Trustworthiness and Expertise: Social Choice and Logic-based Perspectives

A thesis submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy

Joseph Singleton

XXX 2022

Abstract

This thesis studies problems involving unreliable information. We look at how to aggregate conflicting reports from multiple unreliable sources, how to assess the trustworthiness and expertise of sources, and investigate the extent to which the truth can be found with imperfect information. We take a formal approach, developing mathematical frameworks in which these problems can be formulated precisely and their properties studied. The results are of a conceptual and technical nature, which aim to elucidate interesting properties of the problem at the core abstract level.

In the first half we adopt the axiomatic approach of *social choice theory*. We formulate *truth discovery* – the problem of aggregating reports to estimate true information and reliability of the sources – as a social choice problem. We apply the axiomatic method to investigate desirable properties of such aggregation methods, and analyse a specific truth discovery method from the literature. We go on to study ranking methods for *bipartite tournaments*. This setting can be applied to rank sources according to their accuracy on a number of topics, and is also of independent interest.

In the second half we take a logic-based perspective. We use modal logic to formalise the notion of expertise, and explore connections with knowledge and truthfulness of information. We use this as the foundation for a belief change problem, in which reports must be aggregated to form beliefs about the true state of the world and the expertise of the sources. We again take an axiomatic approach – this time in the tradition of belief revision – where several postulates are proposed as rationality criteria. Finally, we address *truth-tracking*: the problem of finding the truth given non-expert reports. Adapting recent work combining logic with formal learning theory, we investigate the extent to which truth-tracking is possible, and how truth-tracking interacts with rationality.

Contents

Contents **i**

Ackowledgements **v**

List of Publications **vi**

I Introduction and Motivation **1**

1 Introduction **2**

2 Thesis Outline **4**

II Social Choice Perspectives **5**

3 Introduction **6**

4 Truth Discovery **7**

4.1 Introduction 7
4.2 An idealised framework for truth discovery 10
4.2.1 Truth discovery networks 10
4.2.2 Truth discovery operators 11
4.3 Examples of truth discovery operators 13
4.3.1 Voting 13
4.3.2 Sums 13
4.3.3 Recursive truth discovery operators 14
4.4 Axioms for truth discovery 15
4.4.1 Coherence 15
4.4.2 Symmetry 16
4.4.3 Fact ranking axioms 17

4.4.4 Independence axioms . . . 18
4.5 Satisfaction of the axioms . . . 21
4.5.1 Voting . . . 21
4.5.2 Iterative and recursive operators . . . 21
4.5.3 Sums . . . 25
4.5.4 Modifying *Voting* and *Sums* . . . 29
4.6 Variable domain truth discovery . . . 37
4.7 Discussion . . . 41
4.8 Related work . . . 43
4.9 Summary . . . 45

5 Bipartite Tournaments **47**
5.1 Introduction . . . 47
5.2 Preliminaries . . . 49
5.2.1 Bipartite Tournaments . . . 49
5.2.2 Chain Graphs . . . 50
5.3 Ranking via Chain Editing . . . 51
5.3.1 Chain-minimal Operators . . . 52
5.3.2 A Maximum Likelihood Interpretation . . . 53
5.4 Axiomatic analysis . . . 58
5.4.1 The Axioms . . . 58
5.4.2 Axiom Compatibility with **chain-min** . . . 60
5.5 Match-preference operators . . . 69
5.6 Relaxing chain-min . . . 73
5.6.1 Chain-definability . . . 73
5.6.2 Interleaving Operators . . . 76
5.6.3 Axiom Compatibility . . . 81
5.6.4 Axiomatic Characterisation of φ_{CI} . . . 86
5.7 Related Work . . . 93
5.8 Conclusion . . . 94

III Logic-based Perspectives **96**

6 Introduction **97**

7 Expertise and Knowledge **98**
7.1 Introduction . . . 98
7.2 Expertise and Soundness . . . 100
7.3 Closure Properties . . . 102
7.4 Connection with Epistemic Logic . . . 106

7.5 Axiomatisation . . . 111
7.6 The Multi-source Case . . . 123
7.6.1 Collective Knowledge . . . 123
7.6.2 Collective Expertise . . . 124
7.7 Dynamic Extension . . . 129
7.7.1 Expertise Increase . . . 129
7.7.2 Sound Announcements . . . 131
7.8 Conclusion . . . 135

8 Belief Change with Non-Expert Sources **136**
8.1 Introduction . . . 136
8.2 The Framework . . . 138
8.3 The Problem . . . 144
8.3.1 Basic Postulates . . . 144
8.3.2 Model-Based Operators . . . 148
8.4 Constructions . . . 149
8.4.1 Conditioning Operators . . . 150
8.4.2 Score-Based Operators . . . 158
8.5 One-Step Revision . . . 160
8.6 Selective Change . . . 168
8.6.1 Case Independence . . . 173
8.6.2 Expertise and Selectivity . . . 175
8.7 Related Work . . . 176
8.8 Conclusion . . . 178

9 Truth-Tracking **179**
9.1 Introduction . . . 179

IV Conclusions **180**

10 Conclusion **181**
10.1 Summary . . . 181
10.2 Future Work . . . 181

Bibliography **182**

A Proofs for Chapter 4 **193**
A.1 Proof of Theorem 4.4.1 . . . 193
A.2 Proof of Theorem 4.4.3 . . . 195
A.3 Proof of Theorem 4.5.1 . . . 197
A.4 Proof of Lemma 4.5.2 . . . 198

A.5 Proof of Theorem 4.5.2 199
A.6 Proof of Theorem 4.5.3 202
A.7 Proof of Theorem 4.5.5 203

Ackowledgements

I would like to thank Bear for being a dog. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

List of Publications

The content of this thesis is derived from the following publications. [**TODO:** Add descriptions and chapter referencesbeneath each citation?]

- Joseph Singleton and Richard Booth. "An Axiomatic Approach to Truth Discovery". In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS 20. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 2011–2013. ISBN: 9781450375184

- Joseph Singleton and Richard Booth. "Rankings for Bipartite Tournaments via Chain Editing". In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '21. Virtual Event, United Kingdom: International Foundation for Autonomous Agents and Multiagent Systems, 2021, pp. 1236–1244. ISBN: 9781450383073

- Joseph Singleton. "A Logic of Expertise". In: *ESSLLI 2021 Student Session* (2021). URL: https://arxiv.org/abs/2107.10832

- Joseph Singleton and Richard Booth. *Who's the Expert? On Multi-source Belief Change*. 2022. DOI: 10.48550/ARXIV.2205.00077. URL: https://arxiv.org/abs/2205.00077

Part I

Introduction and Motivation

1 Introduction

- Overall theme: how should we deal with unreliable information?
- We want to:
  - Aggregate conflicting reports (crowdsourcing, news)
  - Assess the trustworthiness of information sources
  - Understand what reliability, trustworthiness and expertise *mean*
  - Find the truth with imperfect information
- This thesis offers two main perspectives on these general themes
  - **Social choice theory.**
    * By posing the aggregation problem as one of social choice, we can apply the axiomatic method to investigate desirable properties of aggregation methods. We can then analyse and evaluate such methods in a formal and principled way.
    * Related ranking problems can be addressed through the lens of social choice.
  - **Logic and knowledge representation.**
    * We develop a logical system to formalise notions of expertise, and explore connections with knowledge and information.
    * We use these formal notions to express the aggregation problem in logical terms, taking an alternative look at the problems of the first part of the thesis. We use what is essentially still an axiomatic approach, but now in the tradition of knowledge representation and rational belief change.
    * This logical model is well-suited to investigate *truth-tracking*: the question of when we can find the truth given that not all sources are experts.

- Note that while there are many links between the two major parts, they are not tightly connected and may be read independently.

2 Thesis Outline

Part II

Social Choice Perspectives

3 Introduction

- Describe what we mean by social choice?
- Overview of how our stuff will relate to the COMSOC literature?

4 Truth Discovery

4.1 Introduction

There is an increasing amount of data available in today's world, particularly from the web, social media platforms and crowdsourcing systems. The openness of such platforms makes it simple for a wide range of users to share information quickly and easily, potentially reaching a wide international audience. It is inevitable that amongst this abundance of data there are *conflicts*, where data sources disagree on the truth regarding a particular object or entity. For example, low-quality sources may mistakenly provide erroneous data for topics on which they lack expertise.

Resolving such conflicts and determining the true facts is therefore an important task. Truth discovery has emerged as a set of techniques to achieve this by considering the *trustworthiness* of sources [69, 51, 11]. The general principle is that true facts are those claimed by trustworthy sources, and trustworthy sources are those that claim believable facts. Application areas include real-time traffic navigation [34], drug side-effect discovery [74], crowdsourcing and social sensing [110, 98, 73].

For a simple example of a situation where trust can play an important role in conflict resolution, consider the following example.

Example 4.1.1. *Let o and p represent two images for which crowdsourcing workers are asked to provide labels (in the truth discovery terminology, o and p are called* objects*).*

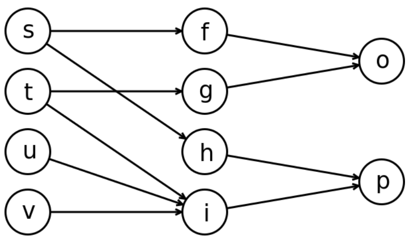


Figure 4.1: Illustrative example of a dataset to which truth discovery can be applied with data sources $\{s, t, u, v\}$, facts $\{f, g, h, i\}$ and objects $\{o, p\}$

Consider workers (the data sources) s, t, u and v who put forward potential labels f, g for o, and h, i for p, as shown in Fig. 4.1; such potential answers are termed facts. *In the graphical representation, sources, facts and objects are shown from left to right, and the edges indicate claims made by sources and the objects to which facts relate.*

Without considering trust information, the label for o appears a tie, with both options f and g receiving one vote from sources s and t respectively.

Taking a trust-aware approach, however, we can look beyond object o to consider the trustworthiness *of s and t. Indeed, when it comes to object p, t agrees with two extra sources u and v, whereas s disagrees with everyone. In principle there could be* hundreds *of extra sources here instead of just two, in which case the effect would be even more striking. We may conclude that s is* less trustworthy *than t. Returning to o, we see that g is supported by a more trustworthy source, and conclude that it should be accepted over f.*

Many truth discovery algorithms have been proposed in the literature with a wide range of techniques used, e.g. iterative heuristic-based methods [83, 42], probabilistic models [108], maximum likelihood estimation and optimisation-based methods [70], and neural network models [65, 75, 99]. It is common for such algorithms to be evaluated empirically by running them against real-world or synthetic datasets for which the true facts are already known; this allows *accuracy* and other metrics to be calculated, and permits comparison between algorithms (see [97] for a systematic empirical evaluation of this kind). This may be accompanied by some theoretical analysis, such as calculating run-time complexity [51], proving convergence of an iterative algorithm [109], or proving convergence to the 'true' facts under certain assumptions on the distribution of source trustworthiness [103, 102, 45].

A limitation of this kind of analysis is that the results only apply narrowly to particular algorithms, due to the assumptions made (for instance, that claims from sources follow a particular probability distribution). Such assumptions can be problematic in domains where the desired truth is somewhat 'fuzzy'; for example, image classification problems and determining the copyright status of books.[1]

In this work we take first steps towards a more general approach, in which we aim to study truth discovery without reference to any specific methodology or probabilistic framework. To do so we note the similarities between truth discovery and problems such as judgment aggregation [36], voting theory [113] ranking and recommendation systems [2, 3, 5, 93] in which the *axiomatic approach* of social choice has been successfully applied. In taking the axiomatic approach one aims to formulate *axioms* that encode intuitively desirable properties that an algorithm may possess. The interaction between these axioms can then be studied; typical results include *impossibility results*, where it is shown that a set of axioms cannot hold

[1] https://www.nytimes.com/2019/08/19/technology/amazon-orwell-1984.html

simultaneously, and *characterisation results*, where it is shown that a set of axioms are uniquely satisfied by a particular algorithm.

Such analysis brings a new *normative* perspective to the truth discovery literature. This complements empirical evaluation: in addition to seeing how well an algorithm performs in practise on test datasets, one can check how well it does against theoretical properties that any 'reasonable' algorithm should satisfy. The satisfaction (or failure) of such properties then shines new light on the *intuitive behaviour* of an algorithm, and may guide development of new ones.

With this in mind, we develop a simplified framework for truth discovery in which axioms can be formulated, and go on to give both an impossibility result and an axiomatic characterisation of a baseline voting algorithm. We also analyse the class of *recursive* truth discovery algorithms, which includes many existing examples from the literature. In particular, we analyse the well-known algorithm *Sums* [83] with respect to the axioms.

However, as a first step towards a social choice perspective of truth discovery, our framework involves a number of simplifying assumptions not commonly made in the truth discovery literature.

- **Manipulation and collusion.** Some of our axioms assume sources are not *manipulative*: they provide claims in good faith, and do not aim to misinform or artificially improve their standing with respect to the truth discovery algorithm. We also assume sources act independently, i.e. they do not *collude* with or *copy* one another.

- **Object correlations.** We do not model correlations between the objects of interest in the truth discovery problem. For example, in a crowdsourcing setting it may be known in advance that two objects o and p are similar, so that the true labels for o and p are correlated; this cannot be expressed in our framework.

- **Ordinal outputs.** For the most part, the outputs of our truth discovery methods consist of *rankings* of the sources and facts. Thus, we describe when a source is considered *more trustworthy* than another, but do not assign precise numerical values representing trustworthiness. This breaks with tradition in the truth discovery literature, but is a common point of view in social choice theory.

At first glance these are strong assumptions, and rule out potential applications of our version of truth discovery. However, we argue that the problem is non-trivial even in this simplified setting, and that interesting axioms can still be put forth. The framework as set out here lays the groundwork for these assumptions to be lifted in future work.

The chapter is organised as follows. Our framework is introduced and formally defined in the next section. Section 4.3 provides examples of truth discovery algorithms from the literature expressed in the framework. In Section 4.4 we formally introduce the axioms and present an impossibility result showing a subset of these cannot all be satisfied simultaneously. The examples of Section 4.3 are then revisited in Section 4.5, where we analyse them with respect to the axioms and propose modifications to resolve some axiom failures. In Section 4.6 we extend the framework to allow variable domains of sources, facts and objects, and give an impossibility result similar to that of Section 4.4. We discuss the axioms in Section 4.7 and related work in Section 4.8. We conclude in Section 4.9. Missing proofs are given in Appendix A.

4.2 An idealised framework for truth discovery

In this section we define our formal framework, which provides the key definitions required for theoretical discussion and analysis of truth discovery methods.

For most of the chapter, we consider a fixed domain of finite and mutually disjoint sets $\mathcal{S}$, $\mathcal{F}$ and $\mathcal{O}$ throughout, called the *sources*, *facts* and *objects* respectively. All definitions and axioms will be stated with respect to these sets.[2]

4.2.1 Truth discovery networks

A core definition of the framework is that of a *truth discovery network*, which represents the input to a truth discovery problem. We model this as a tripartite graph with certain constraints on its structure, in keeping with approaches taken throughout the truth discovery literature [108, 51].

Definition 4.2.1. *A* truth discovery network *(hereafter a* TD network*) is a directed graph $N = (V, E)$ where $V = \mathcal{S} \cup \mathcal{F} \cup \mathcal{O}$, and $E \subseteq (\mathcal{S} \times \mathcal{F}) \cup (\mathcal{F} \times \mathcal{O})$ has the following properties:*

1. *For each $f \in \mathcal{F}$ there is a unique $o \in \mathcal{O}$ with $(f, o) \in E$, denoted $\mathsf{obj}_N(f)$. That is, each fact is associated with exactly one object.*

2. *For $s \in \mathcal{S}$ and $o \in \mathcal{O}$, there is at most one directed path from s to o. That is, sources cannot claim multiple facts for a single object.*

3. *$(\mathcal{S} \times \mathcal{F}) \cap E$ is non-empty. That is, at least one claim is made.*

We will say that s claims *f when $(s, f) \in E$. Let $\mathcal{N}$ denote the set of all TD networks.*

[2] We generalise to variable domains in Section 4.6.

Figure 4.1 (page 7) provides an example of a TD network. Note that there is no requirement that a source makes a claim for *every* object, or even that a source makes any claims at all. This reflects the fact that truth discovery datasets are in practise extremely sparse, i.e. each individual source makes few claims. Conversely, we allow for facts that receive no claims from any sources.

Also note that the object associated with a fact f is not fixed across all networks. This is because we view facts as *labels* for information that sources may claim, not the pieces of information themselves. Similarly, we consider objects simply as labels for real-world entities. Whilst a particular piece of information has a fixed entity to which it pertains, the labels do not.[3]

A special case of our framework is the binary case in which every object has exactly two associated facts. This brings us close to the setting studied in *judgment aggregation* [36] and, specifically (since sources do not necessarily claim a fact associated to every object) to the setting of *binary aggregation with abstentions* [20, 30]. An important difference, however, is that for simplicity we do not assume any *constraints* on the possible configurations of true facts across *different* objects. That is, any combination of facts is feasible. In judgment aggregation such an assumption has the effect of neutralising the impossibility results that arise in that domain (see, e.g., [20]). We shall see that that is not the case in our setting.

To simplify the notation in what follows, for a network $N = (V, E)$ we write $\mathsf{facts}_N(s) = \{f \in \mathcal{F} : (s, f) \in E\}$ for the set of facts claimed by a source s, and $\mathsf{src}_N(f) = \{s \in \mathcal{S} : (s, f) \in E\}$ for the set of sources claiming a fact f.

4.2.2 Truth discovery operators

Having defined the input to a truth discovery problem, the output must be defined. Contrary to many approaches in the truth discovery literature which output numeric *trust scores* for sources and *belief scores* for facts [108, 83, 42, 112, 110, 111], we consider the primary output to be *rankings* of the sources and facts. To the extent that we do consider numeric scores, it is only to induce a ranking. This is because we are chiefly interested in *ordinal properties* rather than quantitative values. Indeed, for the theoretical analysis we wish to perform it is only important that a source is *more trustworthy* than another; the particular numeric scores produced by an algorithm are irrelevant.

Moreover, the scores produced by existing algorithms may have no semantic meaning [83], and so referring to numeric values is not meaningful when comparing across algorithms. In this case it is only the rankings of sources and facts that can

[3] For example, when implementing truth discovery algorithms in practise it is common to assign integer IDs to the 'facts' and 'objects'; the algorithm then operates using only the integer IDs. In this case there is no reason to require that fact 17 is always associated with object 4, for example, and the same principle applies in our framework.

be compared, which is further motivation for our choice. This point of view is also common across the social choice literature.

However, numerical scores do provide valuable information for comparing sources and facts given a *fixed* algorithm. For example, the magnitude of the difference in trust scores for sources s and t tells us something about *confidence*: a small difference indicates low confidence in distinguishing s and t – even if one is ranked above the other – whereas a large difference indicates high confidence. In this sense our decision to primarily deal with ordinal outputs (and ordinal axioms) is another simplifying assumption compared to typical truth discovery settings.

For a set X, let $\mathcal{L}(X)$ denote the set of all total preorders on X, i.e. the set of transitive, reflexive and complete binary relations on X. We define a *truth discovery operator* as a function which maps networks to rankings of sources and facts.

Definition 4.2.2. *An* ordinal truth discovery operator T *(hereafter* TD operator*) is a mapping $T : \mathcal{N} \to \mathcal{L}(\mathcal{S}) \times \mathcal{L}(\mathcal{F})$. We shall write $T(N) = (\sqsubseteq_N^T, \preceq_N^T)$, i.e. $\sqsubseteq_N^T$ is a total preorder on $\mathcal{S}$ and $\preceq_N^T$ is a total preorder on $\mathcal{F}$.*

Intuitively, the relation $\sqsubseteq_N^T$ is a measure of *source trustworthiness* in the network N according to T, and $\preceq_N^T$ is a measure of *fact believability*; $s_1 \sqsubseteq_N^T s_2$ means that source s_2 is *at least as trustworthy* as source s_1, and $f_1 \preceq_N^T f_2$ means fact f_2 is *at least as believable* as fact f_1. The notation $\sqsubset_N^T$ and $\simeq_N^T$ will be used to denote the strict and symmetric orders induced by $\sqsubseteq_N^T$ respectively. For fact rankings, $\prec_N^T$ and $\approx_N^T$ are defined similarly. Note that for simplicity the fact ranking $\preceq_N^T$ compares *all* facts, even those associated with different objects in N.

