr}}_{K_s}(\phi_s) & \text{if } \underline{P}^{\mathrm{irr}}_{K_t}(\phi_t) \leq 0. \end{cases} \tag{7.36}$$

Since we do not know beforehand—when starting the recursion at the leaves of the induced network—what the sign of $\underline{P}^{\mathrm{irr}}_{K_t}(\phi_t)$ will be, it no longer suffices to compute the value of $\underline{P}^{\mathrm{irr}}_{K_s}(\phi_s)$ for all $s \in Ro(G)$. We also need the value of $\overline{P}^{\mathrm{irr}}_{K_s}(\phi_s)$, for all $s \in Ro(G) \setminus \{t\}$.

This problem is a special case—with $P(K_s) = \emptyset$—of the following more general problem, which can be solved recursively: for every $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, we will compute $\underline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s)$ and $\overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s)$. As before, we consider two cases.

This problem is trivial if s is a leaf of the induced forest. In that case, it follows from Equation $(7.21)_{246}$ and conjugacy that

$$\underline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \underline{P}_{s\rfloor x_{P(s)}}(f_s) \text{ and } \overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \overline{P}_{s\rfloor x_{P(s)}}(f_s). \tag{7.37}$$

If s is not a leaf of the induced forest, we can apply Proposition 103_{246} and Equation $(6.14)_{178}$ to find that

$$\underline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \underline{P}_{s\rfloor x_{P(s)}}\left(\underline{P}^{\mathrm{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\phi_s)\right) = \underline{P}_{s\rfloor x_{P(s)}}\left(\underline{h}_{s\rfloor x_{P(K_s)}}\right), \tag{7.38}$$

where $\underline{h}_{s\rfloor x_{P(K_s)}} := \underline{P}^{\mathrm{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\phi_s)$ is a gamble on $\mathscr{X}_s$, given for all $x_s \in \mathscr{X}_s$ by

$$\underline{h}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} f_s(z_s)\,\underline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod\limits_{i\in D(s)} f_i\right) & \text{if } f_s(x_s) \geq 0; \\ f_s(z_s)\,\overline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod\limits_{i\in D(s)} f_i\right) & \text{if } f_s(x_s) \leq 0. \end{cases} \tag{7.39}$$

Similarly, we also find that

$$\overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) = \overline{P}_{s\rfloor x_{P(s)}}\big(\overline{P}^{\mathrm{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\phi_s)\big) = \overline{P}_{s\rfloor x_{P(s)}}\big(\overline{h}_{s\rfloor x_{P(K_s)}}\big), \quad (7.40)$$

where $\overline{h}_{s\rfloor x_{P(K_s)}} := \underline{P}^{\mathrm{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\phi_s)$ is a gamble on $\mathscr{X}_s$, given for all $x_s \in \mathscr{X}_s$ by

$$\overline{h}_{s\rfloor x_{P(K_s)}}(x_s) := \begin{cases} f_s(z_s)\,\overline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } f_s(x_s) \geq 0; \\ f_s(z_s)\,\underline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } f_s(x_s) \leq 0. \end{cases} \quad (7.41)$$

The following equations allow us to evaluate the gambles $\underline{h}_{s\rfloor x_{P(K_s)}}$ and $\overline{h}_{s\rfloor x_{P(K_s)}}$ recursively. For all $x_{P(D(s))} \in \mathscr{X}_{P(D(s))}$, we have that

$$\underline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) = \begin{cases} \underline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \prod_{c\in\tilde{C}(s)\setminus\{t\}} \underline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \underline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \geq 0 \\ \underline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \prod_{c\in\tilde{C}(s)\setminus\{t\}} \overline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \underline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \leq 0 \end{cases} \quad (7.42)$$

and

$$\overline{P}^{\mathrm{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) = \begin{cases} \overline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \prod_{c\in\tilde{C}(s)\setminus\{t\}} \overline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \overline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \geq 0 \\ \overline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \prod_{c\in\tilde{C}(s)\setminus\{t\}} \underline{P}^{\mathrm{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \overline{P}^{\mathrm{irr}}_{K_t\rfloor x_{P(K_t)}}(\phi_t) \leq 0. \end{cases} \quad (7.43)$$

Both equations follow from Proposition 104_{247}, Equation $(6.9)_{173}$ and conjugacy. The node t should be an element of $\tilde{C}(s)$ such that $\phi_c \geq 0$ for all $c \in \tilde{C}(s) \setminus \{t\}$. If $D(s) \cap T = \emptyset$, then any $t \in \tilde{C}(s)$ has this property. Otherwise, Proposition 106_{255} guarantees the existence of a unique $t \in \tilde{C}(s)$ that has this property.

Using these equations, the value of $\underline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$ can be computed recursively, in more or less the same way as before. All we need to do is apply Equation $(7.37)_{\frown}$ in the leaves of the induced forest and then move along its branches towards its roots by means of Equations $(7.38)_{\frown}$–(7.43). A final application of Equation $(7.36)_{\frown}$ will provide us with the value of $\underline{P}^{\mathrm{irr}}_G(\prod_{i\in G} f_i)$.

The corresponding upper prevision $\overline{P}^{\text{irr}}_G(\prod_{i\in G} f_i)$ can be computed in exactly the same way. The only difference is the final step. If $T = \emptyset$, then $\phi_i \geq 0$ for all $i \in G$. In that case, we can apply the simplified methods that were discussed above. If $T \neq 0$, we can combine Equations (7.20)$_{246}$ and (6.9)$_{173}$ with conjugacy to find that

$$\overline{P}^{\text{irr}}_G\left(\prod_{i\in G} f_i\right) = \begin{cases} \overline{P}^{\text{irr}}_{K_t}(\phi_t) \prod_{s\in Ro(G)\setminus\{t\}} \overline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \overline{P}^{\text{irr}}_{K_t}(\phi_t) \geq 0; \\ \overline{P}^{\text{irr}}_{K_t}(\phi_t) \prod_{s\in Ro(G)\setminus\{t\}} \underline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \overline{P}^{\text{irr}}_{K_t}(\phi_t) \leq 0. \end{cases} \tag{7.44}$$

with t as in Equation (7.36)$_{256}$.

Although the expressions in this section are a bit more involved—especially the ones where we partially drop the assumption of non-negativity—the computational complexity of the resulting recursive algorithms is the same as that of the algorithm for sums of univariate gambles discussed in Section 7.5.3$_{249}$. For every $s \in G$, we now need to compute two real-valued functions on $\mathscr{X}_{P(K_s)}$: $\underline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)$ and $\overline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)$, defined by

$$\left(\underline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)\right)(x_{P(K_s)}) := \underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) \text{ for all } x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$$

and

$$\left(\overline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)\right)(x_{P(K_s)}) := \overline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s) \text{ for all } x_{P(K_s)} \in \mathscr{X}_{P(K_s)},$$

whereas before, we only needed to compute one such function. Clearly, such a factor of two does not influence the computational efficiency of the procedure. In fact, in the simplified cases discussed in the beginning of this section, a single such function will suffice. Evaluating one of these functions for some fixed $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$ can be done in constant time, using the recursive expressions provided above. Therefore, the complexity of computing $\underline{P}^{\text{irr}}_G(\prod_{i\in G} f_i)$ and/or $\overline{P}^{\text{irr}}_G(\prod_{i\in G} f_i)$ in this way is linear in the number of nodes $|G|$ and exponential in the parameter $\max_{s\in G}|P(K_s)|$, which implies that it is linear for trees and forests; see Section 7.5.3$_{249}$ for more information.

7.5.5 *A single query node with evidence*

Although the notion of a comparable set T might come across as a fairly abstract concept, it has a number of very simple intuitive cases. Rather amusingly, the most important example is trivial: every singleton $T = \{q\}$, with $q \in G$, is—trivially—a comparable set. A non-trivial example is the set that consists of all the hidden nodes in a hidden Markov model; see Section 7.5.6$_{264}$. For now, we focus on the trivial case. In that case, as we are about to explain, the algorithm in the previous section allows us to compute inferences about a single query variable X_q, conditional on the value $x_E \in \mathscr{X}_E$

of some arbitrary set of evidence variables X_E, with $E \subseteq G \setminus \{q\}$. This is the most import inference problem in credal networks; in fact, most of the existing inference algorithms for credal networks only consider this type of inferences.

For any $f \in \mathscr{G}(\mathscr{X}_q)$, as explained in Section 7.3_{219}, $\underline{P}_q^{\text{irr}}(f\rfloor x_E)$ and $\underline{R}_q^{\text{irr}}(f\rfloor x_E)$ can be approximated from below—and in most cases coincide with—$\underline{E}_q^{\text{irr}*}(f\rfloor x_E)$ and $\underline{R}_q^{\text{irr}*}(f\rfloor x_E)$, respectively. Furthermore, in order to compute $\underline{E}_q^{\text{irr}*}(f\rfloor x_E)$ or $\underline{R}_q^{\text{irr}*}(f\rfloor x_E)$, all that we need is some method for evaluating the real-valued function $\rho_{f,x_E}^{\text{irr}}$, defined by

$$\rho_{f,x_E}^{\text{irr}}(\mu) := \underline{P}_G^{\text{irr}}(\mathbb{I}_{\{x_E\}}[f-\mu]) \text{ for all } \mu \in \mathbb{R}.$$

For some fixed $\mu \in \mathbb{R}$, we can use the algorithm in the previous section for this purpose. It suffices to let $T = \{q\}$ and define

$$f_i := \begin{cases} f-\mu & \text{if } i = q \\ \mathbb{I}_{\{x_i\}} & \text{if } i \in E \\ 1 & \text{otherwise} \end{cases} \qquad \text{for all } i \in G \tag{7.45}$$

because the product $\prod_{i\in G} f_i$ will then be equal to $\mathbb{I}_{\{x_E\}}[f-\mu]$. Evaluating $\rho_{f,x_E}^{\text{irr}}$ for some fixed μ therefore has a computational complexity that is linear in the number of nodes $|G|$ and exponential in the parameter $\max_{s\in G}|P(K_s)|$. As explained in sections $2.7.3_{63}$ and 7.3_{219}, computing $\underline{E}_q^{\text{irr}*}(f\rfloor x_E)$ or $\underline{R}_q^{\text{irr}*}(f\rfloor x_E)$ now amounts to applying the bisection method or some other root-finding procedure to find the highest—rightmost—root of $\rho_{f,x_E}^{\text{irr}}$. For some fixed floating point precision, the number of evaluations of the function $\rho_{f,x_E}^{\text{irr}}$ that is required to find this root is always finite and can furthermore be bounded uniformly from above. This allows us to conclude that the overall complexity of this method for computing $\underline{E}_q^{\text{irr}*}(f\rfloor x_E)$ or $\underline{R}_q^{\text{irr}*}(f\rfloor x_E)$ is linear in the number of nodes $|G|$ and exponential in the parameter $\max_{s\in G}|P(K_s)|$. By conjugacy, $\overline{E}_q^{\text{irr}*}(f\rfloor x_E)$ and $\overline{R}_q^{\text{irr}*}(f\rfloor x_E)$ can be computed similarly.

If the network has a tree topology, this procedure reduces to the inference algorithm for credal trees introduced in Reference [42]. The parameter $\max_{s\in G}|P(K_s)|$ is then equal to one [see Section 7.5.3] and the computational complexity is therefore—in this special case—linear in the number of nodes in the network. Basically, what we have done in this section is to extend the algorithm in Reference [42] from trees to recursively decomposable networks. Reference [42] also discusses some simplifications that can speed up the computation; these apply in our more general case as well. For example, as can be seen from Equation (7.45), only one of the univariate gambles in $\prod_{i\in G} f_i$ depends on μ: the function f_q. As a result, for all $s \in G$ such that $q \notin K_s$, the function ϕ_s will also not depend on μ, which implies that for these nodes s, the value of $\underline{P}_{K_s\rfloor x_{P(K_s)}}^{\text{irr}}(\phi_s)$ and $\overline{P}_{K_s\rfloor x_{P(K_s)}}^{\text{irr}}(\phi_s)$ can be reused while iterating over μ.

The other simplifications in Reference [42] are basically special cases of the general tricks in Section 7.2$_{216}$.

The same algorithm can also be used if we do not know the exact value of X_E but only have partial information about each of the variables X_i, $i \in E$, in the sense that we know that its value belongs to some set $B_i \in \mathscr{P}_\emptyset(\mathscr{X}_i)$. The only difference is that in that case, the univariate gambles f_i are given by

$$f_i := \begin{cases} f - \mu & \text{if } i = q \\ \mathbb{I}_{B_i} & \text{if } i \in E \\ 1 & \text{otherwise} \end{cases} \qquad \text{for all } i \in G.$$

The resulting product $\prod_{i \in G} f_i$ is then equal to $\mathbb{I}_O[f - \mu]$, with $O = \times_{i \in E} B_i$.[22] Again, we can apply the algorithm that we introduced in Section 7.5.4 to compute $\rho^{\text{irr}}_{f,O}(\mu) := \underline{P}^{\text{irr}}_G(\prod_{i \in G} f_i) = \underline{P}^{\text{irr}}_G(\mathbb{I}_O[f - \mu])$ and then apply a root-finding procedure to compute $\underline{E}^{\text{irr}*}_q(f \rfloor O)$ or $\underline{R}^{\text{irr}*}_q(f \rfloor O)$ and, by conjugacy, the corresponding upper previsions.

The algorithms in this section can be used to compute conditional inferences about a single query node in any recursively decomposable credal network. This includes many practically relevant instances. In order to illustrate this, we end this section by briefly discussing some examples.

A first important example is filtering and smoothing in imprecise Hidden Markov models, that is, inference about the last or some intermediate hidden variable, respectively, conditional on evidence about the observed variables. In Figure 7.5$_{238}$, filtering corresponds to letting $q = s_7$ whereas choosing $q = s_4$ corresponds to smoothing; in both cases, E is equal to $\{s_8, \dots, s_{14}\}$. Filtering and smoothing are two important tasks in classical precise hidden Markov models; they are equally important in imprecise hidden Markov models and can be used to tackle the same kinds of practical problems, but more robustly. This example is a bit unfair because—since a hidden Markov model is a special kind of tree—filtering and smoothing can also be dealt with by the algorithm in Reference [42]; our generalisation to recursively decomposable networks is not necessary here. However, this is no longer the case if we consider a hidden Markov model of order $m \geq 2$—Figure 7.6$_{239}$ provides an example with $m = 2$. In that case, our algorithm can still perform the task of filtering and smoothing, whereas the algorithm in Reference [42] no longer applies.

The graphical structure of some well-known classifiers are also decomposable. The simplest example is the naive credal classifier [115], which is the imprecise version of the naive Bayes classifier. Figure 7.10$_{\frown}$ provides an example of a DAG that corresponds to such a naive—Bayes or credal—classifier. The variable X_t that is attached to the root node t is called the class variable. The other variables correspond to so-called features. The goal is to estimate the

[22] We assume that missing evidence is MAR; see Footnote 5$_{216}$.

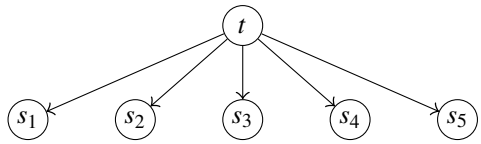


Figure 7.10: Example of a naive classifier

value of the class variable based on—possibly partial—evidence $O = \times_{i \in E} B_i$ about the feature variables, with $E = G \setminus \{t\}$ and, for all $i \in E$, $B_i \in \mathscr{P}_{\emptyset}(\mathscr{X}_i)$. In a naive Bayes classifier, this is done by choosing the class value that has the highest posterior probability conditional on the evidence. In a naive credal classifier, imprecise-probabilistic decision criteria are used instead; see Reference [95] for an overview.

One of the criteria that was used in Reference [115] is maximality.[23] The idea is to consider a partial order over the elements in $\mathscr{X}_t$. For two classes x_t and $\tilde{x}_t$ in $\mathscr{X}_t$, it is said that $\tilde{x}_t$ is better than x_t, denoted by $\tilde{x}_t \succ x_t$, if the probability of $\tilde{x}_t$ is higher than that of x_t for each of the probability mass functions in the updated credal set $\mathrm{marg}_s(\mathscr{F}_G^{\mathrm{irr}} \rfloor O)$, or equivalently, in terms of lower previsions, if $\underline{P}_t^{\mathrm{irr}}\big(\mathbb{I}_{\{\tilde{x}_t\}} - \mathbb{I}_{\{x_t\}} \big\rfloor O\big) > 0$. The classes that are returned by the naive credal classifier are those elements of $\mathscr{X}_t$ that are not dominated under this partial order, in the sense that no other class is better. Since a partial order may have more than one maximal element, this imprecise-probabilistic classifier can be indeterminate.[24] Although this might seem undesirable, it is in fact not because indeterminacy typically occurs in those instances—for those instantiations of the feature variables—where precise-probabilistic classifiers suffer from robustness issues, in the sense that their accuracy drops drastically [13]. Other partial orders can also be considered, resulting in classifiers with similar properties. For example, if we update by means of regular extension, $\tilde{x}_t$ is said to be better than x_t if $\underline{R}_t^{\mathrm{irr}}\big(\mathbb{I}_{\{\tilde{x}_t\}} - \mathbb{I}_{\{x_t\}} \big\rfloor O\big) > 0$. This is often equivalent because, for this kind of classifiers, the lower probability of the evidence O is usually positive, which implies that $\underline{P}_t^{\mathrm{irr}}\big(\mathbb{I}_{\{\tilde{x}_t\}} - \mathbb{I}_{\{x_t\}} \big\rfloor O\big)$ and $\underline{R}_t^{\mathrm{irr}}\big(\mathbb{I}_{\{\tilde{x}_t\}} - \mathbb{I}_{\{x_t\}} \big\rfloor O\big)$ are equal. For the sake of this discussion, let us assume that we use the latter expression.

What is important is that in order to perform this kind of classification, all we need to be able to do is compute $\underline{R}_q^{\mathrm{irr}}(f \rfloor O)$, with $q = t$, $f = \mathbb{I}_{\{\tilde{x}_t\}} - \mathbb{I}_{\{x_t\}}$ and $O = \times_{i \in E} B_i$. Unless the evidence O has upper probability zero, $\underline{R}_q^{\mathrm{irr}}(f \rfloor O)$ will be equal to $\underline{R}_q^{\mathrm{irr}*}(f \rfloor O)$ [see Section 7.3], which implies that we can use the algorithm in this section to perform this classification task. Similar statements

[23] Actually, the authors of Reference [95] speak of credal dominance. However, over the years, maximality seems to have become the preferred name for this criterion.

[24] Any classifier that has this feature of indeterminacy is called a credal classifier [115].

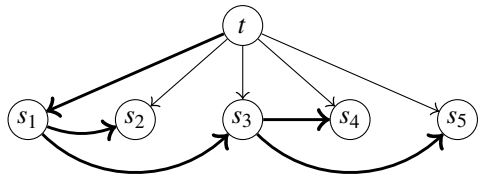


Figure 7.11: Example of a tree-augmented naive classifier

can be made for other decision criteria. For example, if we use interval dominance, $\tilde{x}_t$ is said to be better than x_t if $\underline{R}_t^{\text{irr}}(\tilde{x}_t \rfloor O) > \overline{R}_t^{\text{irr}}(\tilde{x}_t \rfloor O)$. The Γ-maximin and Γ-maximax criteria do not consider a partial order, they simply select the class that maximises $\underline{R}_t^{\text{irr}}(\tilde{x}_t \rfloor O)$ and $\overline{R}_t^{\text{irr}}(\tilde{x}_t \rfloor O)$. Again, the clue is that in order to perform this kind of classification, all we need to be able to do is compute $\underline{R}_q^{\text{irr}}(f \rfloor O)$—or the corresponding upper prevision—with f now equal to the indicator of $\tilde{x}_t$ or x_t. If the evidence O has positive upper probability, then $\underline{R}_q^{\text{irr}}(f \rfloor O)$ is equal to $\underline{R}_q^{\text{irr}*}(f \rfloor O)$ and we can use the algorithm in this section.

Of course, this example is again a bit unfair. First of all, since a naive credal classifier has a tree structure, the algorithm in Reference [42] is also capable of computing $\underline{R}_q^{\text{irr}*}(f \rfloor O)$; our extension to recursively decomposable networks provides no added value. Furthermore, due to the very simple structure of this specific network, applying a general purpose algorithm—even the specialized algorithm in Reference [42]—would be overkill. It is a much better idea to derive explicit expressions for $\underline{R}_q^{\text{irr}*}(f \rfloor O)$. We leave this as an exercise to the reader. For evidence of the form $O = \{x_E\}$, with $x_E \in \mathscr{X}_E$, Zaffalon provides such expressions in Reference [115].[25]

However, this is no longer the case once we move from the naive credal classifier to more involved types of credal classifiers. A first extension is the tree-augmented naive credal classifier, which is the imprecise version of the tree-augmented naive Bayesian classifier. The DAG that corresponds to this type of classifier is identical to that of a naive classifier, but with additional edges such that the features form a tree; see Figure 7.11 for an example. The goal is still to classify X_t, where t is the unique root of the network. As before, we can do this by computing inferences of the form $\underline{R}_q^{\text{irr}*}(f \rfloor O)$, with $q = t$ and $O = \times_{i \in E} B_i$. Since the DAG of a tree-augmented naive classifier is always decomposable but never a tree, our algorithm is capable of computing this type of inference, but the algorithm in Reference [42] is not. Furthermore, since the parameter $\max_{s \in G} |P(K_s)|$ is at most two for this type of DAGs—regardless of the number of nodes and the specific tree-augmented classifier that we are

[25]The naive credal classifier of Zaffalon [115] considered a credal network under strong independence instead of a credal network under epistemic irrelevance. However, as was proved in Reference [67, Corollary 6], in this specific case, the resulting inferences are identical.

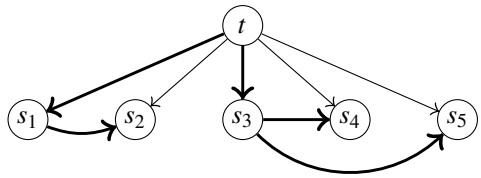


Figure 7.12: Example of a forest-augmented naive classifier

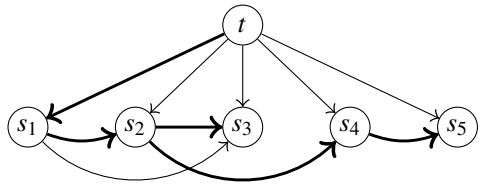


Figure 7.13: Example of an augmented naive classifier that is recursively decomposable but does not belong to any other specific type

considering—the computational complexity of our algorithm will be linear in the number of nodes. For this type of classifiers, there are still alternative methods available. Reference [116] presents an efficient classification algorithm for a tree-augmented credal classifier under strong independence instead of epistemic irrelevance; however, they only consider local models of a specific type—derived from probability intervals—and, for every $i \in E$, B_i should either be a singleton or equal to $\mathscr{X}_i$—every feature should either be known exactly or completely missing.

Clearly, our algorithm can also deal with classifiers whose graphical structure is more complicated than that of a tree-augmented naive classifier. In principle, we can consider any recursively decomposable network, as long as the parameter $\max_{s \in G} |P(K_s)|$ remains small enough. For example, we could consider the recursively decomposable network in Figure 7.9_{240}, with $t = s_4$ as the class node; notice that t is not the root node here. Figure 7.12 depicts an example of a forest-augmented naive classifier, which is a straightforward generalisation of a tree-augmented naive classifier. The DAG in Figure 7.13 is also a recursively decomposable augmented naive classifier, but not of any of the types that were mentioned before. To the best of our knowledge, for these more general types of recursively decomposable networks, there is no other efficient exact algorithm that can compute the inferences that are required to classify X_t.

And, we can even go further than this. If the evidence is of the form $O = \{x_E\}$, with $E = G \setminus \{t\}$ and $x_E \in \mathscr{X}_E$, we can consider any—not necessar-

ily recursively decomposable—network and, for any $t \in G$, efficiently compute the inferences that are required to classify X_t; see Section 7.6.1$_{271}$.

7.5.6 *Multiple query nodes with evidence*

So far, we have restricted attention to inferences with multiple query variables and no evidence—Sections 7.5.3$_{249}$ and 7.5.4$_{252}$—or a single query variable with evidence—Section 7.5.5$_{258}$. We now consider the remaining case: multiple query variables with evidence.

Consider a non-empty comparable set $Q \subseteq G$, some $\hat{x}_Q \in \mathscr{X}_Q$ and, for all $i \in G \setminus Q$, an event $B_i \in \mathscr{P}_\emptyset(\mathscr{X}_i)$. Define $O := \times_{i \in (G \setminus Q)} B_i$. For any $E \subseteq G \setminus Q$ and $x_E \in \mathscr{X}_E$, this includes $O = \{x_E\}$ as a special case. We will compute $\underline{E}_Q^{\text{irr}*}(\hat{x}_Q \rfloor O)$ and $\underline{R}_Q^{\text{irr}*}(\hat{x}_Q \rfloor O)$ and the corresponding upper previsions. The requirement that Q should be a comparable set is not that restrictive. For example, in a hidden Markov model (of order m), the set of hidden nodes is comparable, which means that we can use the algorithm below to compute the lower and upper probability of a hidden state sequence conditional on the value of—or partial information about—the observable states. Arbitrary subsets of the nodes of a Markov chain (of order m) are also comparable. As another example, in Figure 7.9$_{240}$, every subset of $\{s_1, \dots, s_9\}$ is comparable.

We start by computing $\underline{E}_Q^{\text{irr}*}(\hat{x}_Q \rfloor O)$ and $\underline{R}_Q^{\text{irr}*}(\hat{x}_Q \rfloor O)$. As we know from Section 7.3$_{219}$, all we need in order to be able to do that is some method for evaluating the real-valued function $\underline{\rho}^{\text{irr}}_{\hat{x}_Q, O}$, defined by

$$\underline{\rho}^{\text{irr}}_{\hat{x}_Q, O}(\mu) := \underline{P}_G^{\text{irr}}\left(\mathbb{I}_O[\mathbb{I}_{\hat{x}_Q} - \mu]\right) \text{ for all } \mu \in \mathbb{R}.$$

We will now develop a recursive algorithm that can do this efficiently. We start by defining

$$f_i := \begin{cases} 1 & \text{if } i \in Q \\ \mathbb{I}_{B_i} & \text{if } i \notin Q \end{cases} \quad \text{for all } i \in G$$

and

$$\phi_s := \prod_{i \in K_s} f_i \text{ for all } s \in G$$

and

$$\kappa_s := \phi_s[\mathbb{I}_{\hat{x}_{Q \cap K_s}} - \mu] \text{ for all } s \in G \text{ such that } Q \cap K_s \neq 0.$$

Since $Q \neq 0$, we know from Proposition 105$_{255}$ that there is a unique $t \in Ro(G)$ such that $K_t \cap Q \neq \emptyset$. It then follows that

$$\mathbb{I}_O[\mathbb{I}_{\hat{x}_Q} - \mu] = [\mathbb{I}_{\hat{x}_Q} - \mu] \prod_{i \in G} f_i = \kappa_t \prod_{s \in Ro(G) \setminus \{t\}} \phi_s. \tag{7.46}$$

By combining this with Equations (6.9)$_{173}$ and (7.20)$_{246}$, we find that

$$\underline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu) = \begin{cases} \underline{P}^{\text{irr}}_{K_t}(\kappa_t) \prod\limits_{s\in Ro(G)\setminus\{t\}} \underline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \underline{P}^{\text{irr}}_{K_t}(\kappa_t) \geq 0 \\ \underline{P}^{\text{irr}}_{K_t}(\kappa_t) \prod\limits_{s\in Ro(G)\setminus\{t\}} \overline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \underline{P}^{\text{irr}}_{K_t}(\kappa_t) \leq 0. \end{cases} \tag{7.47}$$

In order to compute this number, we need to compute the value of $\underline{P}^{\text{irr}}_{K_t}(\kappa_t)$ and—since we do not know beforehand what the sign of $\underline{P}^{\text{irr}}_{K_t}(\kappa_t)$ will be—for all $s \in Ro(G) \setminus \{t\}$, both $\underline{P}^{\text{irr}}_{K_s}(\phi_s)$ and $\overline{P}^{\text{irr}}_{K_s}(\phi_s)$.

As before, this task is a special case—with $P(K_s) = \emptyset$—of a more general problem, which can be solved recursively. For every $s \in G$ and $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$, we compute $\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s)$ and $\overline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\phi_s)$ and, if $Q \cap K_s \neq \emptyset$, also $\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s)$. For the first two values, we can use the same recursive expressions as before [see Equations $(7.29)_{253}$, $(7.32)_{253}$, $(7.34)_{254}$ and $(7.35)_{254}$]. If $Q \cap K_s \neq \emptyset$, then $\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s)$ can be computed by means of the following expressions.

We consider two cases. If s is a leaf of the induced forest, this problem is trivial. In that case, it follows from Equation $(7.21)_{246}$ that

$$\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s) = \underline{P}_{s\rfloor x_{P(s)}}(\mathbb{I}_{\hat{x}_s} - \mu) \tag{7.48}$$

If s is not a leaf of the induced forest, we can apply Proposition 103_{246} and Equation $(6.14)_{178}$ to find that

$$\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s) = \underline{P}_{s\rfloor x_{P(s)}}\left(\underline{P}^{\text{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\kappa_s)\right) = \underline{P}_{s\rfloor x_{P(s)}}\left(\underline{g}_{s\rfloor x_{P(K_s)}}\right), \tag{7.49}$$

where $\underline{g}_{s\rfloor x_{P(K_s)}} := \underline{P}^{\text{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\kappa_s)$ is a gamble on $\mathscr{X}_s$ that is given by one of the following three expressions. If $s \in Q$ and $Q \cap D(s) \neq \emptyset$, then for all $x_s \in \mathscr{X}_s$:

$$\underline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} \underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu] \prod\limits_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s; \\ -\mu\, \overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod\limits_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \geq 0; \\ -\mu\, \underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod\limits_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \leq 0. \end{cases}$$

If $s \in Q$ and $Q \cap D(s) = \emptyset$, then for all $x_s \in \mathscr{X}_s$:

$$\underline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} (1-\mu)\,\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s \text{ and } \mu \leq 1; \\ (1-\mu)\,\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s \text{ and } \mu \geq 1; \\ -\mu\,\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \geq 0; \\ -\mu\,\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \leq 0. \end{cases}$$

If $s \notin Q$, then $Q \cap D(s) \neq \emptyset$ [becasue $Q \cap K_s \neq \emptyset$] and, for all $x_s \in \mathscr{X}_s$:

$$\underline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} \underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i\right) & \text{if } x_s \in B_s; \\ 0 & \text{if } x_s \notin B_s. \end{cases}$$

If $s \in Q$ and $Q \cap D(s) = \emptyset$, we can already compute this gamble recursively by means of Equations $(7.31)_{253}$ and $(7.33)_{254}$. In the two remaining cases, we have that $Q \cap D(s) \neq \emptyset$. Proposition 106_{255} then guarantees that there is a unique $t \in \tilde{C}(s)$ such that $K_t \cap Q \neq \emptyset$. In combination with Corollary 102_{245}, this implies that

$$[\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i = \kappa_t \prod_{c\in\tilde{C}(s)\setminus\{t\}} \phi_c. \tag{7.50}$$

By applying Proposition 104_{247} and Equation $(6.9)_{173}$, we find that

$$\begin{aligned} &\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i\right) \\ &\qquad = \begin{cases} \underline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t)\prod_{c\in\tilde{C}(s)\setminus\{t\}} \underline{P}^{\text{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \underline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t) \geq 0 \\ \underline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t)\prod_{c\in\tilde{C}(s)\setminus\{t\}} \overline{P}^{\text{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \underline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t) \leq 0 \end{cases} \end{aligned}$$

By combining this expression with Equations $(7.31)_{253}$ and $(7.33)_{254}$, the gamble $\underline{g}_{s\rfloor x_{P(K_s)}}$ can be evaluated recursively in all instances.

The discussion above suggest the following recursive procedure. We start by applying Equations $(7.29)_{253}$ and $(7.34)_{254}$ and, if $s \in Q$, also Equation $(7.48)_{\frown}$ at the leaves s of the induced forest. Next, we move along the branches of the induced forest towards its roots by means of Equations $(7.32)_{253}$ and $(7.35)_{254}$ and, if $Q \cap K_s \neq \emptyset$, also Equations $(7.48)_{\frown}$ and $(7.49)_{\frown}$.

A final application of Equation $(7.47)_{265}$ will provide us with the value of $\underline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu)$.

The computational complexity of this procedure is identical to that of the previous algorithms that we have discussed. For every $s \in G$, we now need to compute at most three real-valued functions on $\mathscr{X}_{P(K_s)}$: $\underline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)$ and $\overline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\phi_s)$ and, if $Q \cap K_s \neq \emptyset$, also $\underline{P}^{\text{irr}}_{K_s\rfloor X_{P(K_s)}}(\kappa_s)$. Evaluating one of these functions for some fixed $x_{P(K_s)} \in \mathscr{X}_{P(K_s)}$ can be done in constant time, using the recursive expressions that are provided above. Therefore, the complexity of computing $\underline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu)$ in this way is linear in the number of nodes $|G|$ and exponential in the parameter $\max_{s\in G}|P(K_s)|$. Again, this implies that for trees and forests, it is linear in the number of nodes. The complexity of computing $\underline{E}^{\text{irr}*}_Q(\hat{x}_Q\rfloor O)$ or $\underline{R}^{\text{irr}*}_Q(\hat{x}_Q\rfloor O)$ is identical because, as explained in Section $7.5.5_{258}$, for some fixed floating point precision, the number of μ's for which we need to compute $\underline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu)$ is bounded.

$\overline{E}^{\text{irr}*}_Q(\hat{x}_Q\rfloor O)$ and $\overline{R}^{\text{irr}*}_Q(\hat{x}_Q\rfloor O)$ can be computed in a similar—and equally efficient—way. By conjugacy, we know that they are equal to $-\underline{E}^{\text{irr}*}_Q(-\mathbb{I}_{\hat{x}_Q}\rfloor O)$ and $-\underline{R}^{\text{irr}*}_Q(-\mathbb{I}_{\hat{x}_Q}\rfloor O)$, so we can focus on computing $\underline{E}^{\text{irr}*}_Q(-\mathbb{I}_{\hat{x}_Q}\rfloor O)$ or $\underline{R}^{\text{irr}*}_Q(-\mathbb{I}_{\hat{x}_Q}\rfloor O)$. As we know from Section 7.3_{219}, all we need to compute these numbers is some method for evaluating the real-valued function $\overline{\rho}^{\text{irr}}_{\hat{x}_Q,O}$, defined by

$$\overline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu) := \underline{P}^{\text{irr}}_G\big(\mathbb{I}_O[-\mathbb{I}_{\hat{x}_Q} - \mu]\big) = -\overline{P}^{\text{irr}}_G\big(\mathbb{I}_O[\mathbb{I}_{\hat{x}_Q} + \mu]\big) \text{ for all } \mu \in \mathbb{R}.$$

For reasons of notational convenience, we will compute $\overline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(-\mu)$ instead of $\overline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(\mu)$; clearly, this makes no difference.