To capture existing truth discovery methods we introduce *numerical operators*, which assign each source a numeric *trust score* and each fact a *belief score*.

Definition 4.2.3. *A* numerical TD operator *is a mapping $T : \mathcal{N} \to \mathbb{R}^{\mathcal{S} \cup \mathcal{F}}$, i.e. T assigns to each TD network N a function $T(N) = T_N : \mathcal{S} \cup \mathcal{F} \to \mathbb{R}$. For $s \in \mathcal{S}$, $T_N(s)$ is the* trust score *for s in the network N according to T; for $f \in \mathcal{F}$, $T_N(f)$ is the* belief score *for f. The set of all numerical TD operators will be denoted by $\mathcal{T}_{Num}$.*

Note that any numerical operator T naturally induces an ordinal operator $\hat{T}$, where $s_1 \sqsubseteq_N^{\hat{T}} s_2$ iff $T_N(s_1) \leq T_N(s_2)$, and $f_1 \preceq_N^{\hat{T}} f_2$ iff $T_N(f_1) \leq T_N(f_2)$. Henceforth we shall write $\sqsubseteq_N^T$, $\preceq_N^T$ without explicitly defining the induced ordinal operator $\hat{T}$.

It is worth noting that yet other truth discovery algorithms output neither rankings nor numeric scores for facts, but only a single 'true' fact for each object [70, 28, 105]. This is also the approach taken in judgment aggregation, where an *aggregation rule* selects which formulas are to be taken as true. In the case of finitely many possible facts, such algorithms can be modelled in our framework as numerical operators where $T_N(f) = 1$ for each identified 'true' fact f, and $T_N(g) = 0$ for other facts g. To go in the reverse direction and obtain the 'true' facts according to an operator, one may simply select the set of facts for each object that rank maximally.

4.3 Examples of truth discovery operators

Our framework can capture some operators that have been proposed in the truth discovery literature. In this section we provide two concrete examples: *Voting*, which is a simple approach commonly used as a baseline method, and *Sums* [83]. We go on to outline the class of *recursive operators* – of which *Sums* is an instance – which contains many more examples from the literature.

4.3.1 Voting

In *Voting*, we consider each source to cast 'votes' for the facts they claim, and facts are ranked according to the number of votes received. Clearly this method disregards the source trustworthiness aspect of truth discovery, as a vote from one source carries as much weight as a vote from any other. As such, *Voting* cannot be considered a serious contender for truth discovery. It is nonetheless useful as a simple baseline method against which to compare more sophisticated methods.

Definition 4.3.1. Voting *is the numerical operator defined as follows: for any network $N \in \mathcal{N}$, $s \in \mathcal{S}$ and $f \in \mathcal{F}$, $T_N(s) = 1$ and $T_N(f) = |\mathsf{src}_N(f)|$.*

Consider the network N shown in Fig. 4.1. Facts f, g and h each receive one vote, whereas i receives 3. The fact ranking induced by *Voting* is therefore $f \approx g \approx h \prec i$. On the other hand, all sources receive a trust score of 1 and therefore rank equally.

4.3.2 Sums

Sums [83] is a simple and well-known operator adapted from the *Hubs and Authorities* [61] algorithm for ranking web pages. The algorithm operates iteratively and recursively, assigning each source and fact a sequences of scores, with the final scores taken as the limit of the sequence.

Initially, scores are fixed at a constant value of $1/2$. The trust score for each source is then updated by summing the belief score of its associated facts. Similarly, belief scores are updated by summing the trust scores of the associated sources. To prevent these scores from growing without bound as the algorithm iterates, they are normalised at each iteration by dividing each trust score by the maximum across all sources (belief scores are normalised similarly).

Expressed in our framework, we have that if T is the (numerical) operator giving the scores at iteration n, then the pre-normalisation scores at iteration $n+1$ are given by T', where

$$T'_N(s) = \sum_{f \in \mathsf{facts}_N(s)} T_N(f); \quad T'_N(f) = \sum_{s \in \mathsf{src}_N(f)} T'_N(s) \tag{4.1}$$

Consider again the network N shown in Fig. 4.1. It can be shown that, with T denoting the limiting scores from *Sums* with normalisation, we have $T_N(s) = 0$, $T_N(t) = 1$, and $T_N(u) = T_N(v) = \sqrt{2}/2$. The induced ranking of sources is therefore $s \sqsubset u \simeq v \sqsubset t$.

For fact scores, we have $T_N(f) = 0$, $T_N(g) = \sqrt{2} - 1$, $T_N(h) = 0$ and $T_N(i) = 1$, so the ranking is $f \approx h \prec g \prec i$. Note that fact g fares better under *Sums* than *Voting*, due to its association with the highly-trusted source t.

4.3.3 Recursive truth discovery operators

The iterative and recursive aspect of *Sums* is hoped to result in the desired mutual dependence between trust and belief scores: namely that sources claiming high-belief facts are seen as trustworthy, and vice versa. In fact, this recursive approach is near universal across the truth discovery literature (see for instance [104, 34, 111, 70, 42, 112]). As such it is appropriate to identify the class of *recursive operators* as an important subset of $\mathcal{T}_{Num}$. To make a formal definition we first define an *iterative operator*.

Definition 4.3.2. *An* iterative operator *is a sequence $(T^n)_{n\in\mathbb{N}}$ of numerical operators. An iterative operator is said to* converge *to a numerical operator T^* if $\lim_{n\to\infty} T^n_N(z) = T^*_N(z)$ for all networks N and $z \in \mathcal{S} \cup \mathcal{F}$. In such case the iterative operator can be identified with the ordinal operator induced by its limit T^*.*

Note that it is possible that an iterative operator $(T^n)_{n\in\mathbb{N}}$ converges for only a subset of networks. In such case we can consider $(T^n)_{n\in\mathbb{N}}$ to converge to a 'partial operator' and identify it with the induced partial ordinal operator; that is, a partial function $\mathcal{N} \to \mathcal{L}(\mathcal{S}) \times \mathcal{L}(\mathcal{F})$. Recursive operators can now be defined as those iterative operators where T^{n+1} can be obtained from T^n.

Definition 4.3.3. *An iterative operator $(T^n)_{n\in\mathbb{N}}$ is said to be* recursive *if there is a function $U : \mathcal{T}_{Num} \to \mathcal{T}_{Num}$ such that $T^{n+1} = U(T^n)$ for all $n \in \mathbb{N}$.*

In this context the mapping $U : \mathcal{T}_{Num} \to \mathcal{T}_{Num}$ is called the *update function*, and the initial operator T^1 is called the *prior operator*. For a prior operator T and update function U, we write $\mathsf{rec}(T, U)$ for the associated recursive operator; that is, $T^1 = T$ and $T^{n+1} = U(T^n)$.

Returning to *Sums*, we see that Eq. (4.1) defines a mapping $\mathcal{T}_{Num} \to \mathcal{T}_{Num}$ and consequently an update function U^{Sums}. The normalisation step can be considered a separate update function norm which maps any numerical operator T to T', where[4]

$$T'_N(s) = \frac{T_N(s)}{\max\limits_{x \in \mathcal{S}} |T_N(x)|}, \quad T'_N(f) = \frac{T_N(f)}{\max\limits_{y \in \mathcal{F}} |T_N(y)|}$$

It can then be seen that *Sums* is the recursive operator $\mathsf{rec}(T^{\text{fixed}}, \mathsf{norm} \circ U^{\mathsf{Sums}})$, where $T_N^{\text{fixed}} \equiv 1/2$.

Many other existing algorithms proposed in the literature can also be realised as recursive operators in the framework, such as *Investment*, *PooledInvestment* [83], *TruthFinder* [108], LDT [111] and others.

4.4 Axioms for truth discovery

Having laid out the formal framework, we now introduce axioms for truth discovery. To start with, we consider axioms which encode a desirable theoretical property that we believe any 'reasonable' operator T should satisfy. Several properties of this nature can be obtained by adapting existing axioms from the social choice literature (e.g. from voting [17], ranking systems [93, 2] and judgement aggregation [36]), to our framework.

However, the correspondence between truth discovery and classical social choice problems – such as voting – has its limits. To show this, we translate the famous Independence of Irrelevant Alternatives (IIA) axiom [6] to our setting, and argue that it is actually an *undesirable* property. Indeed, it will be seen that this translated axiom, in combination with two basic desirable axioms, leads to *Voting*-like behaviour in every network, which is undesirable for the reasons given in Section 4.3.1. Furthermore, a slight strengthening of the IIA axiom completely characterises the fact ranking component of *Voting*. These results formalise the intuition that truth discovery's consideration of source-trustworthiness leads to fundamental differences from classical social choice problems.

Afterwards, we will revisit the specific operators of the previous section to check which axioms are satisfied.

4.4.1 Coherence

As mentioned previously, a guiding principle of truth discovery is that sources claiming highly believed facts should be seen as trustworthy, and that facts backed by highly trusted sources should be seen as believable.

Whilst this intuition is difficult to formalise in general, it is possible to do so in particular cases where there are obvious means by which to compare the set of

[4] If $\max_{x \in \mathcal{S}} |T_N(x)| = 0$ then the above is ill-defined; we set $T'_N(s) = 0$ for all s in this case. Fact belief scores are defined similarly if the maximum is 0.

facts for two sources (and vice versa). This situation is considered in the axiomatic analysis of ranking and reputation systems under the name *Transitivity* [93, 2], and we adapt it to truth discovery in this section. A preliminary definition is required.

Definition 4.4.1. *Let T be a TD operator, N be a TD network and $Y, Y' \subseteq \mathcal{F}$. We say Y is* less believable *than Y' with respect to N and T if there is a bijection $\varphi : Y \to Y'$ such that $f \preceq_N^T \varphi(f)$ for each $f \in Y$, and $\hat{f} \prec_N^T \varphi(\hat{f})$ for some $\hat{f} \in Y$.*

For $X, X' \subseteq \mathcal{S}$ we define X less trustworthy *than X' with respect to N and T in a similar way.*

In plain English, Y less believable than Y' means that the facts in each set can be paired up in such a way that each fact in Y' is at least as believable as its counterpart in Y, and at least one fact in Y' is strictly more believable. Now, consider a situation where $\mathsf{facts}_N(s_1)$ is less believable than $\mathsf{facts}_N(s_2)$. In this case the intuition outlined above tells us that s_2 provides 'better' facts, and should thus be seen as more trustworthy than s_1. A similar idea holds if $\mathsf{src}_N(f_1)$ is less trustworthy than $\mathsf{src}_N(f_2)$ for some facts f_1, f_2. We state this formally as our first axiom.

Axiom 4.4.1 (Coherence)**.** *For any network N, $\mathsf{facts}_N(s_1)$ less believable than $\mathsf{facts}_N(s_2)$ implies $s_1 \sqsubset_N^T s_2$, and $\mathsf{src}_N(f_1)$ less trustworthy than $\mathsf{src}_N(f_2)$ implies $f_1 \prec_N^T f_2$.*

Coherence can be broken down into two sub-axioms: *Source-Coherence*, where the first implication regarding source rankings is satisfied; and *Fact-Coherence*, where the second implication is satisfied. We take Coherence to be a fundamental desirable axiom for TD operators.

4.4.2 Symmetry

Our next axiom requires that rankings of sources and facts should not depend on their 'names', but only on the structure of the network. To state it formally, we need a notion of when two networks are essentially the same but use different names.

Definition 4.4.2. *Two TD networks N and N' are* equivalent *if there is a graph isomorphism π between them that preserves sources, facts and objects, i.e., $\pi(s) \in \mathcal{S}$, $\pi(f) \in \mathcal{F}$ and $\pi(o) \in \mathcal{O}$ for all $s \in \mathcal{S}$, $f \in \mathcal{F}$ and $o \in \mathcal{O}$. In such case we write $\pi(N)$ for N'.*

Axiom 4.4.2 (Symmetry)**.** *Let N and $N' = \pi(N)$ be equivalent networks. Then for all $s_1, s_2 \in \mathcal{S}$, $f_1, f_2 \in \mathcal{F}$, we have $s_1 \sqsubseteq_N^T s_2$ iff $\pi(s_1) \sqsubseteq_{N'}^T \pi(s_2)$ and $f_1 \preceq_N^T f_2$ iff $\pi(f_1) \preceq_{N'}^T \pi(f_2)$.*

In the theory of voting in social choice, Symmetry as above is expressed as two axioms: *Anonymity*, where output is insensitive to the names of voters, and

Neutrality, where output is insensitive to the names of alternatives [113]. Analogous axioms are also used in judgment aggregation.

Symmetry can also be broken down into sub-axioms where the above need only hold for a subset of permutations π satisfying some condition: *Source-Symmetry* (where π must leave facts and objects fixed) and *Fact-Symmetry* (where π leaves sources and objects fixed). For truth discovery we have the additional notion of objects, and thus *Object-Symmetry* can defined be similarly.

4.4.3 Fact ranking axioms

Next, we introduce axioms that dictate the ranking of particular facts in cases where there is an 'obvious' ordering. *Unanimity* and *Groundedness* express the idea that if all sources are in agreement about the status of a fact, then an operator should respect this in its verdict. Two obvious ways in which sources can be in agreement are when *all* sources believe a fact is true, and when *none* believe a fact is true.

Axiom 4.4.3 (Unanimity)**.** *Suppose $N \in \mathcal{N}$, $f \in \mathcal{F}$, and $\mathsf{src}_N(f) = \mathcal{S}$. Then for any other $g \in \mathcal{F}$, $g \preceq_N^T f$.*

Axiom 4.4.4 (Groundedness)**.** *Suppose $N \in \mathcal{N}$, $f \in \mathcal{F}$, and $\mathsf{src}_N(f) = \emptyset$. Then for any other $g \in \mathcal{F}$, $f \preceq_N^T g$.*

That is, f cannot do better than to be claimed by all sources when T satisfies Unanimity, and cannot do worse than to be claimed by none when T satisfies Groundedness. Unanimity here is a truth discovery rendition of the same axiom in judgment aggregation, and can also be compared to the *weak Paretian* property in voting [17]. Groundedness is a version of the same axiom studied in the analysis of collective annotation [67].

The next axiom is a monotonicity property, which states that if f receives extra support from a new source s, then its ranking should receive a strictly positive boost.[5] Note that we do not make any judgement on the new ranking of s.

Axiom 4.4.5 (Monotonicity)**.** *Suppose $N \in \mathcal{N}$, $s \in \mathcal{S}$, $f \in \mathcal{F} \setminus \mathsf{facts}_N(s)$. Write E for the set of edges in N, and let N' be the network in which s claims f; i.e. the network with edge set*

$$E' = \{(s,f)\} \cup E \setminus \{(s,g) : g \neq f, \mathsf{obj}_N(g) = \mathsf{obj}_N(f)\}$$

Then for all $g \neq f$, $g \preceq_N^T f$ implies $g \prec_{N'}^T f$.

Note that the axioms in this section assume sources do not have 'negative' trust levels. That is, we assume that support from even the most untrustworthy source

[5] One could also consider the weak version, in which we only require $g \preceq_{N'}^T f$ in the consequent; we discuss this in Section 4.7.

still has a *positive* effect on the believability of a fact. Consequently, these axioms are not suitable in the presence of knowledgable but malicious sources who always claim false facts. Indeed, otherwise a fact claimed only by a 'negative' source should rank strictly *worse* than a fact with no sources, but this goes against Groundedness. Similarly, receiving extra support from a negative source should worsen a fact's ranking, contrary to Monotonicity. Moreover, Monotonicity implicitly assumes sources act independently, i.e. they do not *collude* with one another.[6]

While these assumptions may appear somewhat strong, we argue that this 'simple' case – with no 'negative' sources or collusion – is already non-trivial and permits interesting axiomatic analysis. We therefore view Unanimity, Groundedness and Monotonicity as desirable properties for TD operators.

4.4.4 Independence axioms

We now come to exploring the differences between truth discovery and other social choice problems via *independence* axioms. In voting, this takes the form of Independence of Irrelevant Alternatives (IIA), which requires that the ranking of two alternatives A and B depends only on the individual assessments of A and B, not on some 'irrelevant' alternative C.

An analogous truth discovery axiom states that the ranking of facts f_1 and f_2 for some object o depends only on the claims relating to o. Intuitively, this is *not* a desirable property. Indeed, we have already seen in Example 4.1.1 that the claims for object p in the network from Fig. 4.1 can play an important role in determining the ranking of f and g for object o, but the adapted IIA axiom precludes this.

This undesirability can be made precise. First, we must state the axiom formally.

Axiom 4.4.6 (Per-object Independence (POI)). *Let $o \in \mathcal{O}$. Suppose N_1, N_2 are networks such that $F_o = \mathsf{obj}_{N_1}^{-1}(o) = \mathsf{obj}_{N_2}^{-1}(o)$ and $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$ for each $f \in F_o$. Then the restrictions of $\preceq_{N_1}^T$ and $\preceq_{N_2}^T$ to F_o are equal; that is, $f_1 \preceq_{N_1}^T f_2$ iff $f_1 \preceq_{N_2}^T f_2$ for all $f_1, f_2 \in F_o$.*

Considering Fig. 4.1 again, POI implies that the ranking of f and g remains the same if the claims for h and i are removed. But in this case, Symmetry implies $f \approx g$. Similarly, the ranking of h and i remains the same if the claims for f and g are removed. In this case, Symmetry together with Monotonicity implies $h \prec i$, since $|\mathsf{src}_N(h)| < |\mathsf{src}_N(i)|$.

This observation forms the basis of the following result, which formalises the undesirability of POI: in the presence of our less controversial requirements of Symmetry and Monotonicity, it forces *Voting*-like behaviour within $\mathsf{obj}_N^{-1}(o)$ for

[6] Note that collusion has been studied in the truth discovery literature (e.g. [31, 8, 32]).

each $o \in \mathcal{O}$. We note that, for the special case of binary networks, similar results have been shown in the literature on binary aggregation with abstentions [20].

Theorem 4.4.1. *Let T be any operator satisfying Symmetry, Monotonicity and POI. Then for any $N \in \mathcal{N}$, $o \in \mathcal{O}$ and $f_1, f_2 \in \mathsf{obj}_N^{-1}(o)$ we have $f_1 \preceq_N^T f_2$ iff $|\mathsf{src}_N(f_1)| \leq |\mathsf{src}_N(f_2)|$.*

Proof (sketch). We will sketch the main ideas of the proof here with some technical details omitted; see Appendix A for the full proof. Let N be a network, o be an object and $f_1, f_2 \in \mathsf{obj}_N^{-1}(o)$. Consider N' obtained by removing from N all claims for objects other than o. By POI, we have $f_1 \preceq_N^T f_2$ iff $f_1 \preceq_{N'}^T f_2$. Since $|\mathsf{src}_N(f_j)| = |\mathsf{src}_{N'}(f_j)|$ also ($j \in \{1, 2\}$), it is sufficient for the proof to show that $f_1 \preceq_{N'}^T f_2$ iff $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$.

For the 'if' direction, first suppose $|\mathsf{src}_{N'}(f_1)| = |\mathsf{src}_{N'}(f_2)|$. Let π be the permutation which swaps f_1 with f_2 and swaps each source in $\mathsf{src}_{N'}(f_1)$ with one in $\mathsf{src}_{N'}(f_2)$; then we have $\pi(N') = N'$, and Symmetry of T gives $f_1 \approx_{N'}^T f_2$. In particular $f_1 \preceq_{N'}^T f_2$ as required.