By combining Equations $(6.9)_{173}$, $(7.20)_{246}$ and 7.46 with conjugacy, we find that

$$\overline{\rho}^{\text{irr}}_{\hat{x}_Q,O}(-\mu) = \begin{cases} -\overline{P}^{\text{irr}}_{K_t}(\kappa_t) \prod_{s\in Ro(G)\setminus\{t\}} \overline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \overline{P}^{\text{irr}}_{K_t}(\kappa_t) \geq 0 \\ -\overline{P}^{\text{irr}}_{K_t}(\kappa_t) \prod_{s\in Ro(G)\setminus\{t\}} \underline{P}^{\text{irr}}_{K_s}(\phi_s) & \text{if } \overline{P}^{\text{irr}}_{K_t}(\kappa_t) \leq 0. \end{cases}$$

Everything else is now completely analogous. The only difference is that for nodes $s \in G$ such that $Q \cap K_s \neq \emptyset$, we now need to compute $\overline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s)$ instead of $\underline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s)$

Again, we consider two cases. If s is a leaf of the induced forest, this problem is trivial. It then follows from Equation $(7.21)_{246}$ that

$$\overline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s) = \overline{P}_{s\rfloor x_{P(s)}}(\mathbb{I}_{\hat{x}_s} - \mu)$$

If s is not a leaf of the induced forest, we can combine Proposition 103_{246} and Equation $(6.14)_{178}$ with conjugacy to find that

$$\overline{P}^{\text{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s) = \overline{P}_{s\rfloor x_{P(s)}}\big(\overline{P}^{\text{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\kappa_s)\big) = \overline{P}_{s\rfloor x_{P(s)}}\big(\overline{g}_{s\rfloor x_{P(K_s)}}\big),$$

where $\overline{g}_{s\rfloor x_{P(K_s)}} := \overline{P}^{\text{irr}}_{D(s)\rfloor(x_{P(D(s))\setminus\{s\}},X_s)}(\kappa_s)$ is a gamble on $\mathscr{X}_s$ that is given by one of the following three expressions. If $s \in Q$ and $Q \cap D(s) \neq \emptyset$, then for all $x_s \in \mathscr{X}_s$:

$$\overline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} \overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s; \\ -\mu\,\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \geq 0. \\ -\mu\,\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \leq 0. \end{cases}$$

If $s \in Q$ and $Q \cap D(s) = \emptyset$, then for all $x_s \in \mathscr{X}_s$:

$$\overline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} (1-\mu)\,\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s \text{ and } \mu \leq 1; \\ (1-\mu)\,\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s = \hat{x}_s \text{ and } \mu \geq 1; \\ -\mu\,\underline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \geq 0; \\ -\mu\,\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left(\prod_{i\in D(s)} f_i\right) & \text{if } x_s \neq \hat{x}_s \text{ and } \mu \leq 0. \end{cases}$$

If $s \notin Q$, then $Q \cap D(s) \neq \emptyset$ [because $Q \cap K_s \neq \emptyset$] and, for all $x_s \in \mathscr{X}_s$:

$$\overline{g}_{s\rfloor x_{P(K_s)}}(x_s) = \begin{cases} \overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i\right) & \text{if } x_s \in B_s; \\ 0 & \text{if } x_s \notin B_s. \end{cases} \tag{7.51}$$

Furthermore, in the two cases where $Q \cap D(s) \neq \emptyset$, we can combine Proposition 104_{247} and conjugacy with Equations $(6.9)_{173}$ and $(7.50)_{266}$ to find that

$$\overline{P}^{\text{irr}}_{D(s)\rfloor x_{P(D(s))}}\left([\mathbb{I}_{\hat{x}_{Q\cap D(s)}} - \mu]\prod_{i\in D(s)} f_i\right)$$
$$= \begin{cases} \overline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t)\prod_{c\in\tilde{C}(s)\setminus\{t\}} \overline{P}^{\text{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \overline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t) \geq 0 \\ \overline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t)\prod_{c\in\tilde{C}(s)\setminus\{t\}} \underline{P}^{\text{irr}}_{K_c\rfloor x_{P(K_c)}}(\phi_c) & \text{if } \overline{P}^{\text{irr}}_{K_t\rfloor x_{P(K_t)}}(\kappa_t) \leq 0. \end{cases}$$

By combining this expression with Equations $(7.31)_{253}$ and $(7.33)_{254}$, the gamble $\overline{g}_{s\rfloor x_{P(K_s)}}$, and therefore also the value of $\overline{P}^{\mathrm{irr}}_{K_s\rfloor x_{P(K_s)}}(\kappa_s)$, can be evaluated recursively in all instances.

7.5.7 *Other types of inferences*

Although the algorithms in the previous sections already cover a number of very important types of inferences in recursively decomposable networks, it is important to realise that they are only examples. The recursive decomposition of $\underline{P}^{\mathrm{irr}}_G(\cdot)$ presented in Section $7.5.2_{246}$ can be used to compute other inferences in recursively decomposable networks as well. The main idea is always the same: plug some function in the recursive equations for $\underline{P}^{\mathrm{irr}}_G(\cdot)$ and use the mathematical properties of marginal extension and independent natural extension to simplify the resulting expressions. For unconditional inferences, this function will be a gamble $f \in \mathscr{G}(\mathscr{X}_G)$ and the goal is then to compute $\underline{P}^{\mathrm{irr}}_G(f)$. For conditional inferences, the function that is plugged into $\underline{P}^{\mathrm{irr}}_G(\cdot)$ is of the form $\mathbb{I}_O[f-\mu]$ and an additional iteration over μ is then needed to compute $\underline{E}^{\mathrm{irr}*}_G(f\rfloor O)$ or $\underline{R}^{\mathrm{irr}*}_G(f\rfloor O)$. We end this section by briefly discussing two additional examples. We provide only very few details; our goal here is to illustrate the range of additional inferences that can be computed by means of the decomposition in Section $7.5.2_{246}$, and not to actually compute them.

As a first—rather arbitrary—example, consider the recursively decomposable network in Figure 7.9_{240} and let $Q = \{s_1, s_i\}$, with $i \in G \setminus \{1\}$. Let E be any subset of $G \setminus Q$ and consider some $f \in \mathscr{G}(\mathscr{X}_Q)$ and $x_E \in \mathscr{X}_E$. Then in order to compute $\underline{E}^{\mathrm{irr}*}_Q(f\rfloor x_E)$ or $\underline{R}^{\mathrm{irr}*}_Q(f\rfloor x_E)$, as we know from Section 7.3_{219}, all that we need is some method for evaluating $\underline{P}^{\mathrm{irr}}_G(\mathbb{I}_{\{x_E\}}[f-\mu])$ for different values of $\mu \in \mathbb{R}$. Since $G = K_{s_1}$, it follows from Equations $(6.14)_{178}$ and $(7.22)_{247}$ that

$$\begin{aligned}\underline{P}^{\mathrm{irr}}_G\left(\mathbb{I}_{\{x_E\}}[f-\mu]\right) &= \underline{P}_{s_1}\left(\underline{P}^{\mathrm{irr}}_{D(s_1)\rfloor X_{s_1}}\left(\mathbb{I}_{\{x_E\}}[f-\mu]\right)\right)\\ &= \underline{P}_{s_1}\left(\sum_{x_{s_1}\in\mathscr{X}_{s_1}} \mathbb{I}_{\{x_{s_1}\}}\underline{P}^{\mathrm{irr}}_{D(s_1)\rfloor x_{s_1}}\left(\mathbb{I}_{\{x_E\}}[f_{x_{s_1}}-\mu]\right)\right)\end{aligned}$$

where, for all $x_{s_1} \in \mathscr{X}_{s_1}$, $f_{x_{s_1}}$ is a gamble on $\mathscr{X}_{s_i}$ that is defined by

$$f_{x_{s_1}}(x_{s_i}) := f(x_{s_1}, x_{s_i}) \text{ for all } x_{s_i} \in \mathscr{X}_{s_i}.$$

The rest of the computation now consists in applying the recursive expressions in Section $7.5.4_{252}$ because $\mathbb{I}_{\{x_E\}}[f_{x_{s_1}}-\mu]$ is a product of univariate functions, only one of which is non-negative. So, although the original problem—computing $\underline{P}^{\mathrm{irr}}_G(\mathbb{I}_{\{x_E\}}[f-\mu])$—cannot be solved by means of the algorithms in the previous sections, we see that with a little extra effort, we obtain a solution method whose computational complexity is similar to that of the algorithms presented before.

As a final example, we consider the problem of state estimation in hidden Markov models. For example, for the hidden Markov model in Figure 7.5$_{238}$, with $Q = \{s_1, \ldots, s_7\}$ and $E = \{s_8, s_{14}\}$, the problem of state estimation consists in estimating the value of X_Q based on the information that $X_E = x_E$, for some sequence of observed states $x_E \in \mathscr{X}_E$. If we use regular extension, this information leads us to consider the updated model $\underline{R}_Q^{\mathrm{irr}}(\cdot\rfloor x_E)$, which, provided that x_E has positive upper probability, is equal to $\underline{R}_Q^{\mathrm{irr}*}(\cdot\rfloor x_E)$. In order to estimate—classify—the value of X_Q, we can now apply an imprecise-probabilistic decision criterion.

If we use interval dominance, Γ-maximin or Γ-maximax [see Section 7.5.5$_{258}$], this requires us to compute inferences of the form $\underline{R}_Q^{\mathrm{irr}*}(x_Q\rfloor x_E)$ and $\overline{R}_Q^{\mathrm{irr}*}(x_Q\rfloor x_E)$, with $x_Q \in \mathscr{X}_Q$. Since Q is a comparable set, we can use the algorithm in Section 7.5.6$_{264}$ to perform these computations efficiently. However, this does not mean that the actual classification problem can be solved efficiently. Since the number of sequences $x_Q \in \mathscr{X}_Q$ for which we need to compute the inferences above is exponential in the length of these sequences, this approach is intractable for long hidden Markov models.

An alternative option is to use the maximality criterion [see Section 7.5.5$_{258}$]. In that case, we need to be able to check inequalities of the form $\underline{R}_Q^{\mathrm{irr}*}(\mathbb{I}_{\{\tilde{x}_Q\}} - \mathbb{I}_{\{x_Q\}}\rfloor x_E) > 0$, with $\tilde{x}_Q, x_Q \in \mathscr{X}_Q$. As we know from Section 7.3$_{219}$, evaluating the left-hand side of such an inequality requires us to compute $\underline{P}_G^{\mathrm{irr}}([\mathbb{I}_{\{\tilde{x}_Q\}} - \mathbb{I}_{\{x_Q\}} - \mu]\mathbb{I}_{\{x_E\}})$ for different values of $\mu \in \mathbb{R}$. Although none of the algorithms that we have introduced is capable of doing that, it should be clear that this problem is very similar to that in Section 7.5.6$_{264}$. It should therefore not be surprising that it is possible to obtain recursive expressions that allow us to compute $\underline{P}_G^{\mathrm{irr}}([\mathbb{I}_{\{\tilde{x}_Q\}} - \mathbb{I}_{\{x_Q\}} - \mu]\mathbb{I}_{\{x_E\}})$ efficiently. We will not derive these expression here; it serves as nice exercise. See our paper [30] for more information. Again, the trick is to consider the decomposition in Section 7.5.2$_{246}$ and exploit the mathematical properties of marginal extension and independent natural extension to simplify the resulting expressions. As before, the fact that we can compute this single inference efficiently does not imply that the complete classification problem can be solved easily; the number of equalities that need to be checked is still exponential in the length of the hidden Markov model. Nevertheless, by applying a number of additional tricks, it is possible to solve this classification problem efficiently: we have designed an algorithm that returns all the maximal hidden sequences; it has a computational complexity that is linear in the number of such sequences. These 'additional tricks' and the resulting algorithm for state estimation in imprecise hidden Markov models are presented in Reference [30], which also discusses an application where we automatically detect and correct the errors that are made by Optical Character Recognition (OCR) software. A detailed discussion of this algorithm and its application falls beyond the scope of this dissertation. For our present purposes, the only point that is really important

here is that the main reason why we were able to solve this state estimation problem efficiently is because the expressions necessary for solving it can be obtained recursively, using methods similar to the ones derived in the previous sections.

We trust that these additional examples, and the algorithms in the previous sections, will inspire others to use similar techniques to solve the inference problems relevant for their particular applications.

7.6 WHAT ABOUT OTHER TYPES OF NETWORKS?

Up to this point, we have presented two types of algorithms for computing inferences in credal networks under epistemic irrelevance. For specific inferences in recursively decomposable networks, we have developed efficient recursive algorithms. For general inferences in general networks, we have introduced—generally intractable—brute force methods and a number of preprocessing tricks to make the inference problem smaller, and to reduce it to the unconditional case. We conclude this chapter by showing that even for general networks that are not recursively decomposable, inferences can sometimes be computed efficiently, either by exploiting the theoretical properties in Chapter 6_{152}, by reducing the network to a recursively decomposable one, or by applying brute force techniques only locally. Except for the result in Section 7.6.1, which can be regarded as the basis for an algorithm, we will not develop generally applicable methods; in fact, we believe that recursively decomposable networks are more or less the largest class of networks for which this is possible. Nevertheless, as we will illustrate by means of examples, it is still possible to devise solution methods on a case-by-case basis. As before, we trust that these examples will inspire others to use similar techniques to solve the inference problems relevant for their particular applications

7.6.1 *A single query node with complete evidence*

There is one specific inference problem that can be solved efficiently in any credal network under epistemic irrelevance: the computation of $\underline{P}_q^{\text{irr}}(f\rfloor x_E)$ or $\underline{R}_q^{\text{irr}}(f\rfloor x_E)$, in the specific case where $E = G \setminus \{q\}$. This means that the value of all the non-query nodes is known; we call this *complete evidence*. This type of inference is important in classifiers, where data is usually not missing. In expert systems however, complete evidence is a rather idealistic situation.

Solving this inference problem is trivial if q is a leaf node. We then have that $PN(q) = E$ and it therefore follows directly from Corollaries 44_{141} and 75_{181} that

$$\underline{P}_q^{\text{irr}}(f\rfloor x_E) = \underline{R}_q^{\text{irr}}(f\rfloor x_E) = \underline{P}_{q\rfloor x_{P(q)}}(f).$$

When q is not a leaf node, we start by applying the tricks at the end of Section 7.2_{216}. For $Q = \{q\}$, the smallest closed subset K of G such that $Q \subseteq K$,

$P(K) \subseteq E$ and $D(K) \cap E = \emptyset$ is clearly $K := K_q = D(q) \cup \{q\}$. Therefore, and because $E \cap K_q = D(q)$ and $PN(K_q) = PN(q)$ and $E \setminus K_q = G \setminus K_q = PN(q)$, it follows from the discussion in Section 7.2$_{216}$ that

$$\underline{P}_q^{\text{irr}}(f \rfloor x_E) = \underline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)})$$

and

$$\underline{R}_q^{\text{irr}}(f \rfloor x_E) = \begin{cases} \underline{R}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)}) & \text{if } \overline{P}_{PN(q)}^{\text{irr}}(x_{PN(q)}) > 0 \\ \underline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)}) & \text{if } \overline{P}_{PN(q)}^{\text{irr}}(x_{PN(q)}) = 0. \end{cases}$$

where, because of Proposition 58$_{159}$:

$$\overline{P}_{PN(q)}^{\text{irr}}(x_{PN(q)}) = \prod_{s \in PN(q)} \overline{P}_{s \rfloor x_{P(s)}}(x_s).$$

In this way, the original inference problem is already reduced to a similar but smaller sized inference problem in the subnetwork that corresponds to K_q and $x_{P(K_q)}$. The remaining task is to compute $\underline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)})$ or $\underline{R}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)})$.

By applying the discussion in Section 7.3$_{219}$ to the subnetwork that corresponds to K_q and $x_{P(K_q)}$, we know that

$$\underline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)}) = \underline{E}_{q \rfloor x_{P(K_q)}}^{\text{irr}*}(f \rfloor x_{D(q)}) \text{ if } \underline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(x_{D(q)}) > 0$$

and

$$\underline{R}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(f \rfloor x_{D(q)}) = \underline{R}_{q \rfloor x_{P(K_q)}}^{\text{irr}*}(f \rfloor x_{D(q)}) \text{ if } \overline{P}_{q \rfloor x_{P(K_q)}}^{\text{irr}}(x_{D(q)}) > 0.$$

If the positivity conditions are not satisfied, the equalities may fail and the right-hand side of these equations then provides a—vacuous—lower approximation of the left-hand side. In the remainder of this section, we compute the right-hand side of these equalities.

As we know from Section 7.3$_{219}$, all that we need in order to compute these right-hand sides—and to check the positive conditions—is an efficient method for evaluating the real-valued function $\rho_{f, x_{K_q} \rfloor x_{P(K_q)}}^{\text{irr}}$, defined by

$$\rho_{f, x_{K_q} \rfloor x_{P(K_q)}}^{\text{irr}}(\mu) := \underline{P}_{K_q \rfloor x_{P(K_q)}}^{\text{irr}}\left([f - \mu]\mathbb{I}_{\{x_{D(q)}\}}\right) \text{ for all } \mu \in \mathbb{R}.$$

Once we can do that, a simple root-finding procedure such as the bisection method provides us with the value of $\underline{E}_{q \rfloor x_{P(K_q)}}^{\text{irr}*}(f \rfloor x_{D(q)})$ and $\underline{R}_{q \rfloor x_{P(K_q)}}^{\text{irr}*}(f \rfloor x_{D(q)})$ that we are after.

So let us fix some $\mu \in \mathbb{R}$. Since q is not a leaf of the network, we know that $D(q) \neq \emptyset$. Therefore, it follows from Proposition 103_{246} and Equation 6.14_{178} that

$$\rho^{\mathrm{irr}}_{f,x_{K_q}\rfloor x_{P(K_q)}}(\mu) = \underline{P}_{q\rfloor x_{P(q)}}(g_\mu),$$

where $g_\mu := \underline{P}^{\mathrm{irr}}_{D(q)\rfloor(x_{P(D(q))\setminus\{q\}},X_q)}\big([f-\mu]\mathbb{I}_{\{x_{D(q)}\}}\big)$ is a gamble on $\mathscr{X}_q$ that can be easily computed because it follows from coherence [$\mathrm{C2}_{48}$], conjugacy and Proposition 58_{159} that

$$g_\mu(x_q) = \begin{cases} [f(x_q)-\mu] \prod_{s\in D(q)} \underline{P}_{s\rfloor x_{P(s)}}(x_s) & \text{if } f(x_q) \geq \mu \\ [f(x_q)-\mu] \prod_{s\in D(q)} \overline{P}_{s\rfloor x_{P(s)}}(x_s) & \text{if } f(x_q) \leq \mu \end{cases} \quad \text{for all } x_q \in \mathscr{X}_q.$$

The computational complexity of evaluating $\rho^{\mathrm{irr}}_{f,x_{K_q}\rfloor x_{P(K_q)}}$ is therefore linear in the number of nodes in $D(q)$.

7.6.2 *Reducing a problem to the recursively decomposable case*

As we have explained in Section 7.2_{216}, it is sometimes possible to reduce an inference problem in a global network to a similar inference problem in a subnetwork. If the subnetwork is recursively decomposable, this trick can turn a seemingly intractable inference problem into an easier problem that can be solved efficiently by means of the recursive algorithms in Section 7.5_{235}. We illustrate this by means of an example.

Consider the DAG in Figure $7.14_{\curvearrowright}$, which is not recursively decomposable because s_1 and s_3 are incomparable and yet they have a common descendant s_{10}, and similarly for s_2 and s_5 and for s_7 and s_8. This can also be seen by looking at the induced DAG, whose edges have been made thicker. Since the induced DAG is not a forest, we know from Proposition 99_{243} that the original DAG is not recursively decomposable.

Consider now a credal network under epistemic irrelevance that has this graphical structure. Let $Q := \{s_2, s_{12}\}$ and $E := \{s_5, s_8, s_{11}, s_{13}\}$ and choose some $\hat{x}_Q \in \mathscr{X}_Q$ and $x_E \in \mathscr{X}_E$. The inference task that we consider is that of computing the value of $\underline{P}^{\mathrm{irr}}_Q(\hat{x}_Q\rfloor x_E)$.