Otherwise, $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Consider N'' where k sources from $\mathsf{src}_{N'}(f_2)$ are removed, and all other claims remain. By Symmetry as above, $f_1 \approx_{N''}^T f_2$. Applying Monotonicity k times we can produce N' from N'' and get $f_1 \prec_{N'}^T f_2$ as desired.

For the 'only if' statement, suppose $f_1 \preceq_{N'}^T f_2$ but, for contradiction, $|\mathsf{src}_{N'}(f_1)| > |\mathsf{src}_{N'}(f_2)|$. Applying Monotonicity again as above we get $f_1 \succ_{N'}^T f_2$ and the required contradiction. □

Recall that Coherence formalises the idea that source-trustworthiness should inform the fact ranking, and vice versa. Clearly *Voting* does not conform to this idea, and in fact even the object-wise voting patterns in Theorem 4.4.1 are incompatible with Coherence. This can easily be seen in the network in Fig. 4.1 where, regarding object p, we have $|\mathsf{src}_N(h)| < |\mathsf{src}_N(i)|$ (hence $h \prec_N^T i$) and, regarding object o, we have $|\mathsf{src}_N(f)| = |\mathsf{src}_N(g)|$ (hence $f \approx_N^T g$). Hence $\mathsf{facts}_N(s)$ is less believable than $\mathsf{facts}_N(t)$. If Coherence held this would give $s \sqsubset_N^T t$, but then $\mathsf{src}_N(f)$ is less trustworthy than $\mathsf{src}_N(g)$, giving $f \prec_N^T g$ – a contradiction. From this discussion and Theorem 4.4.1 we obtain as a corollary the following first impossibility result for truth discovery.

Theorem 4.4.2. *There is no TD operator satisfying Coherence, Symmetry, Monotonicity and POI.*

Given that Theorem 4.4.1 characterises the fact ranking of *Voting* for facts relating to a single object, it is natural to ask if there is a stronger form of independence that guarantees this behaviour across *all* facts. As our next result shows, the answer is *yes*, and the necessary axiom is obtained by ignoring the role of objects altogether for fact ranking.

Axiom 4.4.7 (Strong Independence). *For any networks N_1, N_2 and facts f_1, f_2, if $\mathsf{src}_{N_1}(f_j) = \mathsf{src}_{N_2}(f_j)$ for each $j \in \{1,2\}$ then $f_1 \preceq^T_{N_1} f_2$ iff $f_1 \preceq^T_{N_2} f_2$.*

That is, the ranking of two facts f_1 and f_2 is determined solely by the sources claiming f_1 and f_2. In particular, the fact-object affiliations and claims for facts other than f_1, f_2 are irrelevant when deciding on f_1 versus f_2. Note that Strong Independence implies POI. We have the following result.

Theorem 4.4.3. *Suppose $|\mathcal{O}| \geq 3$. Then an operator T satisfies Strong Independence, Monotonicity and Symmetry if and only if for any network N and $f_1, f_2 \in \mathcal{F}$ we have*

$$f_1 \preceq^T_N f_2 \iff |\mathsf{src}_N(f_1)| \leq |\mathsf{src}_N(f_2)|$$

Theorem 4.4.3 can be seen as a characterisation of the class of TD operators that rank facts in the same way as *Voting*. The proof is similar to that of Theorem 4.4.1, but uses a different transformation to obtain a modified network N' in the first step.

We have established that neither POI nor Strong Independence are satisfactory axioms for truth discovery, and a weaker independence property is required. Figure 4.1 can help us once again in this regard. Whereas POI and Strong Independence would say that facts h and i are irrelevant to f, the argument with Coherence for Theorem 4.4.2 suggests otherwise due the indirect links via the sources. We therefore propose that only when there is no (undirected) path between two nodes can we consider them to be truly irrelevant to each other. That is, nodes are relevant to each other iff they lie in the same *connected component* of the network.

Our final rendering of independence states that the ordering of two facts in the same connected component does not depend on any claims outside of the component, and similarly for sources.

Axiom 4.4.8 (Per-component Independence (PCI)). *For any TD networks N_1, N_2 with a common connected component G, the restrictions of $\sqsubseteq^T_{N_1}$ and $\sqsubseteq^T_{N_2}$ to $G \cap \mathcal{S}$ are equal, and the restrictions of $\preceq^T_{N_1}$ and $\preceq^T_{N_2}$ to $G \cap \mathcal{F}$ are equal; that is, $s_1 \sqsubseteq^T_{N_1} s_2$ iff $s_1 \sqsubseteq^T_{N_2} s_2$ and $f_1 \preceq^T_{N_1} f_2$ iff $f_1 \preceq^T_{N_2} f_2$ for $s_1, s_2 \in G \cap \mathcal{S}$ and $f_1, f_2 \in G \cap \mathcal{F}$.*

In analogy with Source/Fact Coherence and Source/Fact Symmetry, it is possible to split the two requirements of PCI into sub-axioms Source-PCI (in which only the constraint on source ranking is imposed) and Fact-PCI (in which only the fact ranking is constrained).

Note that while our framework can be easily adapted to require *by definition* that a network is itself connected (and therefore has only one connected component), we have found that datasets with multiple connected components do indeed occur in practise.[7] This means that failure of PCI is a real issue, and consequently we consider PCI to be another core axiom that all reasonable operators should satisfy.

4.5 Satisfaction of the axioms

With the axioms formally defined, we can now consider whether they are satisfied by the example operators of Section 4.3. *Voting* can be analysed outright; for *Sums* we require some preliminary results giving sufficient conditions for iterative and recursive operators to satisfy various axioms. It will be seen that neither *Voting* nor *Sums* satisfy all our desirable axioms, but it is possible to modify each operator to gain some improvement with respect to the axioms.

4.5.1 Voting

As the simplest operator, we consider *Voting* first. The following theorem shows that all axioms except Coherence are satisfied. Since Coherence is a fundamental principle of truth discovery, and we actually consider POI and Strong Independence to be *undesirable*, this formally rules out *Voting* as a viable operator.

Theorem 4.5.1. Voting *satisfies Symmetry, Unanimity, Groundedness, Monotonicity, POI, Strong Independence and PCI.* Voting *does not satisfy Coherence.*

The proof is straightforward, and is deferred to Appendix A. Note that once Symmetry, Monotonicity and POI are shown, the fact that *Voting* fails Coherence follows from our impossibility result (Theorem 4.4.2), and Fig. 4.1 serves as an explicit counterexample.

4.5.2 Iterative and recursive operators

In this section we give sufficient conditions for iterative and recursive operators to satisfy various axioms. These results will be useful in what follows when analysing *Sums*, although they may also be applied more generally to other operators.

Coherence. To analyse whether the limit of a recursive operator satisfies Coherence, we consider how the update function U behaves when the difference in belief scores between the facts of s_1 and s_2 is 'small' (and similarly for the sources of f_1, f_2). To that end, we introduce a numerical variant of a set of facts Y being 'less believable' than Y'.

[7] For example, the *Book* and *Restaurant* datasets found at the following web page each contain two connected components: http://lunadong.com/fusionDataSets.htm

Definition 4.5.1. *Let T be a numerical TD operator, N a network, $Y, Y' \subseteq \mathcal{F}$ and $\varepsilon, \rho > 0$. We say Y is (ε, ρ)-*less believable* than Y' with respect to N and T if there is a bijection $\varphi : Y \to Y'$ such that $T_N(f) - T_N(\varphi(f)) \leq \varepsilon$ for all $f \in Y$, and $T_N(\hat{f}) - T_N(\varphi(\hat{f})) \leq \varepsilon - \rho$ for some $\hat{f} \in Y$.*

*For $X, X' \subseteq \mathcal{S}$, we define X (ε, ρ)-*less trustworthy* than X' similarly.*

This generalises Definition 4.4.1 by relaxing the requirement that $f \preceq_N^T \varphi(f)$, and instead requiring that f can only be more believable than $\varphi(f)$ by some threshold $\varepsilon > 0$. Definition 4.4.1 is recovered in the limiting case $\varepsilon \to 0$. We obtain a sufficient condition on the update function U for a recursive operator to satisfy Source-Coherence.

Lemma 4.5.1. *Let $U : \mathcal{T}_{Num} \to \mathcal{T}_{Num}$. For any prior operator T^{prior}, $\mathsf{rec}(T^{prior}, U)$ satisfies Source-Coherence if the following condition is satisfied: there exist $C, D > 0$ such that for all networks N and numerical operators T it holds that if $\mathsf{facts}_N(s_1)$ is (ε, ρ)-less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T, then $T'_N(s_1) - T'_N(s_2) \leq C\varepsilon - D\rho$, where $T' = U(T)$.*

The proof of Lemma 4.5.1 uses the following result, the proof of which is a straightforward application of the definition of the limit.

Lemma 4.5.2. *Let N be a truth discovery network and $(T^n)_{n \in \mathbb{N}}$ be a convergent iterative operator with limit T^*. Then for $f_1, f_2 \in \mathcal{F}$, $f_1 \preceq_N^{T^*} f_2$ if and only if*

$$\forall \varepsilon > 0 \; \exists K \in \mathbb{N} : \forall n \geq K : T_N^n(f_1) - T_N^n(f_2) \leq \varepsilon$$

Also, $f_1 \prec_N^{T^*} f_2$ if and only if

$$\exists \rho > 0 : \forall \varepsilon > 0 \; \exists K \in \mathbb{N} : \forall n \geq K : T_N^n(f_1) - T_N^n(f_2) \leq \varepsilon - \rho$$

Analogous statements for source rankings also hold.

Proof of Lemma 4.5.1. Let N be a network. Suppose U has the stated property and that $\mathsf{rec}(T^{\text{prior}}, U) = (T^n)_{n \in \mathbb{N}}$ converges to T^*. Suppose $\mathsf{facts}_N(s_1)$ is less trustworthy than $\mathsf{facts}_N(s_2)$ with respect to N and T^* under a bijection φ. We must show that $s_1 \sqsubset_N^{T^*} s_2$.

Now, there is some $\hat{f} \in \mathsf{facts}_N(s_1)$ with $\hat{f} \prec_N^{T^*} \varphi(\hat{f})$. The second part of Lemma 4.5.2 therefore applies; let ρ be as given there. Now let $\varepsilon > 0$. Since $f \preceq_N^{T^*} \varphi(f)$ for each $f \in \mathsf{facts}_N(s_1)$, we may apply Lemma 4.5.2 with $f, \varphi(f)$ and $\bar{\varepsilon} = \varepsilon / C$ to get that there is $K \in \mathbb{N}$ such that

$$T_N^n(f) - T_N^n(\varphi(f)) \leq \bar{\varepsilon}$$

and

$$T_N^n(\hat{f}) - T_N^n(\varphi(\hat{f})) \leq \bar{\varepsilon} - \rho$$

for all $n \geq K$. In other words, $\mathsf{facts}_N(s_1)$ is $(\bar{\varepsilon}, \rho)$-less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T^n for all $n \geq K$.

Now, recall that $T^{n+1} = U(T^n)$. For $m \geq K' = K + 1$ we therefore have, applying our condition on U,

$$T_N^m(s_1) - T_N^m(s_2) \leq C\bar{\varepsilon} - D\rho = \varepsilon - D\rho$$

Since $D\rho$ is positive and does not depend on ε, we get $s_1 \sqsubset_N^{T^*} s_2$ by Lemma 4.5.2. This shows that T^* satisfies Source-Coherence. □

A similar result gives conditions under which Fact-Coherence is satisfied.

Lemma 4.5.3. $\mathsf{rec}(T^{prior}, U)$ *satisfies Fact-Coherence if there exist* $E, F > 0$ *such that for all networks* N *and numerical operators* T *it holds that if* $\mathsf{src}_N(f_1)$ *is* (ε, ρ)*-less trustworthy than* $\mathsf{src}_N(f_2)$ *with respect to* N *and* T'*, then* $T'_N(f_1) - T'_N(f_2) \leq E\varepsilon - F\rho$*, where* $T' = U(T)$*.*

Proof. The proof proceeds in an identical way to Lemma 4.5.1; the only difference is that we may simply take $K' = K$ in the final step. □

Note that there is asymmetry between Lemma 4.5.1 and Lemma 4.5.3 – in the condition on U in Lemma 4.5.1 we have $\mathsf{facts}_N(s_1)$ (ε, ρ)-less trustworthy than $\mathsf{facts}_N(s_2)$ with respect to T, whereas in Lemma 4.5.3 the corresponding condition is with respect to $T' = U(T)$. This reflects the manner in which *Sums* and other TD operators are typically defined: source trust scores are updated based on the fact scores of the previous iteration, whereas fact belief scores are updated based on the (new) trust scores in the *current* iteration.

Also note that the above results still hold if U has the stated property only for 'small' ε; that is, if there is a constant $0 < \lambda < 1$ such that the property holds for all ρ and for all $\varepsilon < \lambda\rho$.

Symmetry and PCI. When considering either Symmetry or PCI for an iterative operator $(T^n)_{n\in\mathbb{N}}$, it is not enough to know that each T^n satisfies the relevant axiom. The following example illustrates this fact for Symmetry.

Example 4.5.1. *Fix some* $\hat{f} \in \mathcal{F}$*, and define an iterative operator by*

$$T_N^n(s) = 1$$

$$T_N^n(f) = \begin{cases} |\mathsf{src}_N(f)| + (1 - \frac{1}{n+1}) & \textit{if } |\mathsf{src}_N(f)| = |\mathsf{src}_N(\hat{f})| \\ |\mathsf{src}_N(f)| & \textit{otherwise} \end{cases}$$

That is, each T^n *is a modification of* Voting *in which we boost the score of all facts tied with* $\hat{f}$ *under* Voting *by* $1 - \frac{1}{n+1}$*. Since this additional weight is (strictly) less than 1 for each*

n, the ordinal operator induced by T^n is simply Voting*, and therefore satisfies Symmetry. However, it is easy to see that the limit operator T^* has $T^*_N(\hat{f}) = |\mathsf{src}_N(\hat{f})| + 1$; this means T^* uses extra information beyond the structure of the network N in its ranking (namely, the identity of a selected fact $\hat{f}$) which violates Symmetry.*

Using a similar tactic, one can construct a sequence of numerical operators $(T^n)_{n\in\mathbb{N}}$ such that each T^n satisfies PCI, but the limit operator T^* does not.

Fortunately, there is a natural strengthening of both Symmetry and PCI for numerical operators which *is* preserved in the limit. Let us say that a numerical operator T satisfies *numerical Symmetry* if for any equivalent networks $N, \pi(N)$ we have $T_N(z) = T_{\pi(N)}(\pi(z))$ for all $z \in \mathcal{S} \cup \mathcal{F}$. Similarly, T satisfies *numerical PCI* if for any networks N_1 and N_2 with a common connected component G, we have $T_{N_1}(z) = T_{N_2}(z)$ for all $z \in G \cap (\mathcal{S} \cup \mathcal{F})$. Clearly numerical Symmetry implies Symmetry, and numerical PCI implies PCI. The following result is immediate.

Lemma 4.5.4. *Suppose $(T^n)_{n\in\mathbb{N}}$ converges to T^*. Then*

- *If T^n satisfies numerical Symmetry for each $n \in \mathbb{N}$, then T^* satisfies Symmetry.*
- *If T^n satisfies numerical PCI for each $n \in \mathbb{N}$, then T^* satisfies PCI.*

As a consequence of Lemma 4.5.4, any recursive operator $\mathsf{rec}(T^{\text{prior}}, U)$ satisfies Symmetry whenever T^{prior} satisfies numerical Symmetry and U *preserves* numerical Symmetry, in the sense that $U(T)$ satisfies numerical Symmetry whenever T does (and similarly for PCI).

Unanimity, Groundedness and Monotonicity. In contrast to Symmetry and PCI, both Unanimity and Groundedness *are* preserved when taking the limit of an iterative operator.

Lemma 4.5.5. *Suppose $(T^n)_{n\in\mathbb{N}}$ converges to T^*. Then*

- *If T^n satisfies Unanimity for each $n \in \mathbb{N}$, then T^* satisfies Unanimity.*
- *If T^n satisfies Groundedness for each $n \in \mathbb{N}$, then T^* satisfies Groundedness.*

For Monotonicity, we require the following (stronger) property to hold for each T^n.

Definition 4.5.2. *A numerical operator T satisfies* Improvement *if for each N, N' and f as in the statement of Monotonicity, we have $\delta(f) > \delta(g)$ for all $g \neq f$, where*

$$\delta(g) = T_{N'}(g) - T_N(g)$$

In this case we write $\rho_{N,N'} = \min_{g\neq f}(\delta(f) - \delta(g)) > 0$.

Here $\delta(g)$ is the amount by which the belief score for g increases when going from the network N to N'. Improvement simply says that when adding a new source to a fact f, it is f that sees the largest increase.

Proposition 4.5.1. *Suppose $(T^n)_{n\in\mathbb{N}}$ converges to T^*, and T^n satisfies Improvement for each $n \in \mathbb{N}$. Suppose also that $\inf_{n\in\mathbb{N}} \rho^n_{N,N'} > 0$ for each N, N' arising in the statement of Monotonicity. Then T^* satisfies Monotonicity.*

Proof. Let N, N' and f be as in the statement of Monotonicity, and suppose $g \preceq^{T^*}_N f$ for some $g \neq f$. We will show $g \prec^{T^*}_{N'} f$ using Lemma 4.5.2.

Write $\rho^* = \inf_{n\in\mathbb{N}} \rho^n_{N,N'} > 0$ and let $\varepsilon > 0$. Since $g \preceq^{T^*}_N f$, there is $K \in \mathbb{N}$ such that $T^n_N(g) - T^n_N(f) \leq \varepsilon$ for all $n \geq K$. For such n, we have

$$\begin{aligned}
T^n_{N'}(g) - T^n_{N'}(f) &= (T^n_N(g) + \delta^n(g)) - (T^n_N(f) + \delta^n(f)) \\
&= \underbrace{T^n_N(g) - T^n_N(f)}_{\leq\varepsilon} - \underbrace{(\delta^n(f) - \delta^n(g))}_{\geq \rho^n_{N,N'}} \\
&\leq \varepsilon - \rho^n_{N,N'} \\
&\leq \varepsilon - \rho^*
\end{aligned}$$

By Lemma 4.5.2, we have $g \prec^{T^*}_{N'} f$ as required. □

The requirement that $\inf_{n\in\mathbb{N}} \rho^n_{N,N'} > 0$ is a technical condition which ensures the *strict* inequality $g \prec^{T^*}_{N'} f$ holds in the limit, as required for Monotonicity. If this condition fails T^* still satisfies a natural 'weak Monotonicity' axiom, in which the strict inequality $g \prec^{T^*}_{N'} f$ is replaced with $g \preceq^{T^*}_{N'} f$.

4.5.3 Sums

We come to the axiomatic analysis of *Sums*. Coherence and the simpler axioms are satisfied here, and the undesirable independence axioms (POI and Strong Independence) are not. However, Monotonicity and PCI do *not* hold. Since PCI is one of our most important axioms that we expect any reasonable operator to satisfy, this potentially limits the usefulness of *Sums* in practise.

Theorem 4.5.2. Sums *satisfies Coherence, Symmetry, Unanimity and Groundedness.* Sums *does not satisfy POI, Strong Independence, PCI or Monotonicity.*

Proof (sketch). Symmetry, Unanimity and Groundedness can be easily shown from Lemma 4.5.4 and Lemma 4.5.5; the details can be found in the appendix. In the remainder of the proof, $(T^n)_{n\in\mathbb{N}}$ will denote the iterative operator *Sums*, T^* will denote the limit operator, and $U = \text{norm} \circ U^{\text{Sums}}$ will denote the update function for *Sums*.

Coherence. We will show Source-Coherence using Lemma 4.5.1. The argument for Fact-Coherence is similar (using Lemma 4.5.3) and can be found in the appendix.

Suppose $N \in \mathcal{N}$, $T \in \mathcal{T}_{Num}$, $\varepsilon, \rho > 0$, and $\mathsf{facts}_N(s_1)$ is (ε, ρ)-less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T under a bijection $\varphi : \mathsf{facts}_N(s_1) \to \mathsf{facts}_N(s_2)$. By definition there is $\hat{f} \in \mathsf{facts}_N(s_1)$ such that $T_N(\hat{f}) - T_N(\varphi(\hat{f})) \leq \varepsilon - \rho$. By the remark after the proof of Lemma 4.5.1, we may assume without loss of generality that $\varepsilon < \frac{1}{|\mathcal{F}|}\rho$.

Recall that the update function for *Sums* is $U = \mathsf{norm} \circ U^{\mathsf{Sums}}$. Write $T' = U^{\mathsf{Sums}}(T)$ and $\tilde{T} = U(T) = \mathsf{norm}(U^{\mathsf{Sums}}(T))$ so that $\tilde{T} = \mathsf{norm}(T')$. We must show that $\tilde{T}_N(s_1) - \tilde{T}_N(s_2) \leq C\varepsilon - D\rho$ for some constants $C, D > 0$.

Note at this stage that it is possible to further weaken the hypotheses of Lemma 4.5.1: the result follows if U has the stated property not for *all* operators T, but only for those such that $T = T^n$ for some $n \in \mathbb{N}$. Next, note that if $T'_N(x) = 0$ for all $x \in \mathcal{S}$ then trust and belief scores are 0 in all subsequent iterations, and thus all sources rank equally in the limit T^*. But this means the hypothesis for Source-Coherence cannot be satisfied (there are no strict inequalities). We may therefore assume without loss of generality that $T'_N(x) \neq 0$ for at least one $x \in \mathcal{S}$. Therefore, by definition of norm,

$$\tilde{T}_N(s) = \alpha T'_N(s)$$

where

$$\alpha = \frac{1}{\max_{x \in \mathcal{S}} |T'_N(x)|}$$

Applying the definition of U^{Sums} and using the pairing of $\mathsf{facts}_N(s_1)$ and $\mathsf{facts}_N(s_2)$ via φ, we have

$$\begin{aligned}
\tilde{T}_N(s_1) - \tilde{T}_n(s_2) &= \alpha[T'_N(s_1) - T'_N(s_2)] \\
&= \alpha\left[\sum_{f \in \mathsf{facts}_N(s_1)} T_N(f) - \sum_{f \in \mathsf{facts}_N(s_1)} T_N(\varphi(f))\right] \\
&= \alpha \sum_{f \in \mathsf{facts}_N(s_1)} \Big(T_N(f) - T_N(\varphi(f))\Big) \\
&= \underbrace{\alpha}_{>0}\left[\underbrace{\Big(T_N(\hat{f}) - T_N(\varphi(\hat{f}))\Big)}_{\leq \varepsilon - \rho} + \sum_{f \in \mathsf{facts}_N(s_1) \setminus \{\hat{f}\}} \underbrace{\Big(T_N(f) - T_N(\varphi(f))\Big)}_{\leq \varepsilon}\right] \\
&\leq \alpha\left[\varepsilon - \rho + \sum_{f \in \mathsf{facts}_N(s_1) \setminus \{\hat{f}\}} \varepsilon\right] \\
&\leq \alpha \cdot \underbrace{\Big(|\mathcal{F}|\varepsilon - \rho\Big)}_{<0}
\end{aligned}$$

To complete the proof, we need to find a lower bound for α that is independent of T and N (note that a *lower* bound on α is required since $|\mathcal{F}|\varepsilon - \rho$ is negative). It is here that we use the assumption that $T = T^n$ for some $n \in \mathbb{N}$. Since $T_N^n(x) \in [0, 1]$ for any $n \in \mathbb{N}$ and $x \in S$, we have

$$|T'_N(x)| = T'_N(x) = \sum_{f \in \mathsf{facts}_N(x)} \underbrace{T_N(f)}_{\leq 1} \leq |\mathsf{facts}_N(x)| \leq |\mathcal{F}|$$

and so

$$\alpha = \frac{1}{\max_{x \in \mathcal{S}} |T'_N(x)|} \geq \frac{1}{|\mathcal{F}|}$$

Combining this with the above bound for $\tilde{T}_N(s_1) - \tilde{T}_n(s_2)$, we get

$$\tilde{T}_N(s_1) - \tilde{T}_n(s_2) \leq \frac{1}{|\mathcal{F}|}\Big(|\mathcal{F}|\varepsilon - \rho\Big) = \varepsilon - \frac{1}{|\mathcal{F}|}\rho$$

Taking $C = 1$ and $D = \frac{1}{|\mathcal{F}|}$, the hypotheses of Lemma 4.5.1 are satisfied; thus *Sums* satisfies Source-Coherence.

POI, Strong Independence, PCI and Monotonicity. The remaining axioms are handled by counterexamples derived from the network shown in Fig. 4.2. It can be shown that, if N denotes this network, we have $T_N^*(f) = T_N^*(g) = 0$, so $f \approx_N^{T^*} g$.

Let N' denote the network whose claims are just those of the top connected component. Then it can be shown that $T_{N'}^*(f) = 1$ and $T_{N'}^*(g) = 0$, i.e. $g \prec_{N'}^{T^*} f$. However it is easily verified that our three independence axioms, if satisfied, would each imply $f \preceq_N^{T^*} g$ iff $f \preceq_{N'}^{T^*} g$. Therefore none of POI, Strong Independence and PCI can hold for *Sums*.

For Monotonicity, consider the network N'' obtained from N by removing the edge (u, g). Then we still have $T_{N''}^*(f) = T_{N''}^*(g) = 0$, and in particular $f \preceq_{N''}^{T^*} g$. Returning to N amounts to adding extra support for the fact g. Monotonicity would give $f \prec_N^{T^*} g$ here, but this is clearly false. Hence Monotonicity is not satisfied by *Sums*.

□

The key to the counterexamples derived from Fig. 4.2 in the above proof lies in the lower connected component, which – restricted to $\mathcal{S} \cup \mathcal{F}$ – is a *connected* bipartite graph. That is, each source x_i claims all facts in the component, and each fact y_j is claimed by all sources in the component. Moreover, sources elsewhere in the network claim fewer facts than the x_i, and facts elsewhere are claimed by fewer sources than the y_j.

Since *Sums* assigns scores by a simple sum, this results in the scores for the x_i and y_j dominating those of the other sources and facts. The normalisation step then divides these scores by the (comparatively large) maximum. As the next result

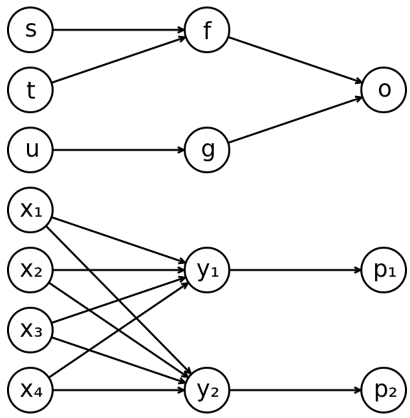


Figure 4.2: Network which yields counterexamples for POI, Strong Independence, PCI and Monotonicity for *Sums*

shows, under certain conditions this causes scores to decrease *exponentially* and become 0 in the limit. In particular, we can generate pathological examples such as Fig. 4.2 where a whole connected component receives scores of 0, which leads to failure of Monotonicity and the independence axioms.

Proposition 4.5.2. *Let N be a network. Suppose there is $X \subseteq \mathcal{S}$, $Y \subseteq \mathcal{F}$ such that*

1. $\mathsf{facts}_N(x) = Y$ *for each* $x \in X$
2. $\mathsf{src}_N(y) = X$ *for each* $y \in Y$
3. $\mathsf{facts}_N(s) \cap Y = \emptyset$ *and* $|\mathsf{facts}_N(s)| \leq \frac{|Y|}{2}$ *for each* $s \in \mathcal{S} \setminus X$
4. $\mathsf{src}_N(f) \cap X = \emptyset$ *and* $|\mathsf{src}_N(f)| \leq \frac{|X|}{2}$ *for each* $f \in \mathcal{F} \setminus Y$

Then, with $(T^n)_{n\in\mathbb{N}}$ *denoting* Sums, *for all* $n > 1$ *we have*

$$T^n_N(s) \leq \frac{1}{2^{n-1}} \quad (s \in \mathcal{S} \setminus X)$$

$$T^n_N(f) \leq \frac{1}{2^{n-1}} \quad (f \in \mathcal{F} \setminus Y)$$

$$T^n_N(x) = 1 \quad (x \in X)$$

$$T^n_N(y) = 1 \quad (y \in Y)$$

In particular, if T^* *denotes the limit of* Sums *then* $T^*_N(s) = T^*_N(f) = 0$ *for all* $s \in \mathcal{S} \setminus X$ *and* $f \in \mathcal{F} \setminus Y$.

Proof. We proceed by induction. The result is easy to show in the base case $n = 2$ since $|\mathsf{facts}_N(s)| \leq \frac{1}{2}|\mathsf{facts}_N(x)|$ for any $x \in X$ and $s \notin X$ (and similarly for facts).

Assume the result holds for some $n > 1$. Write $T' = U^{\mathsf{Sums}}(T^n)$, so that $T^{n+1} = \mathsf{norm}(T')$. If $s \notin X$ then $\mathsf{facts}_N(s) \subseteq \mathcal{F} \setminus Y$, so

$$T'_N(s) = \sum_{f \in \mathsf{facts}_N(s)} \underbrace{T^n_N(f)}_{\leq \frac{1}{2^{n-1}}} \leq \frac{|\mathsf{facts}_N(s)|}{2^{n-1}} \leq \frac{\frac{1}{2}|Y|}{2^{n-1}} = \frac{|Y|}{2^{(n+1)-1}}$$

Similarly, if $f \notin Y$ then $\mathsf{src}_N(f) \subseteq \mathcal{S} \setminus X$, so

$$T'_N(f) = \sum_{s \in \mathsf{src}_N(f)} \underbrace{T'_N(s)}_{\leq \frac{|Y|}{2^{(n+1)-1}}} \leq \frac{|\mathsf{src}_N(f)| \cdot |Y|}{2^{(n+1)-1}} \leq \frac{\frac{1}{2}|X| \cdot |Y|}{2^{(n+1)-1}} = \frac{|X| \cdot |Y|}{2^{(n+2)-1}}$$

On the other hand, the fact that $T^n_N(x) = T^n_N(y) = 1$ for $x \in X$ and $y \in Y$ gives

$$T'_N(x) = \sum_{y \in Y} T^n_N(y) = |Y|$$

$$T'_N(y) = \sum_{x \in X} T'_N(x) = |X| \cdot |Y|$$

Clearly the $x \in X$ and $y \in Y$ are the sources and facts with maximal trust and belief scores, respectively. This means that after normalisation via norm, $T^{n+1}_N(x) = T^{n+1}_N(y) = 1$ and for $s \notin X$ and $f \notin Y$,

$$T^{n+1}_N(s) = \frac{T'_N(s)}{|Y|} \leq \frac{1}{2^{(n+1)-1}}$$

$$T^{n+1}_N(f) = \frac{T'_N(f)}{|X| \cdot |Y|} \leq \frac{1}{2^{(n+2)-1}} \leq \frac{1}{2^{(n+1)-1}}$$

This shows that the claim holds for $n + 1$; by induction, the proof is complete. $\square$

4.5.4 Modifying *Voting* and *Sums*

So far we have seen that neither of the basic operators *Voting* or *Sums* are completely satisfactory with respect to the axioms of Section 4.4. Armed with the knowledge of how and why certain axioms fail, one may wonder whether it is possible to modify the operators accordingly so that the axioms *are* satisfied. Presently we shall show that this is partially possible both in the case of *Voting* and *Sums*.

4.5.4.1 Voting

A core problem with *Voting* is that it fails Coherence. Indeed, all sources are ranked equally regardless of the 'votes' for facts, so in some sense it is obvious that the source ranking does not cohere with the fact ranking.[8] An easy improvement is to explicitly construct the source ranking to guarantee Source-Coherence.

[8] Fact-Coherence is vacuously satisfied, however: since all sources rank equally we can never have $\mathsf{src}_N(f_1)$ less trustworthy than $\mathsf{src}_N(f_2)$.

Definition 4.5.3. *For a network N, define a binary relation $\lhd_N$ on $\mathcal{S}$ by $s_1 \lhd_N s_2$ iff* $\mathsf{facts}_N(s_1)$ *is less believable than* $\mathsf{facts}_N(s_2)$ *with respect to* Voting. *The numerical operator* SC-Voting *(Source-Coherence Voting) is defined by*

$$T_N^{SCV}(s) = |\{t \in \mathcal{S} : t \lhd_N s\}|, \quad T_N^{SCV}(f) = |\mathsf{src}_N(f)|$$

It can be seen that *SC-Voting* satisfies Source-Coherence, which is a significant improvement over regular *Voting*. Since $\lhd_N$ relies on 'global' properties on N, however, this comes at the expense of Source-PCI. Satisfaction of the other axioms is inherited from *Voting*.

Theorem 4.5.3. SC-Voting *satisfies Source-Coherence, Symmetry, Unanimity, Groundedness, Monotonicity, Fact-PCI, POI and Strong Independence. It does not satisfy Fact-Coherence or Source-PCI.*

The following properties of $\lhd_N$ are useful for showing Source-Coherence.

Lemma 4.5.6. *$\lhd_N$ is transitive and irreflexive.*

Proof. For transitivity, suppose $s \lhd_N t$ and $t \lhd_N u$. Then $\mathsf{facts}_N(s)$ is less believable than $\mathsf{facts}_N(t)$ (with respect to *Voting*) via some bijection $\varphi : \mathsf{facts}_N(s) \to \mathsf{facts}_N(t)$, and $\mathsf{facts}_N(t)$ is less believable than $\mathsf{facts}_N(u)$ via some bijection $\psi : \mathsf{facts}_N(t) \to \mathsf{facts}_N(u)$. It is easily seen that $\mathsf{facts}_N(s)$ is less believable than $\mathsf{facts}_N(u)$ via the composition $\theta = \psi \circ \varphi$, so $s \lhd_N u$.

For irreflexivity, suppose for contradiction that $s \lhd_N s$ for some $s \in \mathcal{S}$, i.e. $F = \mathsf{facts}_N(s)$ is less believable than itself under some bijection $\varphi : F \to F$. Then $f \preceq_N^T \varphi(f)$ for each $f \in F$, and there is $\hat{f}$ such that $\hat{f} \prec_N^T \varphi(\hat{f})$. Iterating applications of φ, we get

$$\hat{f} \prec_N^T \varphi(\hat{f}) \preceq_N^T \varphi(\varphi(\hat{f}) \preceq_N^T \cdots \preceq_N^T \varphi^n(\hat{f}) \tag{4.2}$$

for each $n \geq 1$, where φ^n is the n-th iterate of φ and T denotes *Voting*.

Since F is finite, the sequence $\varphi(\hat{f}), \varphi(\varphi(\hat{f})), \ldots$ must repeat at some point, i.e. there is $i < j$ such that $\varphi^i(\hat{f}) = \varphi^j(\hat{f})$. But then injectivity of φ implies that $\hat{f} = \varphi^{j-i}(\hat{f})$. Taking $n = j - i$ in Eq. (4.2) we get $\hat{f} \prec_N^T \hat{f}$ – a contradiction. □

Proof of Theorem 4.5.3 (sketch). Note that *SC-Voting* inherits Unanimity, Groundedness, Monotonicity, Fact-PCI, POI and Strong Independence from *Voting*, since these axioms only refer to the rankings of facts (which is the same for *SC-Voting* as for *Voting*).

We take the remaining axioms in turn. To simplify notation, write $W_N(s) = \{t \in \mathcal{S} : t \lhd_N s\}$ in what follows.

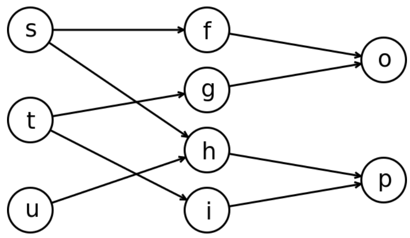


Figure 4.3: Fact-Coherence counterexample for *SC-Voting*

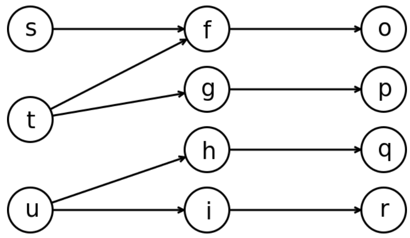


Figure 4.4: Source-PCI counterexample for *SC-Voting*

Source-Coherence. Suppose facts$_N(s_1)$ is less believable than facts$_N(s_2)$ with respect to T^{SCV}. We need to show $s_1 \sqsubset_N^{T^{SCV}} s_2$.

Note that since the fact ranking for T^{SCV} coincides with *Voting*, we have $s_1 \lhd_N s_2$. Transitivity of $\lhd_N$ gives $W_N(s_1) \subseteq W_N(s_2)$. Moreover, $s_1 \in W_N(s_2)$ but by irreflexivity, $s_1 \notin W_N(s_1)$. Therefore $W_N(s_1) \subset W_N(s_2)$, which means $T_N^{SCV}(s_1) = |W_N(s_1)| < |W_N(s_2)| = T_N^{SCV}(s_2)$, i.e. $s_1 \sqsubset_N^{T^{SCV}} s_2$ as required.

Symmetry. Since the fact ranking of T^{SCV} is the same as *Voting*, which satisfies Symmetry, we only need to show that $s_1 \sqsubseteq_N^{T^{SCV}} s_2$ iff $\pi(s_1) \sqsubseteq_{\pi(N)}^{T^{SCV}} \pi(s_2)$ for all equivalent networks $N, \pi(N)$ and $s_1, s_2 \in \mathcal{S}$.

In can be shown, and we do so in the appendix, that the Symmetry of *Voting* implies a symmetry property for $\lhd_N$ and $\lhd_{\pi(N)}$: we have $s_1 \lhd_N s_2$ iff $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$. Consequently, $t \in W_N(s_i)$ iff $\pi(t) \in W_{\pi(N)}(\pi(s_i))$; in particular, $|W_N(s_i)| = |W_{\pi(N)}(\pi(s_i))|$. This means

$$\begin{aligned} s_1 \sqsubseteq_N^{T^{SCV}} s_2 &\iff |W_N(s_1)| \leq |W_N(s_2)| \\ &\iff |W_{\pi(N)}(\pi(s_1))| \leq |W_{\pi(N)}(\pi(s_2))| \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^{SCV}} \pi(s_2) \end{aligned}$$

as required for Symmetry.

Fact-Coherence. Consider the network shown in Fig. 4.3. We have $f \approx g \approx i \prec h$. Source-Coherence between s and t gives $t \sqsubset s$. If Fact-Coherence held we would then get $g \prec f$, but this is not the case.

Source-PCI. Let N_1 denote the top connected component of the network shown in Fig. 4.4, and let N_2 denote the network as a whole. The fact ranking is the same in both networks: $g \approx h \approx i \prec f$. In N_1 sources s and t claim a different number of facts, so neither $s \lhd_{N_1} t$ nor $t \lhd_{N_1} s$. Consequently $W_{N_1}(s) = W_{N_1}(t) = \emptyset$ and $s \simeq^{T^{SCV}}_{N_1} t$.

In N_2 sources t and u can be compared for Source-Coherence, and we see that $u \lhd_{N_2} t$ since $i \preceq^{T^{SCV}}_{N_2} g$ and $h \prec^{T^{SCV}}_{N_2} f$. Hence $W_{N_2}(t) = \{u\}$ and $W_{N_2}(s) = \emptyset$, which means $s \sqsubset^{T^{SCV}}_{N_2} t$. This contradicts Source-PCI, which requires the ranking of s and t to be the same in both networks. □

Note that the idea behind *SC-Voting* can be generalised beyond *Voting*: it is possible to define $\lhd_N$ in terms of *any* operator T, and to construct a new operator T' whose source ranking is given according to $\lhd_N$ as above, and whose fact ranking coincides with that of T. This ensures T' satisfies Source-Coherence whilst keeping the existing fact ranking from T.