We will simplify this problem by means of the techniques in Section 7.2_{216}. The trick is to find a closed subset K of G such that $Q \subseteq K$, $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$. In this case, it is not hard to see that the set $K := \{s_2, s_3, s_4, s_{11}, s_{12}, s_{13}\}$ meets these criteria. In fact, it is the unique smallest set K for which this is the case. Since $P(K) = \{s_5\}$ and $E' := E \cap K = \{s_{11}, s_{13}\}$, it now follows from the discussion in Section 7.2_{216} that

$$\underline{P}^{\mathrm{irr}}_Q(\hat{x}_Q\rfloor x_E) = \underline{P}^{\mathrm{irr}}_{Q\rfloor x_{s_5}}(\hat{x}_Q\rfloor x_{E'}),$$

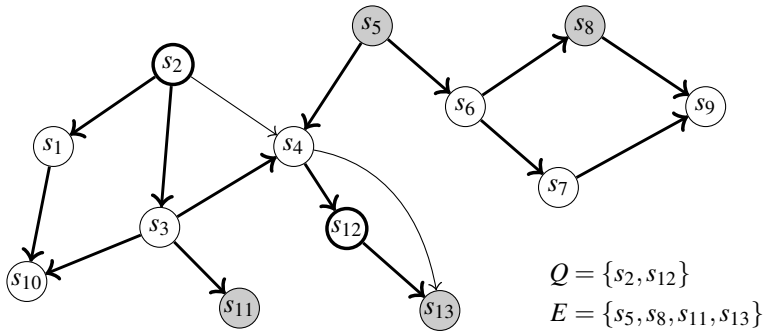


Figure 7.14: Example of a DAG that is not recursively decomposable

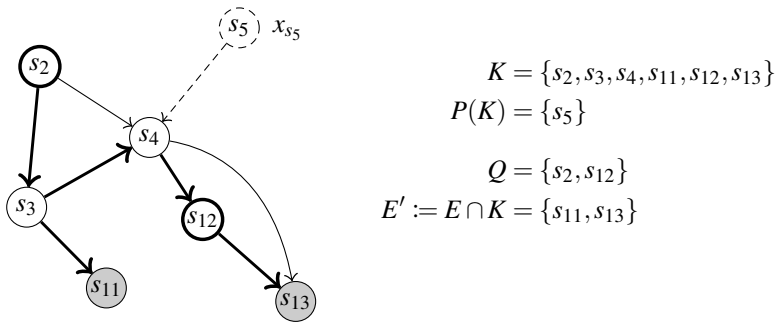


Figure 7.15: Recursively decomposable sub-DAG of the DAG in Figure 7.14

where the right-hand side is computed with respect to the subnetwork that corresponds to K and x_{s_5}; see Figure 7.15 for an illustration.

What is especially nice about this simplification of the problem is that the sub-DAG that corresponds to K is recursively decomposable. Therefore, we can use the algorithm in Section 7.5.4$_{252}$ to compute $\underline{P}^{\text{irr}}_{K}(x_{E'})$. If this lower probability is positive, then as explained in Section 7.3, $\underline{P}^{\text{irr}}_{Q\rfloor x_{s_5}}(\hat{x}_Q \rfloor x_{E'})$ is equal to $\underline{E}^{\text{irr}*}_{Q\rfloor x_{s_5}}(\hat{x}_Q \rfloor x_{E'})$, which, since Q is clearly a comparable set, can be computed efficiently by means of the algorithm in Section 7.5.6$_{264}$.

7.6.3 *Combining brute force techniques with recursion*

The computational methods that we have presented so far consist of two types. On the one hand, we have presented generally applicable brute force methods, which are intractable for large networks. On the other hand, we have presented

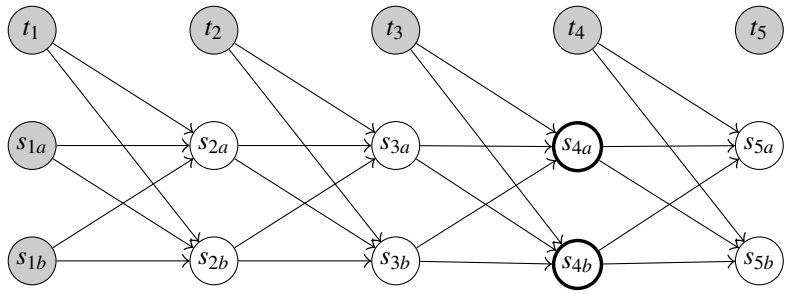


Figure 7.16: Example of a dynamic network that is not recursively decomposable

efficient recursive algorithms that manage to decompose the problem in such a way that all we really ever have to evaluate are the local models. However, these are not the only two options. For some inference problems that cannot be fully decomposed, it may nevertheless be possible to decompose them partially. The idea is to recursively decompose the global problem into subproblems that are as small as possible, and to solve these remaining problems by means of brute force methods. We conclude this chapter with an example that illustrates this method.

Consider a network that consists of $3n$ nodes, with $n \in \mathbb{N}$. For every $i \in \{1, \ldots, n\}$, the variables X_{t_i}, $X_{s_{ia}}$ and $X_{s_{ib}}$ represent domain-specific parameters at time i, whose local models depend on the value of these parameters in the previous time slot. This is a simple example of a dynamic network, which aims to model the evolution of parameters as time evolves. Figure 7.16 depicts an example for $n = 5$. For $n = 14$, this network was used as an example in a recent paper about dynamic credal networks [62]. It is a simplified version of a network that models various aspects of the ripening process of Camembert cheese [5]. It represents the coupled dynamics of a yeast behaviour—Kluyveromyces marxianus concentration $X_{s_{ia}}$—with their substrate consumptions—lactose concentration $X_{s_{ib}}$—influenced by temperature—X_{t_i}. For our present purposes, the local models that are attached to the variables in this network are not really important; see Reference [62] for examples. What we intend to show here is that inference in this network can be performed efficiently.

We focus on one particular inference problem, which is estimating the evolution of the variables $X_{s_{ia}}$ and $X_{s_{ib}}$ through time, based on their value at time 1 and the evolution of the temperature—X_{t_i}—over time. In order to formalise this problem we define

$$S_i := \{s_{ia}, s_{ib}\} \text{ for all } i \in \{1, \ldots, n\}$$

and

$$T_{k:\ell} := \{t_k, \ldots, t_\ell\} \text{ for all } k, \ell \in \{1, \ldots, n\} \text{ such that } k \leq \ell.$$

Let $m \in \{2, \ldots, n\}$ be some arbitrary point in time and consider a gamble $f \in \mathscr{G}(\mathscr{X}_{S_m})$, an initial state $x_{S_1} \in \mathscr{X}_{S_1}$ and a sequence of temperatures $x_{T_{1:n}} \in \mathscr{X}_{T_{1:n}}$. The generic inference problem that we intend to solve is the computation of $\underline{P}^{\text{irr}}_{S_m}(f \rfloor x_{S_1 \cup T_{1:n}})$ or $\underline{R}^{\text{irr}}_{S_m}(f \rfloor x_{S_1 \cup T_{1:n}})$, which—as we will show—are two identical problems. The example in Figure 7.16$_{\frown}$ corresponds to $m = 4$. Reference [62] considered a specific version[26] of this inference problem, for a credal network with the same graphical structure but with other independence assumptions.[27] The authors used an approximate Monte Carlo sampling algorithm to compute their inferences because no exact algorithm was available. As we are about to show, in our case—for a credal network under epistemic irrelevance—these inferences can be computed efficiently and exactly, with a recursive algorithm that has a computational complexity that is linear in m.

The first step is to simplify the inference problem by applying the techniques in Section 7.2$_{216}$. If we define

$$S_{k:\ell} := \cup_{i=k}^{\ell} S_i \text{ for all } k, \ell \in \{1, \ldots, n\} \text{ such that } k \leq \ell,$$

then for all $k, \ell \in \{1, \ldots, n\}$ such that $k \leq \ell$:

$$P(S_{k:\ell}) = S_{k-1} \cup T_{k-1:\ell-1} \text{ and } D(S_{k:\ell}) = S_{\ell+1:n}$$

If we now let $Q := S_m$ and $E := S_1 \cup T_{1:n}$, then it should be clear that $K := S_{2:m}$ is a closed subset of G such that $Q \subseteq K$, $P(K) \subseteq E$ and $D(K) \cap E = \emptyset$. Therefore, it follows from the discussion in Section 7.2$_{216}$ that

$$\underline{P}^{\text{irr}}_{S_m}(f \rfloor x_{S_1 \cup T_{1:n}}) = \underline{R}^{\text{irr}}_{S_m}(f \rfloor x_{S_1 \cup T_{1:n}}) = \underline{P}^{\text{irr}}_{S_m \rfloor x_{P(S_{2:m})}}(f),$$

where the right-hand side is an inference problem in the subnetwork that corresponds to $S_{2:m}$ and $x_{P(S_{2:m})}$. Figure 7.17$_{\frown}$ depicts an example with $m = 4$. The remaining task is to compute $\underline{P}^{\text{irr}}_{S_m \rfloor x_{P(S_{2:m})}}(f)$.

We now consider two cases: $m = 2$ and $m > 2$.

If $m = 2$, then $S_{2:m} = S_2 = \{s_{2a}, s_{2b}\}$. This implies that the subnetwork that corresponds to $S_{2:m}$ and $x_{P(S_{2:m})}$ consists of two disconnected nodes. Therefore, and because $P(S_{2:m}) = P(s_{2a}) = P(s_{2b})$, it follows from the discussion in Section 6.6$_{171}$ that

$$\underline{P}^{\text{irr}}_{S_m \rfloor x_{P(S_{2:m})}}(f) = \underline{P}^{\text{irr}}_{S_2 \rfloor x_{P(S_2)}}(f) = \underline{P}_{s_{2a} \rfloor x_{P(s_{2a})}} \otimes \underline{P}_{s_{2b} \rfloor x_{P(s_{2b})}}(f).$$

[26] Instead of general inferences about X_{S_m}, they computed lower and upper expected values of $X_{s_{ma}}$ and $X_{s_{mb}}$, which corresponds to choosing specific gambles $f \in \mathscr{G}(\mathscr{X}_{S_m})$.

[27] Strong independence and a notion of independence which they call repetitive independence.

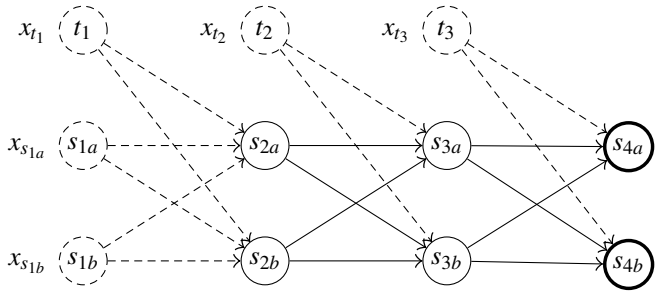


Figure 7.17: Subnetwork of the network in Figure 7.16$_{275}$

Since this computation involves only two local models, it can easily be dealt with by means of brute force techniques. If $X_{s_{2a}}$ and $X_{s_{2b}}$ are binary, we can use the extreme points in Theorem 92$_{233}$. Otherwise, we need to solve a linear program; see Equation 7.15$_{229}$ and Corollary 89$_{229}$. In any case, the important point here is that this computation is local in nature because it involves only two local models. Therefore, if the local models are finitely generated and are based on a reasonable number of assessments, this computation can be expected to be tractable.

If $m > 2$, it follows from Corollary 72$_{180}$ [with $K = S_{2:m}$, $T = S_2$ and $S = S_{3:m}$] that

$$\underline{P}^{\text{irr}}_{S_{2:m}\rfloor x_{P(S_{2:m})}}(\cdot) = \underline{P}^{\text{irr}}_{S_2\rfloor x_{P(S_2)}}(\cdot) \oslash \underline{P}^{\text{irr}}_{S_{3:m}\rfloor (x_{P(S_{3:m})\setminus S_2}, X_{S_2})}(\cdot)$$

Therefore, if we define $g_2 \in \mathscr{G}(\mathscr{X}_{S_2})$ by

$$g_2(x_{S_2}) := \underline{P}^{\text{irr}}_{S_m\rfloor x_{P(S_{3:m})}}(f) \text{ for all } x_{S_2} \in \mathscr{X}_{S_2}, \tag{7.52}$$

where the right-hand side is an inference problem in the subnetwork that corresponds to $S_{3:m}$ and $x_{P(S_{3:m})}$, it follows from Equation (6.14)$_{178}$ that

$$\underline{P}^{\text{irr}}_{S_m\rfloor x_{P(S_{2:m})}}(f) = \underline{P}^{\text{irr}}_{S_2\rfloor x_{P(S_2)}}(g_2) = \underline{P}_{s_{2a}\rfloor x_{P(s_{2a})}} \otimes \underline{P}_{s_{2b}\rfloor x_{P(s_{2b})}}(g_2).$$

As before, this independent natural extension can be computed by means of brute force techniques because it only involves two local models. The remaining problem is now to evaluate g_2. As we can see from Equation (7.52), this inference problem is of the same type as the one that we have considered above. The solution is completely analogous. For all $x_{S_2} \in \mathscr{X}_{S_2}$, we find that

$$g_2(x_{S_2}) = \underline{P}^{\text{irr}}_{S_m\rfloor x_{P(S_{3:m})}}(f) = \underline{P}^{\text{irr}}_{S_3\rfloor x_{P(S_3)}}(g_3) = \underline{P}_{s_{3a}\rfloor x_{P(s_{3a})}} \otimes \underline{P}_{s_{3b}\rfloor x_{P(s_{3b})}}(g_3),$$

where g_3 is a gamble on $\mathscr{X}_{S_3}$ that is equal to f if $m = 3$ and defined by

$$g_3(x_{S_3}) := \underline{P}^{\text{irr}}_{S_m\rfloor x_{P(S_{4:m})}}(f) \text{ for all } x_{S_3} \in \mathscr{X}_{S_3}$$

if $m > 3$. By continuing in this way, we obtain the following recursive solution method. We start by defining $g_m := f$. Next, for all $i \in \{3, \dots, m\}$, we define $g_{i-1} \in \mathscr{G}(\mathscr{X}_{S_{i-1}})$ by

$$g_{i-1}(x_{S_{i-1}}) := \underline{P}_{s_{ia}\rfloor x_{P(s_{ia})}} \otimes \underline{P}_{s_{ib}\rfloor x_{P(s_{ib})}}(g_i) \text{ for all } x_{S_{i-1}} \in \mathscr{X}_{S_{i-1}}$$

The value of the final inference that we are after is then given by

$$\underline{P}^{\text{irr}}_{S_m}(f\rfloor x_{S_1 \cup T_{1:n}}) = \underline{R}^{\text{irr}}_{S_m}(f\rfloor x_{S_1 \cup T_{1:n}}) = \underline{P}_{s_{2a}\rfloor x_{P(s_{2a})}} \otimes \underline{P}_{s_{2b}\rfloor x_{P(s_{2b})}}(g_2).$$

Assuming that the independent natural extension of two local models can be computed in a reasonable—constant—amount of time, this recursive procedure has a computational complexity that is linear in m.

7.A PROOF OF THEOREM 88

Proof of Theorem 88$_{228}$. If $G = \{s\}$, the constraints in this theorem reduce to Equations (7.10)$_{226}$ and (7.11)$_{226}$ and, therefore, they are equivalent to imposing that $p_G \in \mathscr{F}_s$, which implies that p_G is a probability mass function. The result then follows trivially from Corollary 87$_{227}$.

Since the theorem holds for $|G| = 1$, we can now provide a proof by induction. Consider a credal network with $n := |G| > 1$ nodes and assume that the theorem is true for networks with $n-1$ nodes.

Due to Corollary 87$_{227}$, any $p_G \in \mathscr{F}^{\text{irr}}_G \rfloor \mathscr{X}_G$ clearly satisfies the constraints in this theorem. The hard part is the other implication. So, consider any real-valued function p_G on $\mathscr{X}_G$ for which $\sum_{z_G \in \mathscr{X}_G} p_G(z_G) = 1$ and, for all $s \in G$ and $x_{PN(s)} \in \mathscr{X}_{PN(s)}$:

$$\left(\forall \gamma \in \Gamma_{s\rfloor x_{P(s)}}\right) \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p_G(x_{PN(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0. \tag{7.53}$$

We need to prove that $p_G \in \mathscr{F}^{\text{irr}}_G \rfloor \mathscr{X}_G$. In order to do so, it suffices to show that p_G is a probability mass function, because it then follows from Corollary 87$_{227}$ that $p_G \in \mathscr{F}^{\text{irr}}_G \rfloor \mathscr{X}_G$. Since we already know that p_G satisfies the unitary constraint, we are left to prove that $p_G(z_G) \geq 0$ for all $z_G \in \mathscr{X}_G$.

Consider any $\ell \in G$ that is a leaf of the network; this is always possible because every DAG has at least one leaf. Define $K := G \setminus \{\ell\}$ and let q_K be the real-valued function on $\mathscr{X}_K$ that is defined by

$$q_K(z_K) = \sum_{z_\ell \in \mathscr{X}_\ell} p_G(z_K, z_\ell) \text{ for all } z_K \in \mathscr{X}_K.$$

Since $\sum_{z_G \in \mathscr{X}_G} p(z_G) = 1$, we find that $\sum_{z_K \in \mathscr{X}_K} q_K(z_K) = 1$. Since K is clearly an ancestral set, we can consider the irrelevant natural extension $\mathscr{F}^{\text{irr}}_K$ of the corresponding subnetwork. We now prove that $q_K \in \mathscr{F}^{\text{irr}}_K \rfloor \mathscr{X}_K$. Since

$|K| = n-1$, we can use the induction hypothesis for this purpose. It suffices to prove that, for any $s \in K$ and $x_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}$:

$$\big(\forall \gamma \in \Gamma_{s\rfloor x_{P(s)}}\big) \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D_K(s)} \in \mathscr{X}_{D_K(s)}} q_K(x_{PN_K(s)}, z_s, z_{D_K(s)})\gamma(z_s) \geq 0.$$

So fix any $s \in K$, $x_{PN_K(s)} \in \mathscr{X}_{PN_K(s)}$ and $\gamma \in \Gamma_{s\rfloor x_{P(s)}}$. We consider two cases: $D_K(s) = D(s)$ and $D_K(s) \cup \{\ell\} = D(s)$. It follows from Lemma 79(i)$_{185}$ that these two cases are exhaustive. If $D_K(s) = D(s)$, then $PN_K(s) \cup \{\ell\} = PN(s)$ and therefore

$$\begin{aligned}
&\sum_{z_s \in \mathscr{X}_s} \sum_{z_{D_K(s)} \in \mathscr{X}_{D_K(s)}} q_K(x_{PN_K(s)}, z_s, z_{D_K(s)})\gamma(z_s) \\
&\qquad = \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D_K(s)} \in \mathscr{X}_{D_K(s)}} \sum_{x_\ell \in \mathscr{X}_\ell} p_G(x_{PN_K(s)}, z_s, z_{D_K(s)}, x_\ell)\gamma(z_s) \\
&\qquad = \sum_{x_\ell \in \mathscr{X}_\ell} \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p_G(x_{PN(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0,
\end{aligned}$$

where the final inequality is a consequence of Equation (7.53)$_{\frown}$. In case $D_K(s) \cup \{\ell\} = D(s)$, we have that $PN_K(s) = PN(s)$ and therefore

$$\begin{aligned}
&\sum_{z_s \in \mathscr{X}_s} \sum_{z_{D_K(s)} \in \mathscr{X}_{D_K(s)}} q_K(x_{PN_K(s)}, z_s, z_{D_K(s)})\gamma(z_s) \\
&\qquad = \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D_K(s)} \in \mathscr{X}_{D_K(s)}} \sum_{z_\ell \in \mathscr{X}_\ell} p_G(x_{PN_K(s)}, z_s, z_{D_K(s)}, z_\ell)\gamma(z_s) \\
&\qquad = \sum_{z_s \in \mathscr{X}_s} \sum_{z_{D(s)} \in \mathscr{X}_{D(s)}} p_G(x_{PN(s)}, z_s, z_{D(s)})\gamma(z_s) \geq 0,
\end{aligned}$$

where the final inequality again follows from Equation (7.53)$_{\frown}$. Hence, by the induction hypothesis, it follows that $q_K \in \mathscr{F}_K^{\mathrm{irr}} \rfloor \mathscr{X}_K$, which implies that q_K is a probability mass function. We still need to prove that $p_G(z_G) \geq 0$ for all $z_G \in \mathscr{X}_G$.