Moreover we can go in the other direction and ensure *Fact*-Coherence whilst retaining the source ranking of T by defining a relation $\blacktriangleleft_N$ on $\mathcal{F}$ in a analogous manner to $\lhd_N$, and proceeding similarly.

4.5.4.2 Sums

Our main concern with *Sums* is the failure of PCI and Monotonicity. We have seen that this is in some sense caused by the normalisation step: in Fig. 4.2 the scores of f, g etc go to 0 in the limit after dividing the 'global' maximum score across the network, which happens to come from a different connected component.

A natural fix for PCI is to therefore divide by the maximum score *within each component*. In this case the score for a source s depends only on the structure of the connected component in which it lies, which is exactly what is required for PCI.

However, this approach does not negate the undesirable effects of Proposition 4.5.2. Indeed, suppose the network in Fig. 4.2 was modified so that fact y_1 is associated with object o instead of p_1. Clearly Proposition 4.5.2 still applies after this change, and all sources and facts shown now belong to the same connected component. Therefore the 'per-component *Sums*' operator gives the same result as *Sums* itself, and in particular our Monotonicity counterexample still applies. Perhaps even worse, one can show that Coherence fails for this operator. We consider the loss of Coherence too high a price to pay for PCI.

Instead, let us take a step back and consider if normalisation is truly necessary. On the one hand, without normalisation the trust and belief scores are unbounded and therefore do not converge. On the other, we are not interested in the numeric scores for their own sake, but rather for the *rankings* that they induce. It may be possible that whilst the scores diverge without normalisation, the induced rankings *do* converge to a fixed one, which we may take as the 'ordinal limit'. This is in fact the case. We call this new operator *UnboundedSums*.

Definition 4.5.4. UnboundedSums *is the recursive operator* $\mathsf{rec}(T^{prior}, U^{Sums})$ *where* $T_N^{prior}(s) = 1$, $T_N^{prior}(f) = |\mathsf{src}_N(f)|$ *and* U^{Sums} *is defined as in Section 4.3.2.*[9]

Theorem 4.5.4. UnboundedSums *is ordinally convergent in the following sense: there is an ordinal operator* T^* *such that for each network* N *there exists* $J_N \in \mathbb{N}$ *such that* $T_N^n(s_1) \leq T_N^n(s_2)$ *iff* $s_1 \sqsubseteq_N^{T^*} s_2$ *for all* $n \geq J_N$ *and* $s_1, s_2 \in \mathcal{S}$ *(and similarly for facts).*

That is, the rankings induced by T^n *remain constant after* J_N *iterations, and are identical to those of* T^**.*

Proof. The proof will use some results from linear algebra, so we will work with a matrix and vector representation of *UnboundedSums*. Fix an enumeration $\mathcal{S} = \{s_1, \ldots, s_k\}$ of $\mathcal{S}$ and $\mathcal{F} = \{f_1, \ldots, f_l\}$ of $\mathcal{F}$. Write M for the $k \times l$ matrix given by

$$[M]_{ij} = \begin{cases} 1 & \text{if } s_i \in \mathsf{src}_N(f_j) \\ 0 & \text{otherwise} \end{cases} \qquad (1 \leq i \leq k, 1 \leq j \leq l)$$

We also write v_n and w_n for the vectors of trust and belief scores of *UnboundedSums* at iteration n; that is

$$v_n = [T_N^n(s_1), \ldots, T_N^n(s_k)]^\top \in \mathbb{R}^k$$
$$w_n = [T_N^n(f_1), \ldots, T_N^n(f_l)]^\top \in \mathbb{R}^l$$

where $(T^n)_{n \in \mathbb{N}}$ denotes *UnboundedSums*.

Multiplication by M encodes the update step of *UnboundedSums*: it is easily shown that $v_{n+1} = Mw_n$ and $w_{n+1} = M^\top v_{n+1}$. Writing $A = MM^\top \in \mathbb{R}^{k \times k}$, we have $v_{n+1} = Av_n$, and therefore $v_{n+1} = A^n v_1$.

To show that the rankings of *UnboundedSums* remain constant after finitely many iterations, we will show that for each $s_p, s_q \in \mathcal{S}$ there is $J_{pq} \in \mathbb{N}$ such that $\operatorname{sign}([v_n]_p - [v_n]_q)$ is constant for all $n \geq J_{pq}$. Since $[v_n]_p$ and $[v_n]_q$ are the trust scores of s_p and s_q respectively in the n-th iteration, this will show that the ranking of s_p and s_q remains the same after J_{pq} iterations. Since there are only finitely many pairs of sources, we may then take J_N as the maximum value of J_{pq} over all pairs (p, q), and the entire source ranking $\sqsubseteq_N^{T^n}$ of *UnboundedSums* remains constant for $n \geq J_N$. An almost identical argument can be carried out for the fact ranking, and these together will prove the result.

So, fix $s_p, s_q \in \mathcal{S}$. Write $\delta_n = [v_n]_p - [v_n]_q$. First note that $A = MM^\top$ is symmetric, so the *spectral theorem* gives the existence of k orthogonal eigenvectors $x_1, \ldots, x_k$ for

[9] Note that to simplify proof of ordinal convergence we use a different prior operator to *Sums*, but this does not effect the operator in any significant way.

A [7, Theorem 7.29]. Let $\lambda_1, \dots, \lambda_k$ be the corresponding eigenvalues. Form a $(k \times k)$-matrix P whose i-th column is x_i, and let $D = \text{diag}(\lambda_1, \dots, \lambda_k)$. Then A can be diagonalised as $A = PDP^{-1}$. It follows that for any $n \in \mathbb{N}$, $A^n = PD^nP^{-1}$.

Now, since $x_1, \dots, x_k$ are orthogonal, P is an orthogonal matrix, i.e. $P^\top = P^{-1}$. Hence $A^n = PD^nP^\top$. Note that

$$PD^n = \begin{bmatrix} x_1 \mid \dots \mid x_k \end{bmatrix} \begin{bmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_k^n \end{bmatrix} = \begin{bmatrix} \lambda_1^n x_1 \mid \dots \mid \lambda_k^n x_k \end{bmatrix}$$

and

$$P^\top v_1 = \begin{bmatrix} x_1 \\ - \\ \vdots \\ - \\ x_k \end{bmatrix} v_1 = \begin{bmatrix} x_1 \cdot v_1 \\ \vdots \\ x_k \cdot v_1 \end{bmatrix}$$

which means

$$v_{n+1} = A^n v_1 = PD^nP^\top v_1 = \begin{bmatrix} \lambda_1^n x_1 \mid \dots \mid \lambda_k^n x_k \end{bmatrix} \begin{bmatrix} x_1 \cdot v_1 \\ \vdots \\ x_k \cdot v_1 \end{bmatrix} = \sum_{i=1}^{k} (x_i \cdot v_1) \lambda_i^n x_i$$

We obtain an explicit formula for δ_{n+1}:

$$\delta_{n+1} = [v_n]_p - [v_n]_q = \sum_{i=1}^{k} (x_i \cdot v_1) \lambda_i^n \left([x_i]_p - [x_i]_q \right) = \sum_{i=1}^{k} r_i \lambda_i^n \tag{4.3}$$

where $r_i = (x_i \cdot v_1) \left([x_i]_p - [x_i]_q \right)$. Note that r_i does not depend on n.

Now, it is easy to see that $A = MM^\top$ is *positive semi-definite*, which means its eigenvalues $\lambda_1, \dots, \lambda_k$ are all non-negative. We re-index the sum in Eq. (4.3) by grouping together the λ_i which are equal, to get

$$\delta_{n+1} = \sum_{t=1}^{K} R_t \mu_t^n$$

where $K \leq k$, each R_t is a sum of the r_i (whose corresponding λ_i are equal), and the μ_t are distinct and non-negative. Assume without loss of generality that $\mu_1 > \mu_2 > \dots > \mu_K \geq 0$. If $R_t = 0$ for all t, then clearly $\text{sign}(\delta_{n+1}) = \text{sign}(0) = 0$ which is constant, so we are done. Otherwise, let T be the minimal t such that $R_t \neq 0$. We may also assume $\mu_T > 0$ (otherwise we necessarily have $\mu_T = 0, T = K$ and $\text{sign}(\delta_{n+1}) = \text{sign}(R_T \cdot 0^n)$ which is again constant 0). Observe that

$$\delta_{n+1} = R_T \mu_T^n + \sum_{t=T+1}^{K} R_t \mu_t^n = \mu_T^n \left[R_T + \sum_{t=T+1}^{K} R_t \left(\frac{\mu_t}{\mu_T} \right)^n \right]$$

By our assumption on the ordering of the μ_t, we have $\mu_t < \mu_T$ in the sum. Consequently $|\mu_t/\mu_T| < 1$, and $(\mu_t/\mu_T)^n \to 0$ as $n \to \infty$. This means

$$\lim_{n\to\infty}\left[R_T + \sum_{t=T+1}^{K} R_t \underbrace{\left(\frac{\mu_t}{\mu_T}\right)^n}_{\to 0}\right] = R_T \neq 0$$

Since this limit is non-zero, there is $J_{pq} \in \mathbb{N}$ such that the sign of term in square brackets is equal to $S = \operatorname{sign} R_T \in \{1, -1\}$ for all $n \geq J_{pq}$. Finally, for such n we have

$$\operatorname{sign}\delta_{n+1} = \operatorname{sign}\left(\underbrace{\mu_T^n}_{>0}\left[R_T + \sum_{t=T+1}^{K} R_t \left(\frac{\mu_t}{\mu_T}\right)^n\right]\right) = \operatorname{sign}\left(R_T + \sum_{t=T+1}^{K} R_t \left(\frac{\mu_t}{\mu_T}\right)^n\right) = S$$

i.e. $\operatorname{sign}\delta_n$ is constant for $n \geq J_{pq} + 1$. This completes the proof.[10] □

In light of Theorem 4.5.4, we may consider *UnboundedSums* itself as an ordinal operator T^*, where $s \sqsubseteq_N^{T^*} t$ iff $s \sqsubseteq_N^{T^{J_n}} t$ for each network N (and similarly for the fact ranking). Moreover, with the normalisation problems aside, *UnboundedSums* provides an example of a principled operator satisfying our two key axioms – Coherence and PCI. In particular, we escape the undesirable behaviour of *Sums* in Fig. 4.2; whereas *Sums* trivialises the ranking of sources and facts in the upper connected component, *UnboundedSums* allows meaningful ranking (e.g. we have $g \prec f$). In particular, the counterexample for Monotonicity for *Sums* is no longer a counterexample for *UnboundedSums*. We conjecture that *UnboundedSums* also satisfies Monotonicity, but this remains to be proven. For example, we have experimentally verified that *UnboundedSums* satisfies all the specific instances of Monotonicity with the starting network N as in Fig. 4.1.

Theorem 4.5.5. UnboundedSums *satisfies Coherence, Symmetry, Unanimity, Groundedness and PCI.* UnboundedSums *does not satisfy POI and Strong Independence.*

Proof (sketch). The proof that *UnboundedSums* satisfies Symmetry, PCI, Unanimity and Groundedness is routine, and we leave the details to the appendix. In what follows, let $(T^n)_{n\in\mathbb{N}}$ denote *UnboundedSums*, T^* denote the ordinal limit of *UnboundedSums*, and for a network N let J_N be as in Theorem 4.5.4. Then the rankings in N induced by T^n for $n \geq J_N$ are the same as T^*.

[10] The argument which shows that the difference between fact belief scores is also eventually constant in sign is almost identical. Write $B = M^\top M$, and observe that $w_{n+1} = B^n w_1$. Since B is also symmetric and positive semi-definite, the proof goes through as above.

Coherence. First we show Source-Coherence. Let N be a network and suppose $\mathsf{facts}_N(s_1)$ is less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T^*. Let φ and $\hat{f}$ be as in the definition of less believable.

Let $n \geq J_N$. Then $f \preceq_N^{T^*} \varphi(f)$ and $\hat{f} \prec_N^{T^*} \varphi(\hat{f})$ for each $f \in \mathsf{facts}_N(s_1)$ means $T_N^n(f) \leq T_N^n(\varphi(f))$ and $T_N^n(\hat{f}) < T_N^n(\varphi(\hat{f}))$. Hence

$$\begin{aligned}
T_N^{n+1}(s) &= \sum_{f \in \mathsf{facts}_N(s_1)} T_N^n(f) \\
&= T_N^n(\hat{f}) + \sum_{f \in \mathsf{facts}_N(s_1) \setminus \{\hat{f}\}} T_N^n(f) \\
&< T_N^n(\varphi(\hat{f})) + \sum_{f \in \mathsf{facts}_N(s_1) \setminus \{\hat{f}\}} T_N^n(\varphi(f)) \\
&= \sum_{f \in \mathsf{facts}_N(s_1)} T_N^n(\varphi(f)) \\
&= \sum_{g \in \mathsf{facts}_N(s_2)} T_N^n(g) \\
&= T_N^{n+1}(s_2)
\end{aligned}$$

i.e. $T_N^{n+1}(s_1) < T_N^{n+1}(s_2)$. But T_N^{n+1} gives the same ranking as T_N^n and therefore the same ranking as T^*, so we get $s_1 \sqsubset_N^{T^*} s_2$ as required.

For Fact-Coherence, suppose $\mathsf{src}_N(f_1)$ is less trustworthy than $\mathsf{src}_N(f_2)$ with respect to N and T^*. Again, let $n \geq J_N$ and φ, $\hat{s}$ be as in the definition of less trustworthy. Recall that belief scores for facts in T_N^n are obtained from trust scores in T_N^n; applying the same argument as above we get $T_N^n(f_1) < T_N^n(f_2)$ and consequently $f_1 \preceq_N^{T^*} f_2$ as required. Hence T^* satisfies Coherence.

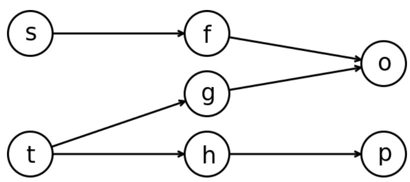


Figure 4.5: Counterexample for POI and Strong Independence for *UnboundedSums*

POI and Strong Independence. To show POI and Strong Independence are not satisfied, consider the network N shown in Fig. 4.5. It can be seen (e.g. by induction) that

$$T_N^n(f) = 1, \quad T_N^n(g) = 2^{n-1}$$

for all $n \in \mathbb{N}$. Consequently $f \prec_N^{T^*} g$.[11]

[11] Note that g ranks higher than f in this network simply because t makes more claims than s, and each fact is claimed only by a single source. This questionable property of *UnboundedSums* is inherited from *Sums*.

Table 4.1: Satisfaction of the axioms for the various operators. Recall that POI and Strong Independence are undesirable properties.

	Voting	SC-Voting	Sums	U-Sums
Source-Coherence	X	✓	✓	✓
Fact-Coherence	✓	X	✓	✓
Symmetry	✓	✓	✓	✓
Unanimity	✓	✓	✓	✓
Ground.	✓	✓	✓	✓
Mon.	✓	✓	X	?
Source-PCI	✓	X	X	✓
Fact-PCI	✓	✓	X	✓
POI	✓	✓	X	X
Str. Indep.	✓	✓	X	X

Now let N' be the network in which the claim (t, h) is removed. Since $\mathsf{src}_N(f) = \mathsf{src}_{N'}(f) = \{s\}$ and $\mathsf{src}_N(g) = \mathsf{src}_{N'}(g) = \{t\}$, both POI and Strong Independence imply $f \preceq_N^{T^*} g$ iff $f \preceq_{N'}^{T^*} g$. Therefore assuming either of POI or Strong Independence we get $f \prec_{N'}^{T^*} g$. However is is also clear that

$$T_{N'}^n(f) = T_{N'}^n(g) = 1$$

for all $n \in \mathbb{N}$, so $f \approx_{N'}^{T^*} g$. This is a contradiction, so neither POI nor Strong Independence are satisfied. □

To summarise this section, Table 4.1 shows which axioms are satisfied by each of the operators.

4.6 Variable domain truth discovery

So far, we have considered an arbitrary but fixed (finite) domain of sources, facts and objects $(\mathcal{S}, \mathcal{F}, \mathcal{O})$. Our operators and axioms were defined with respect to this domain. However, the operators do not *depend* on the domain: they can be defined for *any* choice of $\mathcal{S}$, $\mathcal{F}$ and $\mathcal{O}$. In this section we generalise the framework so that these sets are no longer fixed. This allows new situations to be modelled, such as new sources entering the network. Adapting the definition of a TD operator to this case, we can then see how the ranking of facts changes as new sources are added. Such variable domain operators are then analogues of *variable electorate voting rules* in social theory.

Formally, let $\mathbb{S}$, $\mathbb{F}$ and $\mathbb{O}$ be countably infinite sets of sources, facts and objects respectively. A *domain* is a triple $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, where $\mathcal{S} \subseteq \mathbb{S}$, $\mathcal{F} \subseteq \mathbb{F}$ and $\mathcal{O} \subseteq \mathbb{O}$ are finite, non-empty sets. We think of $\mathbb{S}$, $\mathbb{F}$ and $\mathbb{O}$ as being the 'universe' of possible sources, facts and objects, and a domain as the (finite) sets of entities under consideration in a particular TD problem. Given a domain $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, we define $\mathcal{D}$-networks and $\mathcal{D}$-operators as in Definitions 4.2.1 and 4.2.2.

Definition 4.6.1. *A* variable domain operator T *is a mapping which maps each domain $\mathcal{D}$ to a $\mathcal{D}$-operator $T_{\mathcal{D}}$.*

Note that for a domain $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ and a $\mathcal{D}$-network N, $\sqsubseteq_N^{T_{\mathcal{D}}}$ and $\preceq_N^{T_{\mathcal{D}}}$ are rankings only over the set of sources $\mathcal{S}$ and $\mathcal{F}$ in the domain $\mathcal{D}$, *not* all of $\mathbb{S}$ and $\mathbb{F}$. If $\mathcal{D}$ is clear from context, we write $\sqsubseteq_N^T$ and $\preceq_N^T$ without explicit reference to the domain.

Since all the axioms so far were stated with respect to a fixed but arbitrary domain, they can be easily lifted to the variable domain case. For instance, we say a variable domain operator T satisfies Coherence if $T_{\mathcal{D}}$ satisfies the instance of Coherence for domain $\mathcal{D}$, for all $\mathcal{D}$, and similar for the other axioms.

But we can now go further, and introduce axioms which make use of *several* domains. First, we generalise Symmetry to act across domains. Say networks N, N' in domains $\mathcal{D}, \mathcal{D}'$ respectively are *equivalent* if there is a graph isomorphism π between them such that $\pi(s) \in \mathcal{S}'$, $\pi(f) \in \mathcal{F}'$ and $\pi(o) \in \mathcal{O}'$ for all $s \in \mathcal{S}$, $f \in \mathcal{F}$ and $o \in \mathcal{O}$.

Axiom 4.6.1 (Isomorphism)**.** *Let N and $N' = \pi(N)$ be equivalent networks. Then for all $s_1, s_2 \in \mathcal{S}$, $f_1, f_2 \in \mathcal{F}$, we have $s_1 \sqsubseteq_N^T s_2$ iff $\pi(s_1) \sqsubseteq_{N'}^T \pi(s_2)$ and $f_1 \preceq_N^T f_2$ iff $\pi(f_1) \preceq_{N'}^T \pi(f_2)$.*

Like Symmetry, Isomorphism simply says that operators only care about the *structure* of the network, not the particular 'names' chosen for sources, facts and objects. Symmetry is obtained as the special case where N and N' are equivalent when seen as networks in a common domain $\mathcal{D}$. All the operators of Sections 4.3 and 4.5.4 satisfy Isomorphism.

Next we introduce a new monotonicity property. In what follows, for a network $N = (V, E)$ in domain $(\mathcal{S}, \mathcal{F}, \mathcal{O})$, $f \in \mathcal{F}$ and $\mathcal{S}' \subseteq \mathbb{S}$ finite and disjoint from $\mathcal{S}$, write $N + (\mathcal{S}', f)$ for the network in domain $(\mathcal{S} \cup \mathcal{S}', \mathcal{F}, \mathcal{O})$ with edge set $E \cup \{(s, f) \mid s \in \mathcal{S}'\}$, i.e. the extension of N where a collection of 'fresh' sources $\mathcal{S}'$ each claim f. For example, Fig. 4.6 shows $N + (\mathcal{S}', h)$ for the network N from Fig. 4.1 and new sources $\mathcal{S}' = \{x_1, \ldots, x_4\}$.

Axiom 4.6.2 (Eventual Monotonicity)**.** *Let $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ be a domain and N a $\mathcal{D}$-network. Then for all $f, g \in \mathcal{F}$, $f \neq g$, there is a finite, non-empty set $\mathcal{S}' \subseteq \mathbb{S}$ with $\mathcal{S} \cap \mathcal{S}' = \emptyset$ and $g \prec_{N+(\mathcal{S}',f)}^T f$.*

This axiom says that, given any pair of distinct facts f, g, it is possible to add enough new claims for f to make f strictly more believable than g. Intuitively, this is less demanding that Monotonicity, which requires that f can become strictly more believable than g with the addition of just *one* additional claim. Note that Eventual Monotonicity is not possible to state in the fixed domain case (e.g. consider N already containing claims from all the available sources in $\mathcal{S}$).

When paired with Isomorphism, Eventual Monotonicity takes on a form similar to postulates for *Improvement* and *Majority* operators in belief merging [62, 64]: there is a threshold $n \in \mathbb{N}$ such that f becomes strictly more believable than g after n new claims are added for f. That is, the identities of the new sources $\mathcal{S}'$ are irrelevant; it is just the *size* of $\mathcal{S}'$ that matters. We require a preliminary lemma.

Lemma 4.6.1. *Suppose a variable domain operator T satisfies Isomorphism. Let $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ be a domain, N a network in $\mathcal{D}$ and $f \in \mathcal{F}$. Then for all non-empty, finite sets $\mathcal{S}_1', \mathcal{S}_2' \subseteq \mathbb{S}$ disjoint from $\mathcal{S}$ with $|\mathcal{S}_1'| = |\mathcal{S}_2'|$,*

$$\preceq^T_{N+(\mathcal{S}_1', f)} = \preceq^T_{N+(\mathcal{S}_2', f)}$$

Proof. Write $\mathcal{D}_1 = (\mathcal{S} \cup \mathcal{S}_1', \mathcal{F}, \mathcal{O})$ and $\mathcal{D}_2 = (\mathcal{S} \cup \mathcal{S}_2', \mathcal{F}, \mathcal{O})$. Then $N + (\mathcal{S}_i', f)$ is a network in domain $\mathcal{D}_i$ (for $i \in \{1, 2\}$). Since $|\mathcal{S}_1'| = |\mathcal{S}_2'|$ by assumption, there is a bijection $\varphi : \mathcal{S}_1' \to \mathcal{S}_2'$. Define a mapping π from $\mathcal{D}_1$ to $\mathcal{D}_2$ by

$$\pi(s) = \begin{cases} s, & s \in \mathcal{S} \\ \varphi(s), & s \in \mathcal{S}_1' \end{cases} \quad (s \in \mathcal{S} \cup \mathcal{S}_1')$$

and $\pi(g) = g$, $\pi(o) = o$ for $g \in \mathcal{F}$ and $o \in \mathcal{O}$. Then it is easily verified that π is an isomorphism from $N + (\mathcal{S}_1', f)$ to $N + (\mathcal{S}_2', f)$. For $g_1, g_2 \in \mathcal{F}$, we have $g_1 \preceq^T_{N+(\mathcal{S}_1', f)} g_2$ iff $\pi(g_1) \preceq^T_{N+(\mathcal{S}_2', f)} \pi(g_2)$ by Isomorphism. Since $\pi(g_1) = g_1$ and $\pi(g_2) = g_2$, this shows $\preceq^T_{N+(\mathcal{S}_1', f)} = \preceq^T_{N+(\mathcal{S}_2', f)}$. □

Note that since $\mathbb{S}$ is infinite and domains are finite, for any $n \in \mathbb{N}$ and any domain $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ there is always some $\mathcal{S}' \subseteq \mathbb{S}$, disjoint from $\mathcal{S}$, with $|\mathcal{S}'| = n$.

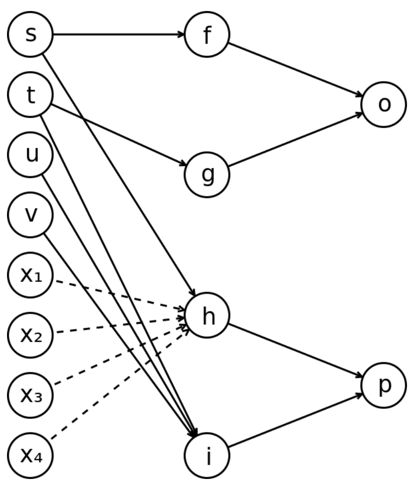


Figure 4.6: $N + (\mathcal{S}', h)$, where N is the network from Fig. 4.1 and $\mathcal{S}' = \{x_1, \ldots, x_4\}$.

For operators T satisfying Isomorphism, write $\preceq^T_{N+(n\times f)}$ for $\preceq^T_{N+(\mathcal{S}',f)}$; Lemma 4.6.1 guarantees this is well-defined (i.e. does not depend on the particular choice of $\mathcal{S}'$). That is, $\preceq^T_{N+(n\times f)}$ is the fact ranking resulting from adding n new claims for f from fresh sources. We obtain an equivalent characterisation of Eventual Monotonicity, whose proof is almost immediate given Lemma 4.6.1.

Proposition 4.6.1. *Suppose T satisfies Isomorphism. Then T satisfies Eventual Monotonicity if and only if for all domains $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, all networks N in $\mathcal{D}$ and distinct $f, g \in \mathcal{F}$, there is $n \in \mathbb{N}$ such that $g \prec^T_{N+(n\times f)} f$.*

Proof. 'if': To show Eventual Monotonicity, take any $\mathcal{S}' \subseteq \mathbb{S} \setminus \mathcal{S}$ of size n.

'Only if': Given that Eventual Monotonicity holds, simply take $n = |\mathcal{S}'|$. □

We can now show that all operators studied so far – when lifted to the variable domain case – satisfy Eventual Monotonicity.

Theorem 4.6.1. Voting, Sums, SC-Voting *and* UnboundedSums *satisfy Eventual Monotonicity.*

Proof (sketch). Let $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ be a domain, N a network in $\mathcal{D}$ and $f, g \in \mathcal{F}$. Given that Isomorphism holds for each operator, we sketch the proof via Proposition 4.6.1.

For *Voting* and *SC-Voting*, we may simply take $n = 1 + |\mathsf{src}_N(g)|$. For *Sums* and *UnboundedSums*, take $n = 2|\mathcal{S}| \cdot |\mathcal{F}|$. Write $N' = N + (\mathcal{S}', f)$ for some $\mathcal{S}' \subseteq \mathbb{S} \setminus \mathcal{S}$ with $|\mathcal{S}'| = n$

If $(T^k)_{k\in\mathbb{N}}$ denotes *Sums*, one can show by induction that $T^k_{N'}(f) = 1$ and $T^k_{N'}(h) \le \frac{1}{2}$ for any $h \neq f$ and $k > 1$, and thus $g \prec^{T^{\textit{Sums}}}_{N'} f$.

Similarly, letting $(T^k)_{k\in\mathbb{N}}$ denote *UnboundedSums*, one can show by induction that $T^k_{N'}(f) > T^k_{N'}(h)$ for $h \neq f$, and thus $g \prec^{T^{\textit{UnboundedSums}}}_{N'} g$. □

To conclude this section, we show that the impossibility result of Theorem 4.4.2 holds in the variable domain case if one replaces Monotonicity with Eventual Monotonicity and Symmetry with Isomorphism.

Theorem 4.6.2. *There is no variable domain operator satisfying Coherence, Isomorphism, Eventual Monotonicity and POI.*

Proof. For contradiction, suppose T is an operator satisfying the stated axioms. Let N be the network from Fig. 4.1, viewed as a network in domain $(\{s,t,u,v\}, \{f,g,h,i\}, \{o,p\})$. Applying Eventual Monotonicity with i and h, we have that there is N' with $i \prec^T_{N'} h$, where $N' = N + (\mathcal{S}', h)$ for some $\mathcal{S}' \subseteq \mathbb{S} \setminus \{s,t,u,v\}$. Since N' only adds claims for p-facts, POI applied to object o and Isomorphism give $f \approx^T_{N'} g$ (e.g. consider π which simply swaps s with t and f with g). From Source-Coherence we get $t \sqsubset^T_{N'} s$. But $\mathsf{src}_{N'}(f) = \{s\}$ and $\mathsf{src}_{N'}(g) = \{t\}$, so Fact-Coherence gives $g \prec^T_{N'} f$: contradiction! □

4.7 Discussion

In this section we discuss the axioms and their limitations. First, the version of Monotonicity we consider is a strict one: a new claim for f gives f a *strictly* positive boost in the fact believability ranking. This is also the case for Eventual Monotonicity in the variable domain case, where we require that some number of new claims make f strictly more believable than any other fact g. As noted in Section 4.4.3, this assumes there is no *collusion* between sources. Indeed, suppose sources s_1, s_2 are in collusion. For example, s_2 may agree to unconditionally back up all claims made by s_1. In this case a claim of f from s_1 alone should carry just as much weight as the claim from both s_1 and s_2. However, Monotonicity requires that f becomes strictly more believable when moving to the latter case.

A natural solution is to simply relax the strictness requirement. We obtain the following weak version of Monotonicity.

Axiom 4.7.1 (Weak Monotonicity)**.** *Let N, s, f, N' be as in the statement of Monotonicity. Then for all $g \neq f$, $g \preceq_N^T$ implies $g \preceq_{N'}^T f$.*

Weak Monotonicity only says says that extra support for a fact f does not make f *less* believable. Clearly Monotonicity implies Weak Monotonicity, but not vice versa. In the collusion example, an operator may select to leave the fact ranking unchanged when a new report of f from s_2 arrives; this is inconsistent with Monotonicity but consistent with Weak Monotonicity. The weak analogue of Eventual Monotonicity can be defined in the same way.

In the same spirit, one could consider versions of Coherence only using weak comparisons. Say $\mathsf{facts}_N(s_1)$ is *weakly less believable* than $\mathsf{facts}_N(s_2)$ iff the condition in Definition 4.4.1 holds, but without the requirement that some $\hat{f} \in \mathsf{facts}_N(s_1)$ is strictly less believable than its counterpart $\varphi(\hat{f})$ in $\mathsf{facts}_N(s_2)$, and define $\mathsf{src}_N(f_1)$ weakly less trustworthy than $\mathsf{src}_N(f_2)$ in a similar way. The weak analogue of Coherence is as follows.

Axiom 4.7.2 (Weak Coherence)**.** *For any network N, $\mathsf{facts}_N(s_1)$ weakly less believable than $\mathsf{facts}_N(s_2)$ implies $s_1 \sqsubseteq_N^T s_2$, and $\mathsf{src}_N(f_1)$ weakly less trustworthy than $\mathsf{src}_N(f_2)$ implies $f_1 \preceq_N^T f_2$.*

Note that Coherence does *not* imply Weak Coherence. This is because Weak Coherence relaxes both the consequent *and the antecedent* in the implications in the statement of the axiom. Whereas Coherence imposes no constraint when $\mathsf{facts}_N(s_1)$ is only weakly less believable than $\mathsf{facts}_N(s_2)$, Weak Coherence requires $s_1 \sqsubseteq_N^T s_2$. Consequently, the 'weakness' of Weak Coherence refers to its use of weak comparisons between sources and facts, not its logical strength in relation to Coherence.

A natural question now arises. Does the impossibility result of Theorem 4.4.2 still hold with these new axioms? We have an answer in the negative: the 'flat' operator, which sets all sources and facts equally ranked in all networks, satisfies all the axioms of the would-be impossibility.

Proposition 4.7.1. *Define an operator T by $s_1 \simeq_N^T s_2$ and $f \approx_N^T f_2$ for all networks N, sources s_1, s_2 and facts f_1, f_2. Then T satisfies Coherence, Weak Coherence, Symmetry, Weak Monotonicity and POI.*

Proof. Coherence holds vacuously since we can never have $\mathsf{facts}_N(s_1)$ less believable than $\mathsf{facts}_N(s_2)$ or $\mathsf{src}_N(f_1)$ less believable than $\mathsf{src}_N(f_2)$. Since *any* weak ranking holds for T, the other axioms are straightforward to see. □

This shows that (strict) Monotonicity is required for the impossibility result, since the result is no longer true when relaxing to Weak Monotonicity.

We now consider the new axioms in relation to the operators. First, Weak Coherence.

Proposition 4.7.2. Voting, Sums *and* UnboundedSums *satisfy Weak Coherence*

Proof (sketch).

Voting. Since $s_1 \sqsubseteq_N^{T^{Voting}} s_2$ always holds, Weak Source-Coherence clearly holds. For Weak Fact-Coherence, suppose $\mathsf{src}_N(f_1)$ is weakly less trustworthy than $\mathsf{src}_N(f_2)$. Then there is a bijection $\varphi : \mathsf{src}_N(f_1) \to \mathsf{src}_N(f_2)$, so $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)|$. By definition of *Voting*, $f_1 \approx_N^{T^{Voting}} f_2$. In particular, $f_1 \prec_N^{T^{Voting}} f_2$.

Sums. First, one may adapt Definition 4.5.1 to a numerical variant of a set of facts Y being *weakly* less believable than Y', by dropping all references to ρ. We then have an analogue of Lemma 4.5.1, and Weak Coherence for *Sums* follows by an argument similar to the one used to show Coherence using Lemma 4.5.1.

UnboundedSums. The proof that *UnboundedSums* satisfies Coherence can be adapted in a straightforward way to show Weak Coherence. □

Proposition 4.7.2 indicates that Weak Coherence may in fact be too weak to capture the original intuition behind Coherence – namely, that there should be a mutual dependence between trustworthy sources and believable facts – since it does not even rule out *Voting*. Instead, Weak Coherence can be seen as a simple requirement which only rules out undesirable behaviour, and complements (strict) Coherence.

Since Monotonicity implies Weak Monotonicity, it is clear that *Voting* satisfies Weak Monotonicity. We conjecture that Weak Monotonicity also holds for *Sums* and *UnboundedSums*, but this remains to be proven.[12]

4.8 Related work

In this section we discuss related work.

Ranking systems. Altman and Tennenholtz [2] initiated axiomatic study of ranking systems. First we discuss their framework in relation to ours, and then turn to their axioms. In their framework, a ranking system F maps any (finite) directed graph $G = (V, E)$ to a total preorder $\leq_G^F$ on the vertex set V. In their view this is a variation of the classical social choice setting, in which the set of voters and alternatives coincide. Nodes $v \in V$ "vote" on their peers in V by a form of approval voting [68]: an edge $v \to u$ is interpret as a vote for u from v. A ranking system then outputs a ranking of V, following the general intuition that the more "votes" v receives (i.e. the more incoming edges), the higher v should rank. As with the ranking of facts in truth discovery, this does not necessarily mean ranking nodes simply by the *number* of votes received, since the *quality* of the voters should also be taken in account. For example, a ranking system may prioritise nodes which receive few votes from highly ranked nodes over those with many votes from lower ranked nodes.

Note that our truth discovery networks N are themselves directed graphs on the vertex set $\mathcal{S} \cup \mathcal{F} \cup \mathcal{O}$. However, naively applying a ranking system to N directly makes little sense: sources never receive any "votes", and the resulting ranking includes objects, which do not need to be ranked in our setting. Perhaps a more sensible approach is to consider the bipartite graph $G_N = (V_N, E_N)$ associated with a network N, where

$$V_N = \mathcal{S} \cup \mathcal{F}, \qquad E_N = \bigcup_{(s,f) \in N} \{(s, f), (f, s)\}.$$

That is, we take the edges from sources to facts together with the reversal of such edges. The edges in G_N have some intuitive interpretation: a source votes for the facts which it claims are true, and a fact votes for the sources who vouch for it. Any ranking system F thus gives rise to a truth discovery operator, where $s_1 \sqsubseteq_N^T s_2$ iff $s_1 \leq_{G_N}^F s_2$, and similar for facts.

However, some characteristic aspects of the truth discovery problem are lost in this translation to ranking systems. Notably, the objects play no role at all in G_N.

[12] Indeed, we conjectured in Section 4.5 that the stronger axiom (strict) Monotonicity holds for *UnboundedSums*. As with Monotonicity, experimental evidence from various starting networks N suggests that Weak Monotonicity is likely to hold.

Sources and facts are also treated symmetrically, where they perhaps should not be. For example, a fact f receiving more claims than g is beneficial for f, all else being equal (see Monotonicity), but a source s claiming more facts than t does not tell us anything about the relative trustworthiness of s and t.

While other choices of G_N may be possible to alleviate some of these problems, we believe the truth discovery is sufficiently specialised beyond general graph ranking so that a bespoke modelling is required to capture its nuances appropriately. Our framework provides this novel contribution.

In [2], Altman and Tennenholtz also introduce axioms for ranking systems. Their first set of axioms deal with the transitive effects of voting when the alternatives are the voters themselves. As mentioned in Section 4.4, these axioms provided the inspiration for Coherence. The core idea is that if the predecessors of a node v are weaker than those of u, then v should be ranked below u. If v additionally has *more* predecessors, v should rank *strictly* below. Coherence applies this idea both in the direction of sources-to-facts (Fact-Coherence) and from facts-to-sources (Source-Coherence). A notable difference is that we only consider the case where the number of sources for two facts (or the number of facts, for two sources) is the same. For example, a source claiming more facts does not give it the strict boost Transitivity would dictate. Under the mapping $N \mapsto G_N$ described above, any ranking system satisfying Transitivity induces a truth discovery operator which satisfies Coherence.

The other axiom in [2] is an independence axiom RIIA (ranked independence of irrelevant alternatives), which adapts the classical IIA axiom from social choice theory to the ranking system setting, although in a different manner to our independence axioms of Section 4.4.4. We describe the axiom in rough terms, deferring to the work itself for the technical details. Suppose the relative ranking of u_1's predecessors compared to u_2's predecessors is the same as that of v_1's compared to v_2's. Then RIIA requires $u_1 \leq u_2$ iff $v_1 \leq v_2$. Informally, "the relative ranking of two agents must only depend on the pairwise comparison of the ranks of their predecessors" [2]. While we do not have an analogous axiom, the idea can be adapted to truth discovery networks. Intuitively, such an axiom would state that the ranking of two facts depends only on comparisons between their corresponding sources (and similar for the ranking of sources).

However, the main result of Altman and Tennenholtz is an impossibility: Transitivity is incompatible with RIIA. Moreover, the result remains true even when restricting to bipartite graphs, such as G_N described above. Accordingly, we can expect a similar impossibility result to hold in the truth discovery setting between Coherence and any analogue of RIIA.

PageRank. PageRank [82] is a well-known algorithm for ranking web pages based on the hyperlink structure of the web, viewed as a directed graph. It has also been studied through the lens of social choice and characterised axiomatically [3, 100].[13] In [3] the authors propose several *invariance axioms*, each of which requires that the ranking of pages is not affected by a certain transformation of the web graph. For example, the axiom *Self Edge* says that adding a self loop from a page a to itself does not change the relative ranking of other pages, and results in a strictly positive boost for a (c.f. Monotonicity). However, if we identify a truth discovery network N with the graph G_N as described above, most of the transformations involved do not respect the bipartite, symmetric structure of G_N. That is, the transformed graph does not correspond to any $G_{N'}$, for a network N'. Consequently, the PageRank axioms have no truth discovery counterpart in our setting. The only exception is *Isomorphism*, where the transformation in question is graph isomorphism; this axiom is analogous to our Symmetry and Isomorphism axioms. However, since PageRank is similar to the *Hubs and Authorities* [61] algorithm on which Sums is based – which also uses the link structure of the web to rank pages – we expect there may be additional axioms which can be expressed both for general graphs *and* truth discovery networks, satisfied by PageRank and Sums. We leave the task of finding such axioms to future work.