Assume *ex absurdo* that there is some $z_G^* \in \mathscr{X}_G$ such that $p_G(z_G^*) < 0$. Since ℓ is a leaf of the network, we know that $D(\ell) = \emptyset$ and $PN(\ell) = K$. It therefore follows from Equation (7.53)$_{\frown}$ that

$$\big(\forall \gamma \in \Gamma_{\ell\rfloor z^*_{P(\ell)}}\big) \sum_{z_\ell \in \mathscr{X}_\ell} p_G(z_K^*, z_\ell)\gamma(z_\ell) \geq 0. \tag{7.54}$$

Choose some arbitrary $p_{\ell\rfloor z^*_{P(\ell)}} \in \mathscr{F}_{\ell\rfloor z^*_{P(\ell)}}$. Since $p_G(z_G^*) < 0$, it is then possible to choose $\varepsilon > 0$ such that $p(z_G^*) + \varepsilon p_{\ell\rfloor z^*_{P(\ell)}}(z_\ell^*) < 0$. Let θ be the real-valued function on $\mathscr{X}_\ell$ that is defined by $\theta(z_\ell) = p_G(z_K^*, z_\ell) + \varepsilon p_{\ell\rfloor z^*_{P(\ell)}}(z_\ell)$ for all $z_\ell \in \mathscr{X}_\ell$. Then $\theta(z_\ell^*) < 0$ and

$$\lambda := \sum_{z_\ell \in \mathscr{X}_\ell} \theta(z_\ell) = \sum_{z_\ell \in \mathscr{X}_\ell} p_G(z_K^*, z_\ell) + \varepsilon \sum_{z_\ell \in \mathscr{X}_\ell} p_{\ell\rfloor z^*_{P(\ell)}}(z_\ell) = q_K(z_K^*) + \varepsilon \geq \varepsilon > 0,$$

where the first inequality follows from the fact that q_K is a probability mass function. Therefore, we can define the real-valued function $\pi = \theta/\lambda$ on $\mathscr{X}_\ell$, for which it should be clear that $\sum_{z_\ell \in \mathscr{X}_\ell} \pi(z_\ell) = 1$ and, because of Equations $(7.54)_{\curvearrowleft}$ and $(7.11)_{226}$, that for all $\gamma \in \Gamma_{\ell \rfloor z^*_{P(\ell)}}$:

$$\sum_{z_\ell \in \mathscr{X}_\ell} \pi(z_\ell)\gamma(z_\ell) = \frac{1}{\lambda} \sum_{z_\ell \in \mathscr{X}_\ell} p(z^*_K, z_\ell)\gamma(z_\ell) + \frac{\varepsilon}{\lambda} \sum_{z_\ell \in \mathscr{X}_\ell} p_{\ell \rfloor z^*_{P(\ell)}}(z_\ell)\gamma(z_\ell) \geq 0.$$

Since $\mathscr{F}_{\ell \rfloor z^*_{P(\ell)}}$ is by assumption completely characterised by Equations $(7.10)_{226}$ and $(7.11)_{226}$, we infer that $\pi \in \mathscr{F}_{\ell \rfloor z^*_{P(\ell)}}$, which implies that $\pi(z^*_\ell) \geq 0$ and therefore also that $\theta(z^*_\ell) \geq 0$, a contradiction. □

7.B PROOF OF THEOREM 92

Proof of Theorem 92$_{233}$. Consider four parameters $\underline{\tau}, \overline{\tau}, \underline{\pi}, \overline{\pi} \in \mathbb{R}$ such that $0 \leq \underline{\tau} \leq \overline{\tau} \leq 1$ and $0 \leq \underline{\pi} \leq \overline{\pi} \leq 1$ and let $a, b, c, d \in \mathbb{R}$ be four unknowns that satisfy the following constraints:

$$a + b + c + d = 1 \tag{7.55}$$

and

$$(1 - \underline{\tau})a \geq \underline{\tau}c \tag{i1}$$
$$\overline{\tau}c \geq (1 - \overline{\tau})a \tag{i2}$$
$$(1 - \underline{\tau})b \geq \underline{\tau}d \tag{i3}$$
$$\overline{\tau}d \geq (1 - \overline{\tau})b \tag{i4}$$
$$(1 - \underline{\pi})a \geq \underline{\pi}b \tag{i5}$$
$$\overline{\pi}b \geq (1 - \overline{\pi})a \tag{i6}$$
$$(1 - \underline{\pi})c \geq \underline{\pi}d \tag{i7}$$
$$\overline{\pi}d \geq (1 - \overline{\pi})c \tag{i8}$$

If the inequalities in Equations (i1)–(i8) are replaced by equalities, we refer to the resulting expressions as $(e1)_{\curvearrowright}$–$(e8)_{\curvearrowright}$:

$$(1-\underline{\tau})a = \underline{\tau}c \tag{e1}$$
$$\overline{\tau}c = (1-\overline{\tau})a \tag{e2}$$
$$(1-\underline{\tau})b = \underline{\tau}d \tag{e3}$$
$$\overline{\tau}d = (1-\overline{\tau})b \tag{e4}$$
$$(1-\underline{\pi})a = \underline{\pi}b \tag{e5}$$
$$\overline{\pi}b = (1-\overline{\pi})a \tag{e6}$$
$$(1-\underline{\pi})c = \underline{\pi}d \tag{e7}$$
$$\overline{\pi}d = (1-\overline{\pi})c \tag{e8}$$

It is a matter of straightforward verification that by replacing the parameters and unknowns in the constraints $(7.55)_{\frown}$ and $(i1)_{\frown}$–$(i8)_{\frown}$ by the ones given in column T1 of Table $(7.6)_{\frown}$, we obtain Equations $(7.17)_{231}$ and $(I1)_{231}$–$(I8)_{231}$. A perhaps more surprising result is that due to the symmetry of this set of constraints, the substitutions in columns T2–T8 of Table $(7.6)_{\frown}$ also yield Equations $(7.17)_{231}$ and $(I1)_{231}$–$(I8)_{231}$.

As explained in Section $7.4.4_{229}$, the extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ can be found by considering every possible subset of three equalities from $(E1)_{231}$–$(E8)_{231}$. For every such combination of three equalities, we need to combine them with the unitary constraint and check whether this results in a unique solution. If this is the case, we need to check whether this unique solution satisfies the inequalities in $(I1)_{231}$–$(I8)_{231}$. If yes, then that unique solution is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

For ease of notation, we will use (E123) to denote the system of constraints consisting of Equations $(E1)_{231}$, $(E2)_{231}$, $(E3)_{231}$ and $(7.17)_{231}$, and similarly for other combinations of Equation $(7.17)_{231}$ with a ternary subset of $(E1)_{231}$–$(E8)_{231}$. In the same way, the combination of Equations (e1), (e2), (e3) and $(7.55)_{\frown}$ is denoted by (e123) and similarly for other combinations of Equation $(7.55)_{\frown}$ with a ternary subset of (e1)–(e8).

We start by considering the following systems of equalities: (E135), (E136), (E247), (E248), (E157), (E368), (E257) and (E468). By applying the substitutions in Table $7.6_{\frown}$, these eight problems all become equivalent to (e135). If this system provides a unique solution, we need to check whether it satisfies $(i1)_{\frown}$–$(i8)_{\frown}$ and if it does, applying each of the substitutions T1–T8 to that solution yields a (possibly different) extreme point.

Equation (e1) tells us that $(1-\underline{\tau})a = \underline{\tau}c$, which is equivalent to $a = \underline{\tau}(a+c)$. Equation (e3) tells us that $(1-\underline{\tau})b = \underline{\tau}d$, which is equivalent to $b = \underline{\tau}(b+d)$. By combining them, we find that

$$a+b = \underline{\tau}(a+b+c+d) = \underline{\tau}, \tag{7.56}$$

where the last equality follows from Equation $(7.55)_{\frown}$. Equation (e5) tells us that $(1-\underline{\pi})a = \underline{\pi}b$, which is equivalent to $a = \underline{\pi}(a+b)$. By combining this

	T1	T2	T3	T4	T5	T6	T7	T8
$\underline{\tau}$	$\underline{p}(h_1)$	$\underline{p}(h_1)$	$\underline{p}(t_1)$	$\underline{p}(t_1)$	$\underline{p}(h_2)$	$\underline{p}(t_2)$	$\underline{p}(h_2)$	$\underline{p}(t_2)$
$\overline{\tau}$	$\overline{p}(h_1)$	$\overline{p}(h_1)$	$\overline{p}(t_1)$	$\overline{p}(t_1)$	$\overline{p}(h_2)$	$\overline{p}(t_2)$	$\overline{p}(h_2)$	$\overline{p}(t_2)$
$1-\overline{\tau}$	$\underline{p}(t_1)$	$\underline{p}(t_1)$	$\underline{p}(h_1)$	$\underline{p}(h_1)$	$\underline{p}(t_2)$	$\underline{p}(h_2)$	$\underline{p}(t_2)$	$\underline{p}(h_2)$
$1-\underline{\tau}$	$\overline{p}(t_1)$	$\overline{p}(t_1)$	$\overline{p}(h_1)$	$\overline{p}(h_1)$	$\overline{p}(t_2)$	$\overline{p}(h_2)$	$\overline{p}(t_2)$	$\overline{p}(h_2)$
$\underline{\pi}$	$\underline{p}(h_2)$	$\underline{p}(t_2)$	$\underline{p}(h_2)$	$\underline{p}(t_2)$	$\underline{p}(h_1)$	$\underline{p}(h_1)$	$\underline{p}(t_1)$	$\underline{p}(t_1)$
$\overline{\pi}$	$\overline{p}(h_2)$	$\overline{p}(t_2)$	$\overline{p}(h_2)$	$\overline{p}(t_2)$	$\overline{p}(h_1)$	$\overline{p}(h_1)$	$\overline{p}(t_1)$	$\overline{p}(t_1)$
$1-\overline{\pi}$	$\underline{p}(t_2)$	$\underline{p}(h_2)$	$\underline{p}(t_2)$	$\underline{p}(h_2)$	$\underline{p}(t_1)$	$\underline{p}(t_1)$	$\underline{p}(h_1)$	$\underline{p}(h_1)$
$1-\underline{\pi}$	$\overline{p}(t_2)$	$\overline{p}(h_2)$	$\overline{p}(t_2)$	$\overline{p}(h_2)$	$\overline{p}(t_1)$	$\overline{p}(t_1)$	$\overline{p}(h_1)$	$\overline{p}(h_1)$
a	$p(h_1,h_2)$	$p(h_1,t_2)$	$p(t_1,h_2)$	$p(t_1,t_2)$	$p(h_1,h_2)$	$p(h_1,t_2)$	$p(t_1,h_2)$	$p(t_1,t_2)$
b	$p(h_1,t_2)$	$p(h_1,h_2)$	$p(t_1,t_2)$	$p(t_1,h_2)$	$p(t_1,h_2)$	$p(t_1,t_2)$	$p(h_1,h_2)$	$p(h_1,t_2)$
c	$p(t_1,h_2)$	$p(t_1,t_2)$	$p(h_1,h_2)$	$p(h_1,t_2)$	$p(h_1,t_2)$	$p(h_1,h_2)$	$p(t_1,t_2)$	$p(t_1,h_2)$
d	$p(t_1,t_2)$	$p(t_1,h_2)$	$p(h_1,t_2)$	$p(h_1,h_2)$	$p(t_1,t_2)$	$p(t_1,h_2)$	$p(h_1,t_2)$	$p(h_1,h_2)$
$(7.55)_{280}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$	$(7.17)_{231}$
$(i1)_{280}$	$(I1)_{231}$	$(I3)_{231}$	$(I2)_{231}$	$(I4)_{231}$	$(I5)_{231}$	$(I6)_{231}$	$(I7)_{231}$	$(I8)_{231}$
$(i2)_{280}$	$(I2)_{231}$	$(I4)_{231}$	$(I1)_{231}$	$(I3)_{231}$	$(I6)_{231}$	$(I5)_{231}$	$(I8)_{231}$	$(I7)_{231}$
$(i3)_{280}$	$(I3)_{231}$	$(I1)_{231}$	$(I4)_{231}$	$(I2)_{231}$	$(I7)_{231}$	$(I8)_{231}$	$(I5)_{231}$	$(I6)_{231}$
$(i4)_{280}$	$(I4)_{231}$	$(I2)_{231}$	$(I3)_{231}$	$(I1)_{231}$	$(I8)_{231}$	$(I7)_{231}$	$(I6)_{231}$	$(I5)_{231}$
$(i5)_{280}$	$(I5)_{231}$	$(I6)_{231}$	$(I7)_{231}$	$(I8)_{231}$	$(I1)_{231}$	$(I3)_{231}$	$(I2)_{231}$	$(I4)_{231}$
$(i6)_{280}$	$(I6)_{231}$	$(I5)_{231}$	$(I8)_{231}$	$(I7)_{231}$	$(I2)_{231}$	$(I4)_{231}$	$(I1)_{231}$	$(I3)_{231}$
$(i7)_{280}$	$(I7)_{231}$	$(I8)_{231}$	$(I5)_{231}$	$(I6)_{231}$	$(I3)_{231}$	$(I1)_{231}$	$(I4)_{231}$	$(I2)_{231}$
$(i8)_{280}$	$(I8)_{231}$	$(I7)_{231}$	$(I6)_{231}$	$(I5)_{231}$	$(I4)_{231}$	$(I2)_{231}$	$(I3)_{231}$	$(I1)_{231}$
(e135)	(E135)	(E136)	(E247)	(E248)	(E157)	(E368)	(E257)	(E468)
(e137)	(E137)	(E138)	(E245)	(E246)	(E357)	(E168)	(E457)	(E268)
(e123)	(E123)	(E134)	(E124)	(E234)	(E567)	(E568)	(E578)	(E678)
(e125)	(E125)	(E346)	(E127)	(E348)	(E156)	(E356)	(E278)	(E478)
(e126)	(E126)	(E345)	(E128)	(E347)	(E256)	(E456)	(E178)	(E378)
(e167)	(E167)	(E358)	(E258)	(E467)	(E235)	(E146)	(E147)	(E238)
(e145)	(E145)	(E236)	(E237)	(E148)	(E158)	(E367)	(E267)	(E458)

Table 7.6: The symmetry in Equations $(7.17)_{231}$ and $(I1)_{231}$–$(I8)_{231}$

with Equation $(7.56)_{281}$ we find that $a = \underline{\tau}\,\underline{\pi}$, which, when we plug it back into Equation $(7.56)_{281}$, in turn implies that $b = \underline{\tau}(1-\underline{\pi})$. If $\underline{\tau} = 0$, then clearly the system (e135) does not define a unique solution. If $\underline{\tau} > 0$, then plugging the result for a in Equation $(e1)_{281}$ yields $c = (1-\underline{\tau})\underline{\pi}$ and plugging the result for b in Equation $(e3)_{281}$ yields $d = (1-\underline{\tau})(1-\underline{\pi})$. We can summarise our findings as follows: if $\underline{\tau} = 0$, then (e135) does not lead to a unique solution and if $\underline{\tau} > 0$, then the unique solution is

$$a = \underline{\tau}\,\underline{\pi},\, b = \underline{\tau}(1-\underline{\pi}),\, c = (1-\underline{\tau})\underline{\pi},\, d = (1-\underline{\tau})(1-\underline{\pi}). \tag{7.57}$$

It is a matter of straightforward verification that this solution satisfies the inequalities in $(i1)_{280}$–$(i8)_{280}$.

By applying each of the substitutions in Table $7.6_{\frown}$ to Equation (7.57), we obtain the following probability mass functions: columns T1 and T5 yield p_{S1}, columns T2 and T6 yield p_{S2}, columns T3 and T7 yield p_{S3} and columns T4 and T8 yield p_{S4}. For each of these substitutions, the corresponding probability mass function is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ if and only if the parameter that was used to substitute $\underline{\tau}$ with differs from zero.

Next, we consider the following systems of equalities: (E137), (E138), (E245), (E246), (E357), (E168), (E457) and (E268). By applying the substitutions in Table $(7.6)_{\frown}$, these eight problems all become equivalent to (e137). If this system provides a unique solution, we need to check whether it satisfies $(i1)_{280}$–$(i8)_{280}$ and if it does, applying each of the substitutions T1–T8 to that solution yields a (possibly different) extreme point.

Equation $(e1)_{281}$ tells us that $(1-\underline{\tau})a = \underline{\tau}c$, which is equivalent to, $c = (1-\underline{\tau})(a+c)$. Equation $(e3)_{281}$ tells us that $(1-\underline{\tau})b = \underline{\tau}d$, which is equivalent to $d = (1-\underline{\tau})(b+d)$. By combining them, we find that

$$c + d = (1-\underline{\tau})(a+b+c+d) = 1-\underline{\tau}, \tag{7.58}$$

where the last equality follows from Equation $(7.55)_{280}$. Equation $(e7)_{281}$ tells us that $(1-\underline{\pi})c = \underline{\pi}d$, which is equivalent to $c = \underline{\pi}(c+d)$. By combining this with Equation (7.58), we find that $c = (1-\underline{\tau})\underline{\pi}$, which, when we plug it back into Equation (7.58), in turn implies that $d = (1-\underline{\tau})(1-\underline{\pi})$. If $\underline{\tau} = 1$, then clearly the system (e137) does not define a unique solution. If $\underline{\tau} < 1$, then plugging the result for c in Equation $(e1)_{281}$ yields $a = \underline{\tau}\,\underline{\pi}$ and plugging the result for d in Equation $(e3)_{281}$ yields $b = \underline{\tau}(1-\underline{\pi})$. We can summarise our findings as follows: if $\underline{\tau} = 1$, then (e137) does not lead to a unique solution and if $\underline{\tau} < 1$, then the unique solution is

$$a = \underline{\tau}\,\underline{\pi},\, b = \underline{\tau}(1-\underline{\pi}),\, c = (1-\underline{\tau})\underline{\pi},\, d = (1-\underline{\tau})(1-\underline{\pi}). \tag{7.59}$$

Since this solution is identical to the solution found in Equation (7.57), we already know that it satisfies the inequalities in $(i1)_{280}$–$(i8)_{280}$.

By applying each of the substitutions in Table $7.6_{\frown}$ to Equation (7.59), we obtain the same probability mass functions that we found when solving (e135):

p_{S1}, p_{S2}, p_{S3} and p_{S4}. However, this time, these probability mass functions are extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ if and only if the parameter used to substitute $\underline{\tau}$ for differs from 1, whereas for the system (e135), they were extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ if and only if that same parameter differed from 0. Since that parameter cannot be 1 and 0 at the same time, we conclude that the probability mass functions p_{S1}, p_{S2}, p_{S3} and p_{S4} are always extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$, regardless of the particular values of the parameters $\underline{p}(h_1)$, $\overline{p}(h_1)$, $\underline{p}(t_1)$, $\overline{p}(t_1)$, $\underline{p}(h_2)$, $\overline{p}(h_2)$, $\underline{p}(t_2)$ and $\overline{p}(t_2)$. These probability mass functions have a special status: they are the extreme points of the strong product $\mathscr{F}_1 \times \mathscr{F}_2$ of $\mathscr{F}_1$ and $\mathscr{F}_2$, as defined by Equation $(7.18)_{234}$.

Consider now the special case where (at least) one of the local models $\mathscr{F}_i$, $i \in \{1,2\}$ is precise—$\underline{p}(h_i) = \overline{p}(h_i)$ and $\underline{p}(t_i) = \overline{p}(t_i)$—or vacuous—$[\underline{p}(h_i), \overline{p}(h_i)] = [\underline{p}(t_i), \overline{p}(t_i)] = [0,1]$. In that case, the strong product $\mathscr{F}_1 \times \mathscr{F}_2$ and independent natural extension $\mathscr{F}_1 \otimes \mathscr{F}_2$ coincide [46, Section 5.5] and therefore have the same extreme points, which implies that p_{S1}, p_{S2}, p_{S3} and p_{S4} are then the only extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$. This already explains the diagram in Figure 7.2_{235}. It is a matter of straightforward verification to see that as soon as (at least) one of the local models is precise, then some of the extreme points p_{S1}, p_{S2}, p_{S3} and p_{S4} coincide; it suffices to compare the formulas in Table 7.1_{233}. If both local models are precise, then all four extreme points even reduce to a single probability mass function, which is just the product of the two local ones.

So, from now on, without loss of generality, we can exclude the cases where (at least) one of the local models is precise or vacuous. Whatever the substitution that is chosen in Table 7.6_{282}, this always implies—among other things—that $\underline{\tau} \neq \overline{\tau}$ and $\underline{\pi} \neq \overline{\pi}$, which in turn implies that $\overline{\tau} \neq 0$, $1 - \underline{\tau} \neq 0$, $\overline{\pi} \neq 0$ and $1 - \underline{\pi} \neq 0$.

With that in mind, let us take a closer look at the systems of equations (e123), (e125) and (e126). Each of these systems corresponds to eight different combinations of the unitary constraint $(7.17)_{231}$ with three equalities out of $(E1)_{231}$–$(E8)_{231}$; see Table 7.6_{282}. The solution of each of these three systems needs to satisfy Equations $(e1)_{281}$ and $(e2)_{281}$, which implies that $(1-\underline{\tau})a = \underline{\tau}c$ and $\overline{\tau}c = (1-\overline{\tau})a$. By multiplying the first equation with $\overline{\tau}$, multiplying the second one with $\underline{\tau}$ and combining the resulting equations, we get $\overline{\tau}(1-\underline{\tau})a = \underline{\tau}(1-\overline{\tau})a$, or equivalently, $(\overline{\tau}-\underline{\tau})a = 0$, which is in turn equivalent with $a = 0$ since we already excluded the case $\underline{\tau} = \overline{\tau}$. By plugging this into Equation $(e2)_{281}$, we find that $\overline{\tau}c = 0$, which implies that $c = 0$ because we have already excluded the case $\overline{\tau} = 0$. So, for each of the systems of equalities (e123), (e125) and (e126), we have that $a = c = 0$.

For the system of equations (e123), Equations $(e1)_{281}$ and $(e2)_{281}$ are combined with Equations $(e3)_{281}$ and $(7.55)_{280}$. Equation $(e3)_{281}$ tells us that $(1-\underline{\tau})b = \underline{\tau}d$, which is equivalent with $b = \underline{\tau}(b+d)$. Combined with the unitary constraint $(7.55)_{280}$, this implies that $b = \underline{\tau}$ because $a = c = 0$. By plugging this back into the unitary constraint $(7.55)_{280}$, we find that $d = 1 - \underline{\tau}$.

The unique solution of (e123) is therefore

$$a = 0, b = \underline{\tau}, c = 0, d = 1 - \underline{\tau}. \tag{7.60}$$

This solution always satisfies Equations $(\text{i1})_{280}$, $(\text{i2})_{280}$, $(\text{i3})_{280}$, $(\text{i4})_{280}$, $(\text{i6})_{280}$ and $(\text{i8})_{280}$. It satisfies Equations $(\text{i5})_{280}$ and $(\text{i7})_{280}$ if and only if $0 \geq \underline{\pi}\,\underline{\tau}$ and $0 \geq \underline{\pi}(1-\underline{\tau})$, or equivalently, if $\underline{\pi} = 0$. The probability mass functions that are obtained by applying the substitutions in Table 7.6_{282} to Equation (7.60) will therefore be an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ if and only if the parameter that is used to substitute $\underline{\pi}$ with is equal to zero. However, these extreme points are already included in the ones that correspond to Equation $(7.59)_{283}$, because $\underline{\pi} = 0$ makes Equations (7.60) and $(7.59)_{283}$ identical. We conclude that the system of equations (e123) does not provide us with additional extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

For the system of equations (e125), Equations $(\text{e1})_{281}$ and $(\text{e2})_{281}$ are combined with Equations $(\text{e5})_{281}$ and $(7.55)_{280}$. Equation $(\text{e5})_{281}$ tells us that $(1-\underline{\pi})a = \underline{\pi}b$, which implies that $\underline{\pi}b = 0$ because $a = 0$. If $\underline{\pi} = 0$, this system has no unique solution. If $\underline{\pi} > 0$, then we find that $b = 0$ and $d = 1$, where the last equality follows from the unitary constraint $(7.55)_{280}$. However, this solution does not satisfy Equation $(\text{i7})_{280}$ because that would require that $(1-\underline{\pi})c \geq \underline{\pi}d$, or equivalently, since $c = 0$ and $d = 1$, that $0 \geq \underline{\pi}$, a contradiction. We conclude that the system of equations (e125) does not provide us with additional extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

For the system of equations (e126), Equations $(\text{e1})_{281}$ and $(\text{e2})_{281}$ are combined with Equations $(\text{e6})_{281}$ and $(7.55)_{280}$. Equation $(\text{e6})_{281}$ tells us that $\overline{\pi}b = (1-\overline{\pi})a$, which implies that $\overline{\pi}b = 0$ because $a = 0$. Since we already excluded $\overline{\pi} = 0$, we find that $b = 0$ and $d = 1$, where the last equality follows from the unitary constraint 7.55_{280}. By plugging this solution into Equations $(\text{i1})_{280}$–$(\text{i8})_{280}$, we find that these inequalities are satisfied if and only if $\underline{\tau} = \underline{\pi} = 0$, which implies that only in that case, the solution

$$a = 0, b = 0, c = 0, d = 1, \tag{7.61}$$

yields extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$, by applying the substitutions in Table 7.6_{282}. However, since $\underline{\tau} = \underline{\pi} = 0$, these extreme points are clearly already included in the extreme points that correspond to Equation $(7.59)_{283}$, because $\underline{\tau} = \underline{\pi} = 0$ makes Equations (7.61) and $(7.59)_{283}$ identical. We conclude that the system of equations (e126) does not provide us with additional extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

Next, we consider the system of equations (e167). The corresponding eight different combinations of the unitary constraint $(7.17)_{231}$ with three equalities out of $(\text{E1})_{231}$–$(\text{E8})_{231}$ can be found in Table 7.6_{282}.

Equation $(\text{e1})_{281}$ tells us that $(1-\underline{\tau})a = \underline{\tau}c$, Equation $(\text{e6})_{281}$ tells us that $\overline{\pi}b = (1-\overline{\pi})a$ and Equation $(\text{e7})_{281}$ tells us that $(1-\underline{\pi})c = \underline{\pi}d$. By combining

these three equations, we find that

$$(1-\underline{\pi})\overline{\pi}(1-\underline{\tau})b = \underline{\pi}(1-\overline{\pi})\underline{\tau}d. \tag{7.62}$$

Since Equation $(\text{i3})_{280}$ must be satisfied, we know that $(1-\underline{\tau})b \geq \underline{\tau}d$, which turns into $(1-\underline{\pi})\overline{\pi}(1-\underline{\tau})b \geq (1-\underline{\pi})\overline{\pi}\underline{\tau}d$ by multiplying both sides with $(1-\underline{\pi})\overline{\pi}$. In combination with Equation (7.62), this implies that $\underline{\pi}(1-\overline{\pi})\underline{\tau}d \geq (1-\underline{\pi})\overline{\pi}\underline{\tau}d$, or equivalently, that $(\underline{\pi}-\overline{\pi})\underline{\tau}d \geq 0$. Since we have already excluded the case $\underline{\pi} = \overline{\pi}$, it must be that $\underline{\tau}d \leq 0$, which implies that $d \leq 0$ or $\underline{\tau} = 0$.

If $d \leq 0$, it follows from Equation $(\text{e7})_{281}$ that $(1-\underline{\pi})c \leq 0$. Since we have already excluded the case $1-\underline{\pi} = 0$, this implies that $c \leq 0$. Due to Equation $(\text{e1})_{281}$, $c \leq 0$ in turn implies that $a \leq 0$ because we have already excluded the case $1-\underline{\tau} = 0$. Finally, due to Equation $(\text{e6})_{281}$, $a \leq 0$ in turn implies that $b \leq 0$ because we have already excluded the case $\overline{\pi} = 0$. We conclude that $a+b+c+d \leq 0$, which contradicts the unitary constraint $(7.55)_{280}$.

If $\underline{\tau} = 0$, it follows from Equation $(\text{e1})_{281}$ that $a = 0$. Therefore, it follows from Equation $(\text{e6})_{281}$ that $b = 0$ because we have already excluded the case $\overline{\pi} = 0$. From Equation $(\text{e7})_{281}$, we now derive that $c = \underline{\pi}(c+d) = \underline{\pi}$, where the last equality is due to the unitary constraint $(7.55)_{280}$, which also tells us that $d = 1-\underline{\pi}$. Therefore, we have found that

$$a = 0,\, b = 0,\, c = \underline{\pi},\, d = 1-\underline{\pi}. \tag{7.63}$$

However, this solution cannot yield additional extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ because, since $\underline{\tau} = 0$, Equations (7.63) and $(7.59)_{283}$ are identical.

At this point, we have already considered 48 of the 56 possible ternary subsets of $(\text{E1})_{231}$–$(\text{E8})_{231}$. So far, our conclusions are that, regardless of the particular values of the parameters $\underline{p}(h_1)$, $\overline{p}(h_1)$, $\underline{p}(t_1)$, $\overline{p}(t_1)$, $\underline{p}(h_2)$, $\overline{p}(h_2)$, $\underline{p}(t_2)$ and $\overline{p}(t_2)$, the probability mass functions p_{S1}, p_{S2}, p_{S3} and p_{S4} are always (possibly coinciding) extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$ and, as soon as (at least) one of the local models is either precise or vacuous, these are the only extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

Finally, we consider the system of equations (e145). The corresponding eight different combinations of the unitary constraint $(7.17)_{231}$ with three equalities out of $(\text{E1})_{231}$–$(\text{E8})_{231}$ are (E145), (E236), (E237), (E148), (E158), (E367), (E267) and (E458); see Table 7.6_{282}.

Equation $(\text{e1})_{281}$ says that $(1-\underline{\tau})a = \underline{\tau}c$, which is equivalent to $a = \underline{\tau}(a+c)$. Equation $(\text{e4})_{281}$ says that $\overline{\tau}d = (1-\overline{\tau})b$, which is equivalent to $b = \overline{\tau}(b+d)$. Equation $(\text{e5})_{281}$ tells us that $(1-\underline{\pi})a = \underline{\pi}b$. Combined with the unitary constraint $(7.55)_{280}$, this implies that

$$\begin{aligned}\underline{\tau}\overline{\tau}\underline{\pi} &= \underline{\tau}\overline{\tau}\underline{\pi}(a+b+c+d)\\ &= \underline{\tau}\overline{\tau}\underline{\pi}(b+d) + \underline{\tau}\overline{\tau}\underline{\pi}(a+c) = \underline{\tau}\,\underline{\pi}b + \overline{\tau}\underline{\pi}a = \underline{\tau}(1-\underline{\pi})a + \overline{\tau}\underline{\pi}a\\ &= (\underline{\tau}(1-\underline{\pi}) + \overline{\tau}\underline{\pi})a. \end{aligned} \tag{7.64}$$

Since we have already excluded the cases $1-\underline{\tau}=0$ and $1-\underline{\pi}=0$, the system (e145) clearly does not lead to a unique solution if $\underline{\tau}=\underline{\pi}=0$. If we assume that $\underline{\tau}+\underline{\pi}>0$, then since we have already excluded the cases $1-\underline{\pi}=0$ and $\overline{\tau}=0$, we find that $\underline{\tau}(1-\underline{\pi})+\overline{\tau}\underline{\pi}>0$, which allows us to infer from Equation $(7.64)_{\curvearrowleft}$ that

$$a=\frac{\underline{\tau}\overline{\tau}\underline{\pi}}{\underline{\tau}(1-\underline{\pi})+\overline{\tau}\underline{\pi}}. \tag{7.65}$$

If $\underline{\pi}>0$, we can combine this with Equation $(\text{e5})_{281}$ to find that

$$b=\frac{\underline{\tau}\overline{\tau}(1-\underline{\pi})}{\underline{\tau}(1-\underline{\pi})+\overline{\tau}\underline{\pi}}. \tag{7.66}$$

If $\underline{\pi}=0$, then $\underline{\tau}>0$ [because $\underline{\tau}+\underline{\pi}>0$] and $a=0$ [because of Equation (7.65)] and therefore, we infer from Equation $(\text{e1})_{281}$ that $c=0$. Since the unitary constraint $(7.55)_{280}$ then tells us that $b+d=0$, it follows from Equation $(\text{e4})_{281}$ that $b=\overline{\tau}$, which, with $\underline{\pi}=0$ and $\underline{\tau}>0$, is a special case of Equation (7.66). Since the cases $\underline{\pi}>0$ and $\underline{\pi}=0$ are exhaustive, it follows that—under the assumption that $\underline{\tau}+\underline{\pi}>0$—Equation (7.66) always holds.

If $\underline{\tau}>0$, we can combine Equation (7.65) and $(\text{e1})_{281}$ to find that

$$c=\frac{(1-\underline{\tau})\overline{\tau}\underline{\pi}}{\underline{\tau}(1-\underline{\pi})+\overline{\tau}\underline{\pi}}. \tag{7.67}$$

If $\underline{\tau}=0$, then $\underline{\pi}>0$ [because $\underline{\tau}+\underline{\pi}>0$] and $a=0$ [because of Equation (7.65)] and therefore, we infer from Equation $(\text{e5})_{281}$ that $b=0$. Since we have already excluded the case $\overline{\tau}=0$, Equation $(\text{e4})_{281}$ now tells us that $d=0$. By applying the unitary constraint $(7.55)_{280}$, it follows that $c=1$, which, with $\underline{\tau}=0$, $\underline{\pi}>0$ and $\overline{\tau}>0$, is a special case of Equation (7.67). Since the cases $\underline{\tau}>0$ and $\underline{\tau}=0$ are exhaustive, it follows that—under the assumption that $\underline{\tau}+\underline{\pi}>0$—Equation (7.67) always holds.

Finally, since we have already excluded the case $\overline{\tau}=0$, it follows from Equations (7.66) and $(\text{e4})_{281}$ that

$$d=\frac{\underline{\tau}(1-\overline{\tau})(1-\underline{\pi})}{\underline{\tau}(1-\underline{\pi})+\overline{\tau}\underline{\pi}}. \tag{7.68}$$

We can summarise our findings as follows. If $\underline{\tau}=\underline{\pi}=0$, then (e145) does not lead to a unique solution. If $\underline{\tau}+\underline{\pi}>0$ ($\underline{\tau}>0$ and/or $\underline{\pi}>0$), the unique solution of (e145) is given by Equations (7.65), (7.66), (7.66) and (7.68). Furthermore, in that case, by substituting this solution into Equations $(\text{i1})_{280}$–$(\text{i8})_{280}$, one can see that Equations $(\text{i1})_{280}$–$(\text{i7})_{280}$ are all satisfied and that Equation $(\text{i8})_{280}$ is satisfied if and only if

$$\underline{\tau}(1-\overline{\tau})\overline{\pi}(1-\underline{\pi})\geq\overline{\tau}(1-\underline{\tau})\underline{\pi}(1-\overline{\pi}). \tag{7.69}$$

Therefore, if this inequality holds and if $\underline{\tau}+\underline{\pi}>0$, then by applying the substitutions in Table 7.6$_{282}$ to Equations (7.65)$_{\frown}$, (7.66)$_{\frown}$, (7.66)$_{\frown}$ and (7.68)$_{\frown}$, we obtain eight new extreme points of $\mathscr{F}_1 \otimes \mathscr{F}_2$.