4.9 Summary

In this chapter we formalised a mathematical framework for truth discovery. While a number of simplifying assumptions were made compared to the mainstream truth discovery literature, we are able to express several algorithms in the framework. This provided the setting for the axiomatic method of social choice to be applied. To our knowledge, this is the first such axiomatic treatment in this context.

It was possible to adapt many axioms from social choice theory and related areas. In particular, the *Transitivity* axiom studied in the context of ranking systems [93, 2] took on new life in the form of Coherence, which we consider a core axiom for TD operators. We proceeded to establish the differences between our setting and classical social choice by considering independence axioms. This led to an impossibility result and an axiomatic characterisation of the baseline *Voting* method.

On the practical side, we analysed the existing TD algorithm *Sums* and found that, surprisingly, it fails PCI. This is a serious issue for *Sums* which has not been discussed in the literature to date, and its discovery here highlights the benefits of

[13] Ws and Skibski [100] axiomatise the *numerical scores* of PageRank, whereas Altman and Tennenholtz [3] axiomatise the resulting ranking. Moreover, Ws and Skibski point out that Altman and Tennenholtz in fact only consider a simplified version of PageRank called *Katz prestige*, defined only for strongly connected graphs.

the axiomatic method. To resolve this, we suggested a modification to *Sums* – which we call *UnboundedSums* – for which PCI *is* satisfied. However, while *UnboundedSums* resolves axiomatic problems with *Sums*, it may introduce computational difficulties (since the numeric scores involved grow without bound). We leave further investigation of such issues to future work.

A restriction of our analysis is that only one 'real-world' algorithm was considered. Further axiomatic analysis of algorithms provides a deeper understanding of how algorithms operate on an intuitive level, but is made difficult by the complexity of the state-of-the-art truth discovery methods. New techniques for establishing the satisfaction (or otherwise) of axioms would be helpful in this regard.

There is also scope for extensions to our model of truth discovery in the framework itself. For example, even in the variable domain setting of Section 4.6, we make the somewhat simplistic assumption that there are only finitely many possible facts for sources to claim. This effectively means we can only consider *categorical values*; modelling an object whose domain is the set of real numbers, for example, is not straightforward in our framework.

Next, our model does not account for any associations or constraints between objects, whereas in reality the belief in a fact for one object may strengthen or weaken our belief in other facts for related objects. These types of constraints or correlations have been studied both on the theoretical side (e.g. in judgment aggregation) and practical side in truth discovery [104].

The axioms can also be further refined to relax some of the simplifying assumptions we make regarding source attitudes; e.g. that they do not collude or attempt to manipulate. Most notably, Monotonicity should be weakened to account for such sources.

Finally, it may be argued that truth discovery as formulated in this chapter risks simply to find *consensus* among sources, rather than the *truth*. To remedy this, the framework could be extended to model the possible states of the world and thus the *ground truth* (c.f. [76]). Upon doing so one could investigate how well, and under what conditions, an operator is able to recover the truth from a TD network. Such truth-tracking methods have also been studied in judgment aggregation and belief fusion [39, 54].

5 Bipartite Tournaments

5.1 Introduction

A tournament consists of a finite set of players equipped with a *beating relation* describing pairwise comparisons between each pair of players. Determining a ranking of the players in a tournament has applications in voting in social choice [18] (where players represent alternatives and x beats y if a majority of voters prefer x over y), paired comparisons analysis [47] (where players may represent products and the beating relation the preferences of a user), search engines [92], sports tournaments [15] and other domains.

In this chapter we introduce *bipartite tournaments*, which consist of two disjoint sets of players A and B such that comparisons only take place between players from opposite sets. We consider ranking methods which produce two rankings for each tournament – one for each side of the bipartition. Such tournaments model situations in which two different kinds of entity compete *indirectly* via matches against entities of the opposite kind. The notion of competition may be abstract, which allows the model to be applied in a variety of settings. An important example is education [58], where A represents students, B exam questions, and student a 'beats' question b by answering it correctly. Here the ranking of students reflects their performance in the exam, and the ranking of questions reflects their *difficulty*. The simultaneous ranking of both sides allows one ranking to influence the other; e.g. so that students are rewarded for correctly answering difficult questions. This may prove particularly useful in the context of crowdsourced questions provided by students themselves, which may vary in their difficulty (see for example the PeerWise system [27]).

A related example is *truth discovery* [69, 88]: the task of finding true information on a number of topics when faced with conflicting reports from sources of varying (but unknown) reliability. Many truth discovery algorithms operate iteratively, alternately estimating the reliability of sources based on current estimates of the true information, and obtaining new estimates of the truth based on source reliability levels. The former is an instance of a bipartite tournament; similar to the education example, A represents data sources, B topics of interest, and a defeats b by providing true information on topic b (according to the current estimates of the truth). Applying a bipartite tournament ranking method at this step may therefore facilitate development of *difficulty-aware* truth discovery algorithms, which reward sources for providing accurate information on difficult topics [43]. Other application domains include the evaluation of generative models in machine learning [79] (where A represents generators and B discriminators) and solo sports contests (e.g. where A represents golfers and B golf courses).

In principle, bipartite tournaments are a special case of *generalised* tournaments [47, 91, 23], which allow intensities of victories and losses beyond a binary win or loss (thus permitting draws or multiple comparisons), and drop the requirement that every player is compared to all others. However, many existing ranking methods in the literature do not apply to bipartite tournaments due to the violation of an *irreducibility* requirement, which requires that the tournament graph be strongly connected. In any case, bipartite tournament ranking presents a unique problem – since we aim to rank players with only indirect information available – which we believe is worthy of study in its own right.

In this work we focus particularly on ranking via *chain graphs* and *chain editing*. A chain graph is a bipartite graph in which the neighbourhoods of vertices on one side form a chain with respect to set inclusion. A (bipartite) tournament of this form represents an 'ideal' situation in which the capabilities of the players are perfectly nested: weaker players defeat a subset of the opponents that stronger players defeat. In this case a natural ranking can be formed according to the set of opponents defeated by each player. These rankings respect the tournament results in an intuitive sense: if a player a defeats b and b' ranks worse than b, then a must defeat b' also. Unfortunately, this perfect nesting may not hold in reality: a weak player may win a difficult match by coincidence, and a strong player may lose a match by accident. With this in mind, Jiao, Ravi, and Gatterbauer [58] suggested an appealing ranking method for bipartite tournaments: apply *chain editing* to the input tournament – i.e. find the minimum number of edge changes required to form a chain graph – and output the corresponding rankings. Whilst their work focused on algorithms for chain editing and its variants, we look to study the properties of the ranking method itself through the lens of computational social choice.

Contribution. Our primary contribution is the introduction of a class of ranking mechanisms for bipartite tournaments defined by chain editing. We also provide a new probabilistic characterisation of chain editing via maximum likelihood estimation. To our knowledge this is the first in-depth study of chain editing as a ranking mechanism. Secondly, we introduce a new class of 'chain-definable' mechanisms by relaxing the minimisation constraint of chain editing in order to obtain tractable algorithms and to resolve the failure of an important anonymity axiom. We present a concrete example of such an algorithm, and characterise it axiomatically.

Chapter outline. In Section 5.2 we define the framework for bipartite tournaments and introduce chain graphs. Section 5.3 outlines how one may use chain editing to rank a tournament, and characterises the resulting mechanisms in a probabilistic setting. Axiomatic properties are considered in Section 5.4. Section 5.5 defines a concrete scheme for producing chain-editing-based rankings. Section 5.6 introduces new ranking methods by relaxing the chain editing requirement. Related work is discussed in Section 5.7, and we conclude in Section 5.8.

5.2 Preliminaries

In this section we define our framework for bipartite tournaments, introduce chain graphs and discuss the link between them.

5.2.1 Bipartite Tournaments

Following the literature on generalised tournaments [47, 91, 23], we represent a tournament as a matrix, whose entries represent the results of matches between participants. In what follows, $[n]$ denotes the set $\{1, \ldots, n\}$ whenever $n \in \mathbb{N}$.

Definition 5.2.1. *A* bipartite tournament *– hereafter simply a* tournament *– is a triple (A, B, K), where $A = [m]$ and $B = [n]$ for some $m, n \in \mathbb{N}$, and K is an $m \times n$ matrix with $K_{ab} \in \{0, 1\}$ for all $(a, b) \in A \times B$. The set of all tournaments will be denoted by $\mathcal{K}$.*

Here A and B represent the two sets of players in the tournament.[1] An entry K_{ab} gives the result of the match between $a \in A$ and $b \in B$: it is 1 if a defeats b and 0 otherwise. Note that we do not allow for the possibility of draws, and every $a \in A$ faces every $b \in B$. When there is no ambiguity we denote a tournament simply by K, with the understanding that $A = [\text{rows}(K)]$ and $B = [\text{columns}(K)]$.

The *neighbourhood* of a player $a \in A$ in K is the set $K(a) = \{b \in B \mid K_{ab} = 1\} \subseteq B$, i.e. the set of players which a defeats. The neighbourhood of $b \in B$ is the set $K^{-1}(b) = \{a \in A \mid K_{ab} = 1\} \subseteq A$, i.e. the set of players defeating b.

Given a tournament K, our goal is to place a ranking on each of A and B. We define a ranking *operator* for this purpose.

[1] Note that A and B are not disjoint as sets: 1 is always contained in both A and B, for instance. This poses no real problem, however, since we view the number 1 merely a *label* for a player. It will always be clear from context whether a given integer should be taken as a label for a player on the A side or the B side.

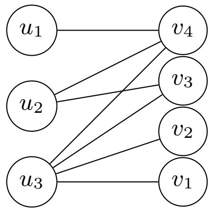


Figure 5.1: An example of a chain graph

Definition 5.2.2. *An* operator φ *assigns each tournament* K *a pair* $\varphi(K) = (\preceq_K^\varphi, \sqsubseteq_K^\varphi)$ *of total preorders on* A *and* B *respectively.*[2]

For $a, a' \in A$, we interpret $a \preceq_K^\varphi a'$ to mean that a' is ranked *at least as strong* as a in the tournament K, according to the operator φ (similarly, $b \sqsubseteq_K^\varphi b'$ means b' is ranked at least as strong as b). The strict and symmetric parts of $\preceq_K^\varphi$ are denoted by $\prec_K^\varphi$ and $\approx_K^\varphi$.

As a simple example, consider φ_{count}, where $a \preceq_K^{\varphi_{\mathsf{count}}} a'$ iff $|K(a)| \leq |K(a')|$ and $b \sqsubseteq_K^{\varphi_{\mathsf{count}}} b'$ iff $|K^{-1}(b)| \geq |K^{-1}(b')|$. This operator simply ranks players by number of victories. It is a bipartite version of the *points system* introduced by Rubinstein [85], and generalises *Copeland's rule* [18].

5.2.2 Chain Graphs

Each bipartite tournament K naturally corresponds to a bipartite graph G_K, with vertices $A \sqcup B$ and an edge between a and b whenever $K_{ab} = 1$.[3] The task of ranking a tournament admits a particularly simple solution if this graph happens to be a *chain graph*.

Definition 5.2.3 ([106])**.** *A bipartite graph* $G = (U, V, E)$ *is a* chain graph *if there is an ordering* $U = \{u_1, \ldots, u_k\}$ *of* U *such that* $N(u_1) \subseteq \cdots \subseteq N(u_k)$*, where* $N(u_i) = \{v \in V \mid (u_i, v) \in E\}$ *is the neighbourhood of* u_i *in* G*.*

In other words, a chain graph is a bipartite graph where the neighbourhoods of the vertices on one side can be ordered so as to form a chain with respect to set inclusion. It is easily seen that this nesting property holds for U if and only if it holds for V. Figure 5.1 shows an example of a chain graph.

Now, as our terminology might suggest, the neighbourhood $K(a)$ of some player $a \in A$ in a tournament K coincides with the neighbourhood of the corresponding vertex in G_K. If G_K is a chain graph we can therefore enumerate A as $\{a_1, \ldots, a_m\}$ such that $K(a_i) \subseteq K(a_{i+1})$ for each $1 \leq i < m$. This indicates that each a_{i+1} has performed *at least as well* as a_i in a strong sense: every opponent which a_i defeated

[2] A total preorder is a transitive and complete binary relation.

[3] $A \sqcup B$ is the *disjoint union* of A and B, which we define as $\{(a, \mathcal{A}) \mid a \in A\} \cup \{(b, \mathcal{B}) \mid b \in B\}$, where $\mathcal{A}$ and $\mathcal{B}$ are constant symbols.

was also defeated by a_{i+1}, and a_{i+1} may have additionally defeated opponents which a_i did not.[4] It seems only natural in this case that one should rank a_i (weakly) below a_{i+1}. Appealing to transitivity and the fact that each $a \in A$ appears as *some* a_i, we see that any tournament K where G_K is a chain graph comes pre-equipped with a natural total preorder on A, where a' ranks higher than than a if and only if $K(a) \subseteq K(a')$. The duality of the neighbourhood-nesting property for chain graphs implies that B can also be totally preordered, with b' ranked higher than b if and only if $K^{-1}(b) \supseteq K^{-1}(b')$.[5] Moreover, these total preorders relate to the tournament results in an important sense: if a defeats b and b' ranks worse than b, then a must defeat b' also. That is, the neighbourhood of each $a \in A$ is *downwards closed* w.r.t the ranking of B, and the neighbourhood of each $b \in B$ is *upwards closed* in A.

Tournaments corresponding to chain graphs will be said to satisfy the *chain property*, and will accordingly be called *chain tournaments*. We give a simpler (but equivalent) definition which does not refer to the underlying graph G_K. First, define relations $\leqslant_K^{\mathcal{A}}, \leqslant_K^{\mathcal{B}}$ on A and B respectively by $a \leqslant_K^{\mathcal{A}} a'$ iff $K(a) \subseteq K(a')$ and $b \leqslant_K^{\mathcal{B}} b'$ iff $K^{-1}(b) \supseteq K^{-1}(b')$, for any tournament K.

Definition 5.2.4. *A tournament K has the* chain property *if $\leqslant_K^{\mathcal{A}}$ is a total preorder.*

According to the duality principle mentioned already, the chain property implies that $\leqslant_K^{\mathcal{B}}$ is also a total preorder. Note that the relations $\leqslant_K^{\mathcal{A}}$ and $\leqslant_K^{\mathcal{B}}$ are analogues of the *covering relation* for non-bipartite tournaments [18].

Example 5.2.1. *Consider $K = \begin{bmatrix} 1&0&0&0 \\ 1&1&0&0 \\ 1&1&1&1 \end{bmatrix}$. Then $K(1) \subset K(2) \subset K(3)$, so K has the chain property. In fact, K is the tournament corresponding to the chain graph G from Figure 5.1.*

5.3 Ranking via Chain Editing

We have seen that chain tournaments come equipped with natural rankings of A and B. Such tournaments represent an 'ideal' situation, wherein the abilities of the players on both sides of the tournament are perfectly nested. Of course this may not be so in reality: the nesting may be broken by some $a \in A$ winning a match it ought not to by chance, or by losing a match by accident.

One idea for recovering a ranking in this case, originally suggested by Jiao, Ravi, and Gatterbauer [58], is to apply *chain editing*: find the minimum number of edge changes required to convert the graph G_K into a chain graph. This process can be seen as correcting the 'noise' in an observed tournament K to obtain an ideal

[4] Note that this is a more robust notion of performance than comparing the neighbourhoods of a_i and a_{i+1} by *cardinality*, which may fail to account for differences in the strength of opponents when counting wins and losses.

[5] Note that the ordering of the Bs is reversed compared to the As, since the larger $K^{-1}(b)$ the *worse* b has performed.

ranking. In this section we introduce the class of operators producing rankings in this way.

5.3.1 Chain-minimal Operators

To define chain-editing in our framework we once again present an equivalent definition which does not refer to the underlying graph G_K: the number of edge changes between graphs can be replaced by the *Hamming distance* between tournament matrices.

Definition 5.3.1. *For $m, n \in \mathbb{N}$, let $\mathcal{C}_{m,n}$ denote the set of all $m \times n$ chain tournaments. For an $m \times n$ tournament K, write $\mathcal{M}(K) = \arg\min_{K' \in \mathcal{C}_{m,n}} d(K, K') \subseteq \mathcal{K}$ for the set of chain tournaments closest to K w.r.t the Hamming distance $d(K, K') = |\{(a, b) \in A \times B \mid K_{ab} \neq K'_{ab}\}|$. Let $m(K)$ denote this minimum distance.*

Note that chain editing, which is NP-hard in general [58], amounts to finding a single element of $\mathcal{M}(K)$.[6] We comment further on the computational complexity of chain editing in Section 5.7. The following property characterises chain editing-based operators φ.

Axiom 5.3.1 (chain-min)**.** *For every tournament K there is $K' \in \mathcal{M}(K)$ such that $\varphi(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$.*

That is, the ranking of K is obtained by choosing the neighbourhood-subset rankings for some closest chain tournament K'. Operators satisfying **chain-min** will be called *chain-minimal*.

Example 5.3.1. *Consider $K = \begin{bmatrix} 1&0&1&0 \\ 1&1&0&0 \\ 0&1&1&1 \end{bmatrix}$. K does not have the chain property, since neither $K(1) \subseteq K(2)$ nor $K(2) \subseteq K(1)$. The set $\mathcal{M}(K)$ consists of four tournaments a distance of 2 from K:*

$$\mathcal{M}(K) = \left\{ \begin{bmatrix} 1&\mathbf{1}&1&0 \\ 1&1&0&0 \\ \mathbf{1}&1&1&1 \end{bmatrix}, \begin{bmatrix} 1&0&\mathbf{0}&0 \\ 1&1&0&0 \\ \mathbf{1}&1&1&1 \end{bmatrix}, \begin{bmatrix} 1&0&1&0 \\ 1&\mathbf{0}&0&0 \\ \mathbf{1}&1&1&1 \end{bmatrix}, \begin{bmatrix} 1&0&1&0 \\ 1&1&\mathbf{1}&0 \\ \mathbf{1}&1&1&1 \end{bmatrix} \right\}$$

The corresponding rankings are $(213, \{12\}34)$, $(123, 12\{34\})$, $(213, 13\{24\})$ and $(123, \{13\}24)$.[7]

Example 5.3.1 shows that there is no unique chain-minimal operator, since for a given tournament K there may be several closest chain tournaments to choose from. In Section 5.5 we introduce a principled way to single out a *unique* chain tournament and thereby construct a well-defined chain-minimal operator.

[6] The decision problem associated with chain editing – which in tournament terms is the question of whether $m(K) \leq k$ for a given integer k – is NP-complete [33].

[7] Here $a_1 a_2 a_3$ is shorthand for the ranking $a_1 \prec a_2 \prec a_3$ of A, and similar for B. Elements in brackets are ranked equally.

5.3.2 A Maximum Likelihood Interpretation

So far we have motivated **chain-min** as a way to fix errors in a tournament and recover the ideal or *true* ranking. In this section we make this notion precise by defining a probabilistic model in which chain-minimal rankings arise as maximum likelihood estimates. The maximum likelihood approach has been applied for (non-bipartite) tournaments (e.g. the Bradley-Terry model [16, 47]), voting in social choice theory [35], truth discovery [98], belief merging [38] and other related problems.

In this approach we take an epistemic view of tournament ranking: it is assumed there exists a true 'state of the world' which determines the tournament results along with objective rankings of A and B. A given tournament K is then seen as a *noisy observation* derived from the true state, and a *maximum likelihood estimate* is a state for which the probability of observing K is maximal.