Using the substitutions in columns T1, T2, T3 and T4, we find that if

$$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2) \geq \overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2) \tag{7.70}$$

then

- p_{A1} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(h_1)=0$ and $\underline{p}(h_2)=0$;
- p_{A2} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(h_1)=0$ and $\underline{p}(t_2)=0$;
- p_{A3} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(t_1)=0$ and $\underline{p}(h_2)=0$;
- p_{A4} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(t_1)=0$ and $\underline{p}(t_2)=0$.

Similarly, using the substitution in column T5, T6, T7 and T8, we find that if

$$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2) \leq \overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2) \tag{7.71}$$

then

- p_{B1} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(h_1)=0$ and $\underline{p}(h_2)=0$;
- p_{B2} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(h_1)=0$ and $\underline{p}(t_2)=0$;
- p_{B3} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(t_1)=0$ and $\underline{p}(h_2)=0$;
- p_{B4} is an extreme point of $\mathscr{F}_1 \otimes \mathscr{F}_2$ unless $\underline{p}(t_1)=0$ and $\underline{p}(t_2)=0$.

We can now explain the second part of the diagram, which is depicted in Figures 7.3$_{236}$ and 7.4$_{237}$. As we have already explained, for this second part, we already know that p_{S1}, p_{S2}, p_{S3} and p_{S4} are extreme points $\mathscr{F}_1 \otimes \mathscr{F}_2$. Furthermore, we may assume that neither of the local models is precise or vacuous. For each $i \in \{1,2\}$, this implies that $\overline{p}(h_i)>0$ and $\overline{p}(t_i)>0$ and that $\underline{p}(h_i)$ and $\underline{p}(t_i)$ cannot both be zero.

If $\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2) > \overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2)$, then Equation (7.70) is satisfied and Equation (7.71) is not. The left-hand side of this inequality can not be zero since this would force it to be equal to the right-hand side. Therefore we have that $\underline{p}(h_1)>0$ and $\underline{p}(t_1)>0$. This implies that $\mathscr{F}_1 \otimes \mathscr{F}_2$ has four additional extreme points: p_{A1}, p_{A2}, p_{A3} and p_{A4}. If $\underline{p}(h_2)=0$, then one can check in Tables 7.1$_{233}$ and 7.2$_{234}$ that p_{A1} coincides with p_{S3} and that p_{A3} coincides with p_{S1}, which leaves only p_{A2} and p_{A4} as actual additional extreme points. If $\underline{p}(t_2)=0$, then p_{A2} coincides with p_{S4} and p_{A4} coincides with p_{S2}, which leaves only p_{A1} and p_{A3} as actual additional extreme points. $\underline{p}(h_2)$ and $\underline{p}(t_2)$ cannot both be zero because this would imply that $\mathscr{F}_2$ is a vacuous model, which is a case that was already excluded earlier on in the diagram.

If $\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2) < \overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2)$, then Equation $(7.71)_{\frown}$ is satisfied and Equation $(7.70)_{\frown}$ is not. The right-hand side of this inequality can not be zero since this would force it to be equal to the left-hand side. Therefore we have that $\underline{p}(h_2) > 0$ and $\underline{p}(t_2) > 0$. This implies that $\mathscr{F}_1 \otimes \mathscr{F}_2$ has four additional extreme points: $p_{\mathrm{B}1}$, $p_{\mathrm{B}2}$, $p_{\mathrm{B}3}$ and $p_{\mathrm{B}4}$. If $\underline{p}(h_1) = 0$, then one can check in Tables 7.1_{233} and 7.2_{234} that $p_{\mathrm{B}1}$ coincides with $p_{\mathrm{S}2}$ and that $p_{\mathrm{B}2}$ coincides with $p_{\mathrm{S}1}$, which leaves only $p_{\mathrm{B}3}$ and $p_{\mathrm{B}4}$ as actual additional extreme points. If $\underline{p}(t_1) = 0$, then $p_{\mathrm{B}3}$ coincides with $p_{\mathrm{S}4}$ and $p_{\mathrm{B}4}$ coincides with $p_{\mathrm{S}3}$, which leaves only $p_{\mathrm{B}1}$ and $p_{\mathrm{B}2}$ as actual additional extreme points. $\underline{p}(h_1)$ and $\underline{p}(t_1)$ cannot both be zero because this would imply that $\mathscr{F}_1$ is a vacuous model, which is a case that was already excluded earlier on in the diagram.

If

$$\underline{p}(h_1)\underline{p}(t_1)\overline{p}(h_2)\overline{p}(t_2) = \overline{p}(h_1)\overline{p}(t_1)\underline{p}(h_2)\underline{p}(t_2), \tag{7.72}$$

then Equations $(7.70)_{\frown}$ and $(7.71)_{\frown}$ are both satisfied. The diagram distinguishes between three cases: $\underline{p}(h_1) = 0$ or $\underline{p}(t_1) = 0$ or neither of them equal to zero. The case $\underline{p}(h_1) = \underline{p}(t_1) = 0$ is excluded because this would imply that $\mathscr{F}_1$ is a vacuous model.

If $\underline{p}(h_1) \neq 0$ and $\underline{p}(t_1) \neq 0$, then since we already excluded $\overline{p}(h_2) = 0$ and $\overline{p}(t_2) = 0$, it must hold that $\underline{p}(h_2) \neq 0$ and $\underline{p}(t_2) \neq 0$ because otherwise the right-hand side of Equation (7.72) would be zero whereas the left-hand side is not. We find that $\mathscr{F}_1 \otimes \mathscr{F}_2$ has eight additional extreme points: $p_{\mathrm{A}1}$, $p_{\mathrm{A}2}$, $p_{\mathrm{A}3}$, $p_{\mathrm{A}4}$, $p_{\mathrm{B}1}$, $p_{\mathrm{B}2}$, $p_{\mathrm{B}3}$ and $p_{\mathrm{B}4}$. However, some of them coincide: it follows from Equation (7.72) that $p_{\mathrm{A}1} = p_{\mathrm{A}4}$, $p_{\mathrm{A}2} = p_{\mathrm{A}3}$, $p_{\mathrm{B}1} = p_{\mathrm{B}4}$ and $p_{\mathrm{B}2} = p_{\mathrm{B}3}$. If we also have that $\mathscr{F}_1 = \mathscr{F}_2$, then $\underline{p}(h_1) = \underline{p}(h_2)$, $\overline{p}(h_1) = \overline{p}(h_2)$, $\underline{p}(t_1) = \underline{p}(t_2)$ and $\overline{p}(t_1) = \overline{p}(t_2)$. It is a matter of straightforward verification to see that then $p_{\mathrm{A}1} = p_{\mathrm{A}4} = p_{\mathrm{B}1} = p_{\mathrm{B}4}$ and $p_{\mathrm{A}2} = p_{\mathrm{A}3} = p_{\mathrm{B}2} = p_{\mathrm{B}3}$.

If $\underline{p}(h_1) = 0$ or $\underline{p}(t_1) = 0$, then in order for Equation (7.72) to hold, the right-hand side of that equality must be zero. Since we already excluded $\overline{p}(h_1) = 0$ and $\overline{p}(t_1) = 0$, this means that $\underline{p}(h_2) = 0$ or $\underline{p}(t_2) = 0$. They cannot both be zero because this would imply that $\mathscr{F}_2$ is a vacuous model, which is a case that was already excluded earlier on in the diagram. There are now four options left.

If $\underline{p}(h_1) = 0$ and $\underline{p}(h_2) = 0$, we find that $\mathscr{F}_1 \otimes \mathscr{F}_2$ has six additional extreme points: $p_{\mathrm{A}2}$, $p_{\mathrm{A}3}$, $p_{\mathrm{A}4}$, $p_{\mathrm{B}2}$, $p_{\mathrm{B}3}$ and $p_{\mathrm{B}4}$. However, some of them coincide with each other and/or with the extreme points that we already know. It follows from $\underline{p}(h_1) = 0$ and $\underline{p}(h_2) = 0$ that $p_{\mathrm{S}1} = p_{\mathrm{A}2} = p_{\mathrm{A}3} = p_{\mathrm{B}2} = p_{\mathrm{B}3}$ and $p_{\mathrm{A}4} = p_{\mathrm{B}4}$.

If $\underline{p}(h_1) = 0$ and $\underline{p}(t_2) = 0$, then $\mathscr{F}_1 \otimes \mathscr{F}_2$ has the following six additional extreme points: $p_{\mathrm{A}1}$, $p_{\mathrm{A}3}$, $p_{\mathrm{A}4}$, $p_{\mathrm{B}1}$, $p_{\mathrm{B}3}$ and $p_{\mathrm{B}4}$. Again, some of them coincide with each other and/or with the extreme points that we already know. It follows from $\underline{p}(h_1) = 0$ and $\underline{p}(t_2) = 0$ that $p_{\mathrm{S}2} = p_{\mathrm{A}1} = p_{\mathrm{A}4} = p_{\mathrm{B}1} = p_{\mathrm{B}4}$ and $p_{\mathrm{A}3} = p_{\mathrm{B}3}$.

If $\underline{p}(t_1) = 0$ and $\underline{p}(h_2) = 0$, then $\mathscr{F}_1 \otimes \mathscr{F}_2$ has the following six additional extreme points: p_{A1}, p_{A2}, p_{A4}, p_{B1}, p_{B2} and p_{B4}. Again, some of them coincide with each other and/or with the extreme points that we already know. It follows from $\underline{p}(t_1) = 0$ and $\underline{p}(h_2) = 0$ that $p_{\mathrm{S3}} = p_{\mathrm{A1}} = p_{\mathrm{A4}} = p_{\mathrm{B1}} = p_{\mathrm{B4}}$ and $p_{\mathrm{A2}} = p_{\mathrm{B2}}$.

If $\underline{p}(t_1) = 0$ and $\underline{p}(t_2) = 0$, then $\mathscr{F}_1 \otimes \mathscr{F}_2$ has the following six additional extreme points: p_{A1}, p_{A2}, p_{A3}, p_{B1}, p_{B2} and p_{B3}. Again, some of them coincide with each other and/or with the extreme points that we already know. It follows from $\underline{p}(t_1) = 0$ and $\underline{p}(t_2) = 0$ that $p_{\mathrm{S4}} = p_{\mathrm{A2}} = p_{\mathrm{A3}} = p_{\mathrm{B2}} = p_{\mathrm{B3}}$ and $p_{\mathrm{A1}} = p_{\mathrm{B1}}$. □

8

CONCLUSIONS

"The only relevant thing is uncertainty—the extent of our knowledge and ignorance. The actual fact of whether or not the events considered are in some sense determined, or known by other people, and so on, is of no consequence."

Bruno de Finetti

The main conclusion of this dissertation is that credal networks under epistemic irrelevance satisfy surprisingly many powerful theoretical properties, and that these properties can be exploited to develop efficient exact inference algorithms, for large classes of inference problems that were previously presumed to be intractable. Since many of these inference problems are NP-hard in credal networks under strong independence, our results turn credal networks under epistemic irrelevance into a serious, practically feasible alternative that should enable practitioners to solve real-life problems for which the corresponding necessary inferences were hitherto regarded as intractable.

More generally, there are also some interesting lessons to be drawn from how we obtained our results. The most important lesson is that there is much to be gained from considering different imprecise probability frameworks simultaneously. Each of these frameworks has its own specific features and, by combining them, we can take advantage of their respective merits within a single theory.

The framework of sets of desirable gambles has been our theoretical and philosophical workhorse. On the one hand, it endows lower previsions with

a clear behavioural interpretation that does not require an assumption of ideal precision. As explained in Chapter 3_{69}, this interpretation can be used to justify the use of regular and natural extension as conservative updating rules. On the other hand, sets of desirable gambles have proved to be an extremely powerful theoretical tool that excels both in elegance and mathematical generality. Perhaps most importantly: the usual issues with conditioning on events with probability zero are completely non-existing in this framework. It should therefore not be surprising that all of our main technical results were first proved in terms of sets of desirable gambles, only to be—often trivially—translated to other frameworks afterwards. Working the other way around would, in all likelihood, have been much more difficult.

Sets of probabilities—credal sets—reside at the other end of the spectrum. Centuries of probability theory have endowed us with an instinctive intuition about them that is hard to compete with. Given that practically everyone grasps the concept of probability—or at least some primitive notion of it—the educational purpose of this framework should clearly not be underestimated. Although I have done my very best to promote various alternatives that have clear advantages, I must admit that at the end of the day, I often think in terms of probabilities. They provide many of the concepts in this dissertation with a valuable intuition and will most likely remain the most important tool for elicitation for a long time to come. On a more technical side, in our specific case, they have enabled us to reformulate inference as a linear optimisation task and to characterise the independent natural extension of two binary models by means of the extreme points of its credal set.

Sets of linear previsions are the least important framework in this dissertation and, with hindsight, perhaps redundant. Their main advantage is that they allow us to clarify—intuitively—and simplify—mathematically—the link between credal sets and lower previsions, by serving as an intermediate step.

Lower previsions have been our main algorithmic tool. Given that inference is all about computing bounds, having a language that has these bounds as its primitive concepts is clearly beneficial in order to develop and present inference algorithms. It is therefore no coincidence that the efficient recursive algorithms in this dissertation are all expressed in terms of lower previsions. On the philosophical side, the main advantage of this framework is that these 'bounds'—lower and upper previsions—can be interpreted in two different ways. On the one hand, they are lower and upper expectations—lower and upper bounds on precise expectations. On the other hand, they are supremum buying prices and infimum selling prices. In this way, lower and upper previsions provide a unifying language that connects the framework of sets of desirable gambles to that of credal sets.

The features of these four different frameworks are not specific to our particular context. We are convinced that combining their respective advantages within a single theory could prove beneficial in other contexts as well, such as, for example, statistics with imprecise probabilities.

We conclude this dissertation by briefly discussing some avenues for further research that we consider to be promising, ranging from applications of our results, over possible extensions of the ideas in this dissertation, to fundamentally new concepts.

As far as applications of our results are concerned, we see plenty of directions for further research. The inference algorithms in this dissertation should already allow practitioners to solve large classes of inference problems that are relevant to their applications. To give but one example: the algorithms in Sections $7.5.5_{258}$ and $7.6.1_{271}$ can be directly applied to solve classification problems. However, this requires two additional steps. First of all, it would be necessary to implement our algorithms and to develop user-friendly software that computes inferences with them; no such software currently exists. Secondly, it should be thoroughly tested whether the inferences that are computed by our algorithms are informative enough to be useful in practice. Since epistemic irrelevance imposes less stringent constraints than complete, strong, or epistemic independence, the inferences of a credal network under epistemic irrelevance will be more conservative than those that correspond to other types of credal networks, and possibly too conservative to be of practical use. Although we did not observe such behaviour in the OCR application that was mentioned in Section $7.5.7_{269}$ [30], it remains to be seen whether this will be the case in other applications as well.

Besides applying the algorithms in this dissertation to real-life problems, another important avenue for future research would be to develop new algorithms, for inference problems that were not yet considered here. The theoretical tools in Chapter 6_{152}, together with the examples in Chapter 7_{211}, should enable theoretically oriented researchers to develop such algorithms. However, given that general inference in credal networks under epistemic irrelevance is $\mathrm{NP}^{\mathrm{PP}}$-hard [67], there will be inference problems for which it is not possible to develop efficient exact algorithms. In order to be able to deal with these inference problems anyway, the development of approximate algorithms is crucial. A first idea could be to add additional arrows to a network until it becomes recursively decomposable. It is not hard to see that this is always possible. In the case of Figure 1.1_{26}, we could for example add an arrow from 'Hayfever' to 'Flu', thereby making this simple network recursively decomposable. This arrow should be virtual, in the sense that it does not affect the local models. Its only purpose is to remove some of the assessments of irrelevance that are imposed by the graph. For example, although the local model for 'Flu' would then formally depend on the value of 'Hayfever' and 'Season', it would in practice only depend on the value of 'Season'—would remain constant if the value of 'Hayfever' is changed. The resulting irrelevant natural extension will produce inferences that provide lower bounds for those of the original model and, since the network is now recursively decomposable, this approximation can be computed recursively. By reversing the direction of this new virtual arrow, we obtain a second approximation. For every inference, the best of these

two approximations—the highest lower bound—can then be considered. This idea can be extended beyond this simple example and, in this way, it is possible to develop approximate algorithms. The main bottlenecks are that (a) there might be exponentially many different ways in which these virtual arrows can be added and (b) the number of virtual arrows that needs to be added could be large, which would increase the value of the parameter $\max_{s \in G} |P(K_s)|$ and, therefore, decrease the computational efficiency of this approximation method. Nevertheless, I believe that this idea is worth exploring.

Instead of developing additional—exact or approximate—inference algorithms for the notion of irrelevant natural extension that we considered here, it would also be interesting to try and generalise this notion itself, for example by relaxing the assumption that the variables in the network can only take a finite number of values. Since variables that can take infinitely many values—for example natural- or real-valued variables—are frequently used in real-life applications, being able to deal with them would clearly be beneficial. One way of doing so could be to focus on some finite partition of the state space. However, it would be more elegant to deal with infinite state spaces directly. I believe that investigating whether or not this is possible and, if yes, to what extent, could be a very interesting line of future research. I do not expect major problems in defining the irrelevant natural extension. I believe that in the infinite case, the choices that we have made here, which are using epistemic h-irrelevance instead of epistemic value-irrelevance, and (a slightly adapted version of) W-coherence (Williams-coherence) instead of Walley-coherence, will enable us to avoid the typical problems with Walley-coherence (most importantly the fact that the natural extension might not exist [106, Appendix K]) as well as the problem that the notion of natural extension that corresponds to W-coherence is sometimes too weak; see Footnote 9_{116}. Extending some of the theoretical properties in Chapter 6_{152}—for example Theorem 53_{156}—will most likely be much harder. Defining conditioning rules and justifying them as updating rules could also be tricky, especially if the state space is uncountably infinite and the conditioning event is a singleton; in that case, I expect this to require a limit argument, as it does in standard measure-theoretic probability. The most challenging task would be the development of inference algorithms for the infinite case; this seems far from trivial and I expect this to require significant amounts of additional research. Nevertheless, if successful, it would be well worth the effort.

A final avenue for future research, which I believe most promising, is to develop a new type of credal networks, called *mixed credal networks*. Instead of adopting a single notion of independence, as current theories of credal networks do, the basic idea is to use a mix of different notions of independence. By allowing assessments of different notions of independence to be combined within a single model, we obtain a richer theory that is able to represent a wider range of structural assessments. At first sight, this might come across as a nice but nevertheless completely unrealistic idea. Given that mixed credal networks

include conventional credal networks as special cases, it would seem obvious that inference is bound to become even harder than it already is for the types of credal networks that have been considered so far. However, I believe that this is not the case. In fact, I am convinced that the standard types of credal networks that are currently being used are among the hard instances, and that there are specific types of mixed credal networks for which inference will turn out to be a much simpler problem, to the point that it will be possible to develop inference algorithms whose computational complexity is comparable to that of inference algorithms for Bayesian networks. The use of the specific types of mixed credal networks for which I believe this to be possible can even be justified on philosophical grounds. I hope that I will be given the opportunity to further explore this idea in the context of postdoctoral research.

BIBLIOGRAPHY

[1] Alessandro Antonucci, Ralph Brühlmann, Alberto Piatti, and Marco Zaffalon. Credal networks for military identification problems. *International Journal of Approximate Reasoning*, 50(4):666–679, 2009.

[2] Alessandro Antonucci, Cassio P. de Campos, and Marco Zaffalon. Probabilistic graphical models. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 207–229. John Wiley & Sons, Chichester, 2014.

[3] Alessandro Antonucci, Andrea Salvetti, and Marco Zaffalon. Credal networks for hazard assessment of debris flows. In Jürgen P. Kropp and Jürgen Scheffran, editors, *Advanced Methods for Decision Making and Risk Management in Sustainability Science*, pages 237–256. NOVA Science Publishers, New York, 2007.

[4] Thomas Augustin, Gero Walter, and Frank P. A. Coolen. Statistical inference. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. John Wiley & Sons, Chichester, 2014.

[5] Cédric Baudrit, Mariette Sicard, Pierre-Henri Wuillemin, and Nathalie Perrot. Towards a global modelling of the Camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic Bayesian networks. *Journal of Food Engineering*, 98(3):283–293, 2010.

[6] Alessio Benavoli, Marco Zaffalon, and Enrique Miranda. Robust filtering through coherent lower previsions. *IEEE Transactions on Automatic Control*, 56(7):1567–1581, 2011.

[7] James O. Berger. *Statistical decision theory and Bayesian analysis*. Springer series in statistics. Springer, New York, NY, 1985.

[8] Jean-Marc Bernard. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning*, 39(2-3):123–150, 2005.

[9] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to linear optimization*. Athena Scientific, 1997.

[10] Richard Bradley. The kinematics of belief and desire. *Synthese*, 156(3):513–535, 2007.

[11] Seamus Bradley and Katie Steele. Uncertainty, learning, and the "problem" of dilation. *Erkenntnis*, 79(6):1287–1303, 2013.

[12] Marco E. G. V. Cattaneo. A continuous updating rule for imprecise probabilities. In Anne Laurent, Olivier Strauss, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 426–435. Springer International Publishing, 2014.

[13] Giorgio Corani, Joaquín Abellán, Andrés Masegosa, Serafín Moral, and Marco Zaffalon. Classification. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 230–257. John Wiley & Sons, Chichester, 2014.

[14] Inés Couso and Serafín Moral. Sets of desirable gambles: conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52(7):1034–1055, 2011.

[15] Inés Couso, Serafín Moral, and Peter Walley. Examples of Independence for Imprecise Probabilities. In Gert de Cooman, Fabio G. Cozman, Serafín Moral, and Peter Walley, editors, *ISIPTA '99: Proceedings of the First International Symposium on Imprecise Probabilities and their Applications*, pages 121–130. Ghent, 1999.

[16] Fabio G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[17] Fabio G. Cozman. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, 39(2):167–184, 2005.

[18] Fabio G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[19] Fabio G. Cozman. Independence for sets of full conditional probabilities, sets of lexicographic probabilities, and sets of desirable gambles. In Fabio G. Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors, *ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*, pages 87–97. SIPTA, Compiègne, 2013.

[20] Fabio G. Cozman and Cassio P. de Campos. Kuznetsov independence for interval-valued expectations and sets of probability distributions: Properties and algorithms. *International Journal of Approximate Reasoning*, 55(2):666–682, 2014.

[21] Fabio G. Cozman and Teddy Seidenfeld. Independence for full conditional measures and their graphoid properties. In Benedikt Löwe, Eric Pacuit, and Jan-Willem Romeijn, editors, *Foundations of the Formal Sciences VI: Reasoning about Probabilities and Probabilistic Reasoning*, pages 1–29. College Publications, London, 2010.

[22] Fabio G. Cozman and Peter Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45(1-2):173–195, 2005.

[23] A. Philip Dawid. Conditional independence. In *Encyclopedia of Statistical Sciences*, pages 146–155. Wiley-Interscience, 1997.

[24] Jasper De Bock, Cassio P. de Campos, and Alessandro Antonucci. Global sensitivity analysis for MAP inference in graphical models. In *Advances in Neural Information Processing Systems 27*, pages 2690–2698. 2014.

[25] Jasper De Bock and Gert de Cooman. State sequence prediction in imprecise hidden Markov models. In Frank P. A. Coolen, Gert de Cooman, Thomas Fetz, and Michael Oberguggenberger, editors, *ISIPTA '11: Proceedings of the Seventh International Symposium on Imprecise Probability Theories and Applications*, pages 149–158. SIPTA, Innsbruck, 2011.

[26] Jasper De Bock and Gert de Cooman. Imprecise Bernoulli processes. In Salvatore Greco, Bernadette Bouchon-Meunier, Giulianella Coletti, Mario Fedrizzi, Benedetto Matarazzo, and Ronald R. Yager, editors, *Advances in Computational Intelligence (Proceedings of IPMU 2012)*, pages 400–409. Springer, Berlin, 2012.

[27] Jasper De Bock and Gert de Cooman. Allowing for probability zero in credal networks under epistemic irrelevance. In Fabio G. Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors, *ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*, pages 109–118. SIPTA, Compiègne, 2013.

[28] Jasper De Bock and Gert de Cooman. Credal networks under epistemic irrelevance using sets of desirable gambles. In Fabio G. Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors,

ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications, pages 99–108. SIPTA, Compiègne, 2013.

[29] Jasper De Bock and Gert de Cooman. Extreme lower previsions and Minkowski indecomposability. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (Proceedings of ECSQARU 2013)*, pages 157–168. Springer, Berlin, 2013.

[30] Jasper De Bock and Gert de Cooman. An efficient algorithm for estimating state sequences in imprecise hidden Markov models. *Journal of Artificial Intelligence Research*, 50:189–233, 2014.

[31] Jasper De Bock and Gert de Cooman. Credal networks under epistemic irrelevance: the sets of desirable gambles approach. *International Journal of Approximate Reasoning*, 56:178–207, 2015.

[32] Jasper De Bock and Gert de Cooman. Extreme lower previsions. *Journal of Mathematical Analysis and Applications*, 421(2):1042–1080, 2015.

[33] Jasper De Bock, Arthur Van Camp, Márcio Alves Diniz, and Gert de Cooman. Representation theorems for partially exchangeable random variables. 2014. Accepted for publication in Fuzzy Sets and Systems.

[34] Cedric De Boom, Jasper De Bock, Arthur Van Camp, and Gert de Cooman. Robustifying the Viterbi algorithm. In *Probabilistic Graphical Models*, pages 160–175. Springer International Publishing, 2014.

[35] Cassio P. de Campos and Fabio G. Cozman. Computing lower and upper expectations under epistemic independence. *International Journal of Approximate Reasoning*, 44(3):244–260, 2007.

[36] Cassio P. de Campos, Lei Zhang, Yan Tong, and Qiang Ji. Semi-qualitative probabilistic networks in computer vision problems. *Journal of Statistical Theory and Practice*, 3(1):197–210, 2009.

[37] Luis M. de Campos and Serafín Moral. Independence concepts for convex sets of probabilities. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*, pages 108–115. Morgan Kaufmann, Montreal, 1995.

[38] Gert de Cooman. Belief models: an order-theoretic investigation. *Annals of Mathematics and Artificial Intelligence*, 45(1-2):5–34, 2005.

[39] Gert de Cooman, Jasper De Bock, and Márcio Alves Diniz. Predictive inference under exchangeability, and the imprecise Dirichlet multinomial model. In *Proceedings of the 12th Brazilian Meeting on Bayesian Statistics (EBEB)*, pages 13–33. 2014.

[40] Gert de Cooman, Jasper De Bock, and Márcio Alves Diniz. Coherent predictive inference under exchangeability with imprecise probabilities. *Journal of Artificial Intelligence Research*, 52:1–95, 2015.

[41] Gert de Cooman, Jasper De Bock, and Arthur Van Camp. Recent advances in imprecise-probabilistic graphical models. In *Frontiers in Artificial Intelligence and Applications, Volume 242 (Proceedings of ECAI 2012)*, pages 27–32. 2012.

[42] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9):1029–1052, 2010.

[43] Gert de Cooman, Filip Hermans, and Erik Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009.

[44] Gert de Cooman and Enrique Miranda. Forward irrelevance. *Journal of Statistical Planning and Inference*, 139(2):256–276, 2009.

[45] Gert de Cooman and Enrique Miranda. Irrelevant and independent natural extension for sets of desirable gambles. *Journal of Artificial Intelligence Research*, 45:601–640, 2012.

[46] Gert de Cooman, Enrique Miranda, and Marco Zaffalon. Independent natural extension. *Artificial Intelligence*, 175(12):1911–1950, 2011.

[47] Gert de Cooman and Erik Quaeghebeur. Exchangeability and sets of desirable gambles. *International Journal of Approximate Reasoning*, 53(3):363–395, 2012.

[48] Gert de Cooman, Matthias C. M. Troffaes, and Enrique Miranda. n-Monotone lower previsions. *Journal of Intelligent and Fuzzy Systems*, 16(4):253–263, 2005.

[49] Gert de Cooman and Marco Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1-2):75–125, 2004.

[50] Bruno De Finetti. *Teoria delle probabilità*. Einaudi, Turin, 1970. English translation: [51].

[51] Bruno De Finetti. Theory of probability. John Wiley & Sons, New York, 1974.

[52] Arthur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967.

[53] Sébastien Destercke and Didier Dubois. Other uncertainty theories based on capacities. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 93–113. John Wiley & Sons, Chichester, 2014.

[54] Lester E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3(1):89–99, 1975.

[55] Enrico Fagiuoli and Marco Zaffalon. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.

[56] Peter Gärdenfors. *Knowledge in flux: modeling the dynamics of epistemic states*. MIT Press, Cambridge, 1988.

[57] Dan Geiger, Thomas Verma, and Judea Pearl. d-Separation: from theorems to algorithms. In *Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence*, pages 139–148. Elsevier, 1990.

[58] Dan Geiger, Thomas Verma, and Judea Pearl. Identifying independence in Bayesian networks. *Networks*, 20(5):507–534, 1990.

[59] Itzhak Gilboa and David Schmeidler. Updating ambiguous beliefs. *Journal of Economic Theory*, 59(1):33–49, 1993.

[60] Branko Grünbaum. *Convex polytopes*. Springer Science & Business Media, second edition, 2003.

[61] Peter J. Hammond. Elementary non-archimedean representations of probability for decision theory and games. In *Patrick Suppes: Scientific Philosopher*, pages 25–61. Springer, Dordrecht, 1994.

[62] Matthieu Hourbracq, Cédric Baudrit, Pierre-Henri Wuillemin, and Sébastien Destercke. Dynamic credal networks: introduction and use in robustness analysis. In Fabio G. Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors, *ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*, pages 159–169. SIPTA, Compiègne, 2013.

[63] Peter J. Huber. *Robust statistics*. John Wiley & Sons, New York, 1981.

[64] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

[65] Elmar Kriegler. Updating under unknown unknowns: an extension of Bayes' rule. *International Journal of Approximate Reasoning*, 50(4):583–596, 2009.

[66] Isaac Levi. *The enterprise of knowledge: an essay on knowledge, credal probability, and chance*. MIT Press, 1983.

[67] Denis D. Mauá, Cassio P. de Campos, Alessio Benavoli, and Alessandro Antonucci. Probabilistic inference in credal networks: new complexity results . *Journal of Artificial Intelligence Research*, 50:603–637, 2014.

[68] Enrique Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[69] Enrique Miranda. Updating coherent previsions on finite spaces. *Fuzzy Sets and Systems*, 160(9):1286–1307, 2009.

[70] Enrique Miranda and Gert de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 46(1):188–225, 2007.

[71] Enrique Miranda and Gert de Cooman. Lower previsions. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 28–55. John Wiley & Sons, Chichester, 2014.

[72] Enrique Miranda and Gert de Cooman. Structural judgements. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 56–78. John Wiley & Sons, Chichester, 2014.

[73] Enrique Miranda and Sébastien Destercke. Extreme points of the credal sets generated by comparative probabilities. *Journal of Mathematical Psychology*, 64–65:44–57, 2015.

[74] Enrique Miranda and Marco Zaffalon. Notes on desirability and conditional lower previsions. *Annals of Mathematics and Artificial Intelligence*, 60(3-4):251–309, 2011.

[75] Enrique Miranda and Marco Zaffalon. Independent products in infinite spaces. *Journal of Mathematical Analysis and Applications*, 425(1):460–488, 2014.

[76] Serafín Moral. Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics and Artificial Intelligence*, 45(1-2):197–214, 2005.

[77] Serafín Moral and Nic Wilson. Revision rules for convex sets of probabilities. In *Mathematical models for handling partial knowledge in artificial intelligence*, pages 113–127. Springer, Boston, 1995.

[78] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, San Mateo, 1988.

[79] Judea Pearl and Thomas Verma. The logic of representing dependencies by directed graphs. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 374–379. 1987.

[80] Renato Pelessoni and Paolo Vicig. Williams coherence and beyond. *International Journal of Approximate Reasoning*, 50(4):612–626, 2009.

[81] Alberto Piatti, Alessandro Antonucci, and Marco Zaffalon. Building knowledge-based systems by credal networks: a tutorial. In Albert R. Baswell, editor, *Advances in Mathematics Research, Volume 11*, pages 227–279. Nova Science Publishers, 2010.

[82] Olivier Pourret, Patrick Naïm, and Bruce Marcot. *Bayesian networks: a practical guide to applications*. John Wiley & Sons, Chichester, 2008.

[83] Erik Quaeghebeur. The CONEstrip algorithm. In *Synergies of Soft Computing and Statistics for Intelligent Data Analysis*, pages 45–54. Springer, Berlin, 2013.

[84] Erik Quaeghebeur. Desirability. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 1–27. John Wiley & Sons, Chichester, 2014.

[85] Erik Quaeghebeur, Gert de Cooman, and Filip Hermans. Accept & reject statement-based uncertainty models. *International Journal of Approximate Reasoning*, 57:69–102, 2015.

[86] Ralph Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[87] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[88] Glenn Shafer. Bayes's two arguments for the rule of conditioning. *The Annals of Statistics*, 10(4):1075–1089, 1982.

[89] Glenn Shafer. Conditional probability. *International Statistical Review*, 53(3):261–275, 1985.

[90] Glenn Shafer. Advances in the understanding and use of conditional independence. *Annals of Mathematics and Artificial Intelligence*, 21(1):1–11, 1997.

[91] Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33(1):1–49, 2003.

[92] Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76(1):107–133, 2011.

[93] Michael Smithson. Elicitiation. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 318–328. John Wiley & Sons, Chichester, 2014.

[94] Matthias C. M. Troffaes. Conditional lower previsions for unbounded random quantities. In *Soft Methods for Integrated Uncertainty Modelling*, pages 201–209. Springer, Berlin, 2006.

[95] Matthias C. M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29, 2007.

[96] Matthias C. M. Troffaes and Gert de Cooman. *Lower previsions*. John Wiley & Sons, 2014.

[97] Matthias C. M. Troffaes and Michael Goldstein. A note on the temporal sure preference principle and the updating of lower previsions. In Fabio G. Cozman, Thierry Denœux, Sébastien Destercke, and Teddy Seidenfeld, editors, *ISIPTA '13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theory and Applications*, pages 319–328. SIPTA, Compiègne, 2013.

[98] Matthias C. M. Troffaes and Robert Hable. Computation. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 329–337. John Wiley & Sons, Chichester, 2014.

[99] Arthur Van Camp, Gert de Cooman, Enrique Miranda, and Erik Quaeghebeur. Modelling Indifference with Choice Functions. 2015. Accepted for publication in ISIPTA '15: Proceedings of the Ninth International Symposium on Imprecise Probability: Theory and Applications.

[100] Barbara Vantaggi. Graphical models for conditional independence structures. In *ISIPTA '01: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 332–341. 2001.

[101] Barbara Vantaggi. The L-separation criterion for description of cs-independence models. *International Journal of Approximate Reasoning*, 29(3):291–316, 2002.

[102] Barbara Vantaggi. Conditional independence structures and graphical models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(5):545–571, 2003.

[103] Barbara Vantaggi. Graphical representation of asymmetric graphoid structures. In *ISIPTA '03: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 560–574. SIPTA, 2003.

[104] Paolo Vicig, Marco Zaffalon, and Fabio G. Cozman. Notes on "Notes on conditional previsions". *International Journal of Approximate Reasoning*, 44(3):358–365, 2007.

[105] Frans Voorbraak. Probabilistic belief expansion and conditioning. Technical report, Department of Mathematics, Computer Science, Physics and Astronomy, University of Amsterdam, 1996.

[106] Peter Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[107] Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58:3–57, 1996.

[108] Peter Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83(1):1–58, 1996.

[109] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2-3):125–148, 2000.

[110] Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10(3):741–761, 1982.

[111] Peter Walley, Renato Pelessoni, and Paolo Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126(1):119–151, 2004.

[112] Peter M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, 1975. Revised journal version: [113].

[113] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[114] Nic Wilson and Serafín Moral. A logical view of probability. In Anthony G. Cohn, editor, *Proceedings of the Eleventh European Conference on Artificial Intelligence*, pages 386–390. John Wiley and Sons, Amsterdam, 1994.

[115] Marco Zaffalon. The naive credal classifier. *Journal of Statistical Planning and Inference*, 105(1):5–21, 2002.

[116] Marco Zaffalon and Enrico Fagiuoli. Tree-based credal networks for classification. *Reliable computing*, 9(6):487–509, 2003.

[117] Marco Zaffalon and Enrique Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34(2):757, 2009.

[118] Marco Zaffalon and Enrique Miranda. Probability and time. *Artificial Intelligence*, 198:1–51, 2013.

[119] Marco Zaffalon, Keith Wesnes, and Orlando Petrini. Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, 29(1-2):61–79, 2003.