More specifically, a state of the world is represented as a vector of *skill levels* for the players in A and B.[8]

Definition 5.3.2. *For a fixed size $m \times n$, a* state of the world *is a tuple $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, where $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{y} \in \mathbb{R}^n$ satisfies the following properties:*

$$\forall a, a' \in A \quad (x_a < x_{a'} \implies \exists b \in B : x_a < y_b \leq x_{a'}) \tag{5.1}$$

$$\forall b, b' \in B \quad (y_b < y_{b'} \implies \exists a \in A : y_b \leq x_a < y_{b'}) \tag{5.2}$$

where $A = [m]$, $B = [n]$. Write $\Theta_{m,n}$ for the set of all $m \times n$ states.

For $a \in A$, x_a is the *skill level* of a in state θ (and similarly for y_b). These skill levels represent the true capabilities of the players in A and B in state θ: a is capable of defeating b if and only if $x_a \geq y_b$. Note that (5.1) suggests a simple form of *explainability*: a' can only be strictly more skilful than a if there is some $b \in B$ which *explains* this fact, i.e. some b which a' can defeat but a cannot ((5.2) is analogous for the Bs). These conditions are intuitive if we assume that skill levels are relative to the sets A and B currently under consideration (i.e. they do not reflect the abilities of players in future matches against new contenders outside of A or B). Finally note that our states of the world are *richer* than the output of an operator, in contrast to other work in the literature [16, 47, 35]. Specifically, a state θ contains extra information in the form of comparisons between A and B.

Noise is introduced in the observed tournament K via *false positives* (where $a \in A$ defeats a more skilled $b \in B$ by accident) and *false negatives* (where $a \in A$ is defeated by an inferior $b \in B$ by mistake).[9] The noise model is therefore parametrised by the

[8] For simplicity we use numerical skill levels here, although it would suffice to have a partial preorder on $A \sqcup B$ such that each $a \in A$ is comparable with every $b \in B$.

[9] Note that a false positive for a is a false negative for b and vice versa.

false positive and false negative rates $\boldsymbol{\alpha} = \langle \alpha_+, \alpha_- \rangle \in [0,1]^2$, which we assume are the same for all $a \in A$.[10] We also assume that noise occurs independently across all matches.

Definition 5.3.3. *Let $\boldsymbol{\alpha} = \langle \alpha_+, \alpha_- \rangle \in [0,1]^2$. For each $m, n \in \mathbb{N}$ and $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \Theta_{m,n}$, consider independent binary random variables X_{ab} representing the outcome of a match between $a \in [m]$ and $b \in [n]$, where*

$$P_{\boldsymbol{\alpha}}(X_{ab} = 1 \mid \theta) = \begin{cases} \alpha_+, & x_a < y_b \\ 1 - \alpha_-, & x_a \geq y_b \end{cases} \tag{5.3}$$

$$P_{\boldsymbol{\alpha}}(X_{ab} = 0 \mid \theta) = \begin{cases} 1 - \alpha_+, & x_a < y_b \\ \alpha_-, & x_a \geq y_b \end{cases} \tag{5.4}$$

This defines a probability distribution $P_{\boldsymbol{\alpha}}(\cdot \mid \theta)$ over $m \times n$ tournaments by

$$P_{\boldsymbol{\alpha}}(K \mid \theta) = \prod_{(a,b) \in [m] \times [n]} P_{\boldsymbol{\alpha}}(X_{ab} = K_{ab} \mid \theta)$$

Here $P_{\boldsymbol{\alpha}}(K \mid \theta)$ is the probability of observing the tournament results K when the false positive and negative rates are given by $\boldsymbol{\alpha}$ and the true state of the world is θ. Note that the four cases in (5.3) and (5.4) correspond to a false positive, true positive, true negative and false negative respectively. We can now define a maximum likelihood operator.

Definition 5.3.4. *Let $\boldsymbol{\alpha} \in [0,1]^2$ and $m, n \in \mathbb{N}$. Then $\theta \in \Theta_{m,n}$ is a* maximum likelihood estimate *(MLE) for an $m \times n$ tournament K w.r.t $\boldsymbol{\alpha}$ if $\theta \in \arg\max_{\theta' \in \Theta_{m,n}} P_{\boldsymbol{\alpha}}(K \mid \theta')$. An operator φ is a* maximum likelihood operator *w.r.t $\boldsymbol{\alpha}$ if for any $m, n \in \mathbb{N}$ and any $m \times n$ tournament K there is an MLE $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \Theta_{m,n}$ for K such that $a \preceq_K^{\varphi} a'$ iff $x_a \leq x_{a'}$ and $b \sqsubseteq_K^{\varphi} b'$ iff $y_b \leq y_{b'}$.*

To help analyse MLE operators, we consider the tournament K_θ associated with each state $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, given by $[K_\theta]_{ab} = 1$ if $x_a \geq y_b$ and $[K_\theta]_{ab} = 0$ otherwise. Note that K_θ is the unique tournament with non-zero probability when there are no false positive or false negatives. The following technical lemma obtains an expression for $P_{\boldsymbol{\alpha}}(K \mid \theta)$ in terms of K_θ and K.

Lemma 5.3.1. *Let K be an $m \times n$ tournament, $\boldsymbol{\alpha} \in [0,1]^2$ and $\theta \in \Theta_{m,n}$. Then*

$$\begin{aligned} P_{\boldsymbol{\alpha}}(K \mid \theta) = \prod_{a \in A} & \alpha_+^{|K(a) \setminus K_\theta(a)|} (1 - \alpha_-)^{|K(a) \cap K_\theta(a)|} \\ & (1 - \alpha_+)^{|B \setminus (K(a) \cup K_\theta(a))|} \alpha_-^{|K_\theta(a) \setminus K(a)|}. \end{aligned}$$

[10] This is a strong assumption, and it may be more realistic to model the false positive/negative rates as a function of x_a. We leave this to future work.

Proof. Write $p_{ab,K}$ for $P_{\boldsymbol{\alpha}}(X_{ab} = K_{ab} \mid \theta)$. Expanding the product in Definition 5.3.3, we have

$$P_{\boldsymbol{\alpha}}(K \mid \theta) = \prod_{a \in A} \prod_{b \in B} p_{ab,K}.$$

Let $a \in A$. Note that B can be written as the disjoint union $B = B_1 \cup B_2 \cup B_3 \cup B_4$, where

$$\begin{aligned}
B_1 &= K(a) \setminus K_\theta(a) \\
B_2 &= K(a) \cap K_\theta(a) \\
B_3 &= B \setminus (K(a) \cup K_\theta(a)) \\
B_4 &= K_\theta(a) \setminus K(a).
\end{aligned}$$

Recall that $b \in K_\theta(a)$ iff $x_a \geq y_b$ (where $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$). It follows that

- $b \in B_1$ iff $K_{ab} = 1$ and $x_a < y_b$
- $b \in B_2$ iff $K_{ab} = 1$ and $x_a \geq y_b$
- $b \in B_3$ iff $K_{ab} = 0$ and $x_a < y_b$
- $b \in B_4$ iff $K_{ab} = 0$ and $x_a \geq y_b$

Note that this correspond exactly to the four cases in (5.3) and (5.4) which define $p_{ab,K}$; we have

$$p_{ab,K} = \begin{cases} \alpha_+, & b \in B_1 \\ 1 - \alpha_-, & b \in B_2 \\ 1 - \alpha_+, & b \in B_3 \\ \alpha_-, & b \in B_4. \end{cases}$$

Consequently

$$\begin{aligned}
\prod_{b \in B} p_{ab,K} &= \left(\prod_{b \in B_1} \alpha_+ \right) \left(\prod_{b \in B_2} (1 - \alpha_-) \right) \left(\prod_{b \in B_3} (1 - \alpha_+) \right) \left(\prod_{b \in B_4} \alpha_- \right) \\
&= \alpha_+^{|B_1|} (1 - \alpha_-)^{|B_2|} (1 - \alpha_+)^{|B_3|} \alpha_-^{|B_4|} \\
&= \alpha_+^{|K(a) \setminus K_\theta(a)|} (1 - \alpha_-)^{|K(a) \cap K_\theta(a)|} \\
&\qquad (1 - \alpha_+)^{|B \setminus (K(a) \cup K_\theta(a))|} \alpha_-^{|K_\theta(a) \setminus K(a)|}.
\end{aligned}$$

Taking the product over all $a \in A$ we reach the desired expression for $P_{\boldsymbol{\alpha}}(K \mid \theta)$. □

Expressed in terms of K_θ, the MLEs take a particularly simple form if $\alpha_+ = \alpha_-$, i.e. if false positives and false negatives occur at the same rate.

Lemma 5.3.2. *Let $\boldsymbol{\alpha} = \langle \beta, \beta \rangle$ for some $\beta < \frac{1}{2}$. Then θ is an MLE for K if and only if $\theta \in \arg\min_{\theta' \in \Theta_{m,n}} d(K, K_{\theta'})$.*

Proof. Let K be an $m \times n$ tournament. By Lemma 5.3.1,

$$P_{\boldsymbol{\alpha}}(K \mid \theta) = \Big(\prod_{a \in A} \alpha_+^{|K(a) \setminus K_\theta(a)|} (1-\alpha_-)^{|K(a) \cap K_\theta(a)|} \\ (1-\alpha_+)^{|B \setminus (K(a) \cup K_\theta(a))|} \alpha_-^{|K_\theta(a) \setminus K(a)|} \Big).$$

Plugging in $\alpha_+ = \alpha_- = \beta$ and simplifying, one can obtain

$$P_{\boldsymbol{\alpha}}(K \mid \theta) = c \prod_{a \in A} \left(\frac{\beta}{1-\beta} \right)^{|K(a) \triangle K_\theta(a)|},$$

where $X \triangle Y = (X \setminus Y) \cup (Y \setminus X)$ is the symmetric difference of two sets X and Y, and $c = (1-\beta)^{|A| \cdot |B|}$ is a positive constant that does not depend on θ. Now, $P_{\boldsymbol{\alpha}}(K \mid \theta)$ is positive, and is maximal when its logarithm is. We have

$$\begin{aligned} \log P_{\boldsymbol{\alpha}}(K \mid \theta) &= \log c + \log \left(\frac{\beta}{1-\beta} \right) \sum_{a \in A} |K(a) \triangle K_\theta(a)| \\ &= \log c + \log \left(\frac{\beta}{1-\beta} \right) d(K, K_\theta) \end{aligned}.$$

Since $\log c$ is constant and $\beta < 1/2$ implies $\log \left(\frac{\beta}{1-\beta} \right) < 0$, it follows that $\log P_{\boldsymbol{\alpha}}(K \mid \theta)$ is maximised exactly when $d(K, K_\theta)$ is minimised, which proves the result. □

This result characterises the MLE states for K as those for which K_θ is the closest to K. As it turns out, the tournaments K_θ that arise in this way are exactly those with the chain property.

Lemma 5.3.3. *Let $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \Theta_{m,n}$. Then for all $a, a' \in A$ and $b, b' \in B$:*

1. $K_\theta(a) \subseteq K_\theta(a')$ *iff* $x_a \leq x_{a'}$
2. $K_\theta^{-1}(b) \supseteq K_\theta^{-1}(b')$ *iff* $y_b \leq y_{b'}$.

Proof. We prove (1); (2) is shown similarly. Let $a, a' \in A$. First suppose $x_a \leq x_{a'}$. Let $b \in K_\theta(a)$. Then $y_b \leq x_a \leq x_{a'}$, so $b \in K_\theta(a')$ also. This shows $K_\theta(a) \subseteq K_\theta(a')$.

Now suppose $K_\theta(a) \subseteq K_\theta(a')$. For the sake of contradiction, suppose $x_a > x_{a'}$. By (5.1) in the definition of a state (Definition 5.3.2), there is $b \in B$ such that $x_{a'} < y_b \leq x_a$. But this means $b \in K_\theta(a) \setminus K_\theta(a')$, which contradicts $K_\theta(a) \subseteq K_\theta(a')$. Thus (1) is proved. □

Lemma 5.3.4. *An $m \times n$ tournament K has the chain property if and only if $K = K_\theta$ for some $\theta \in \Theta_{m,n}$.*

Proof. The "if" direction follows from Lemma 5.3.3 part (1): if $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ and $a, a' \in A$ then either $x_a \leq x_{a'}$ – in which case $K_\theta(a) \subseteq K_\theta(a')$ – or $x_{a'} < x_a$ – in which case $K_\theta(a') \subseteq K_\theta(a)$. Therefore K_θ has the chain property.

For the "only if" direction, suppose K has the chain property. Define $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ by

$$x_a = |\{a' \in A \mid K(a') \subseteq K(a)\}|$$
$$y_b = \begin{cases} \min\{x_a \mid a \in K^{-1}(b)\}, & K^{-1}(b) \neq \emptyset \\ 1 + |A|, & K^{-1}(b) = \emptyset \end{cases}$$

It is easily that since the neighbourhood-subset relation $\leqslant_K^A$ is a total preorder, we have $K(a) \subseteq K(a')$ if and only if $x_a \leq x_{a'}$. First we show that $K_\theta = K$ by showing that $K_{ab} = 1$ if and only if $[K_\theta]_{ab} = 1$. Suppose $K_{ab} = 1$. Then $a \in K^{-1}(b)$, so $y_b = \min\{x_{a'} \mid a' \in K^{-1}(b)\} \leq x_a$ and consequently $[K_\theta]_{ab} = 1$.

Now suppose $[K_\theta]_{ab} = 1$. Then $x_a \geq y_b$. We must have $K^{-1}(b) \neq \emptyset$; otherwise $y_b = 1 + |A| > |A| \geq x_a$. We can therefore take $\hat{a} \in \arg\min_{a' \in K^{-1}(b)} x_{a'}$. By definition of y_b, $x_{\hat{a}} = y_b \leq x_a$. But $x_{\hat{a}} \leq x_a$ implies $K(\hat{a}) \subseteq K(a)$; since $\hat{a} \in K^{-1}(b)$ this gives $b \in K(\hat{a})$ and $b \in K(a)$, i.e. $K_{ab} = 1$. This completes the claim that $K = K_\theta$.

It only remains to show that θ satisfies conditions (5.1) and (5.2) of Definition 5.3.2. For (5.1), suppose $x_a < x_{a'}$. Then $K(a) \subset K(a')$, i.e there is $b \in K(a') \setminus K(a) = K_\theta(a') \setminus K_\theta(a)$. But $b \in K_\theta(a')$ gives $y_b \leq x_{a'}$, and $b \notin K_\theta(a)$ gives $x_a < y_b$; this shows that (5.1) holds.

For (5.2), suppose $y_b < y_{b'}$. Clearly $K^{-1}(b) \neq \emptyset$ (otherwise $y_b = 1 + |A|$ is maximal). Thus there is $a \in K^{-1}(b)$ such that $y_b = x_a$. This of course means $x_a < y_{b'}$; in particular we have $y_b \leq x_a < y_{b'}$ as required for (5.2).

We have shown that $K = K_\theta$ and that $\theta \in \Theta_{m,n}$, and the proof is complete. □

Note that the proof of Lemma 5.3.3 relies crucially on (5.1) and (5.2) in the definition of a state. Combining all the results so far we obtain our first main result: the maximum likelihood operators for $\boldsymbol{\alpha} = \langle \beta, \beta \rangle$ are exactly the chain-minimal operators.

Theorem 5.3.1. *Let $\boldsymbol{\alpha} = \langle \beta, \beta \rangle$ for some $\beta < \frac{1}{2}$. Then φ is a maximum likelihood operator w.r.t $\boldsymbol{\alpha}$ if and only if φ satisfies **chain-min**.*

Proof. First we show that for any $m, n \in \mathbb{N}$ and any $m \times n$ tournament K it holds that θ is an MLE state for K if and only if $K_\theta \in \mathcal{M}(K)$.

Indeed, fix some m, n and K. Write $\mathcal{K}_{\Theta_{m,n}} = \{K_\theta \mid \theta \in \Theta_{m,n}\}$. By Lemma 5.3.2, θ is an MLE if and only if $d(K, K_\theta) \leq d(K, K_{\theta'})$ for all $\theta' \in \Theta_{m,n}$, i.e. $K_\theta \in \arg\min_{K' \in \mathcal{K}_{\Theta_{m,n}}} d(K, K')$. But by Lemma 5.3.4, $\mathcal{K}_{\Theta_{m,n}}$ is just $\mathcal{C}_{m,n}$, the set of all

$m \times n$ tournaments with the chain property. We see that $\arg\min_{K' \in \mathcal{K}_{\Theta_{m,n}}} d(K, K') = \arg\min_{K' \in \mathcal{C}_{m,n}} d(K, K') = \mathcal{M}(K)$ by definition of $\mathcal{M}(K)$. This shows that θ is an MLE iff $K_\theta \in \mathcal{M}(K)$.

Now, by definition, φ satisfies **chain-min** iff for every tournament K there is $K' \in \mathcal{M}(K)$ such that $\varphi(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$. Using Lemma 5.3.4 and the above result, $K' \in \mathcal{M}(K)$ if and only if $K' = K_\theta$ for some MLE θ for K. We see that **chain-min** can be equivalently stated as follows: for all tournament K there exists an MLE θ such that $\varphi(K) = (\leqslant^{\mathcal{A}}_{K_\theta}, \leqslant^{\mathcal{B}}_{K_\theta})$. But by Lemma 5.3.3 we have $a \leqslant^{\mathcal{A}}_{K_\theta} a'$ iff $x_a \leq x_{a'}$ and $b \leqslant^{\mathcal{B}}_{K_\theta} b'$ iff $y_b \leq y_{b'}$ (where $\theta = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$). The above reformulation of **chain-min** now coincides with the definition of a maximum likelihood operator, and we are done. $\square$

Similar results can be obtained for other limiting values of $\boldsymbol{\alpha}$. If $\alpha_+ = 0$ and $\alpha_- \in (0, 1)$ then the MLE operators correspond to *chain completion*: finding the minimum number of edge *additions* required to make G_K a chain graph. This models situations where false positives never occur, although false negatives may (e.g. numerical entry questions in the case where A represents students and B exam questions [58]). Similarly, the case $\alpha_- = 0$ and $\alpha_+ \in (0, 1)$ corresponds to *chain deletion*, where edge additions are not allowed.

5.4 Axiomatic analysis

Chain-minimal operators have theoretical backing in a probabilistic sense due to the results of Section 5.3.2, but are they appropriate ranking methods in practise? To address this question we consider the *normative* properties of chain-minimal operators via the axiomatic method of social choice theory. We formulate several axioms for bipartite tournament ranking and assess whether they are compatible with **chain-min**. It will be seen that an important *anonymity* axiom fails for all chain-minimal operators; later in Section 5.5 we describe a scenario in which this is acceptable and define a class of concrete operators for this case, and in Section 5.6 we relax the **chain-min** requirement in order to gain anonymity.

5.4.1 The Axioms

We will consider five axioms – mainly adaptations of standard social choice properties to the bipartite tournament setting.

Symmetry Properties. We consider two symmetry properties. The first is a classic *anonymity* axiom, which says that an operator φ should not be sensitive to the 'labels' used to identify participants in a tournament. Axioms of this form are standard in social choice theory; a tournament version goes at least as far back as [85].

We need some notation: for a tournament K and permutations $\sigma : A \to A$, $\pi : B \to B$, let $\sigma(K)$ and $\pi(K)$ denote the tournament obtained by permuting the rows and columns of K by σ and π respectively, i.e. $[\sigma(K)]_{ab} = K_{\sigma^{-1}(a),b}$ and $[\pi(K)]_{ab} = K_{a,\pi^{-1}(b)}$. Note that in the statement of the axioms we omit universal quantification over K, $a, a' \in A$ and $b, b' \in B$ for brevity.

Axiom 5.4.1 (anon). *Let $\sigma : A \to A$ and $\pi : B \to B$ be permutations. Then $a \preceq^{\varphi}_{K} a'$ iff $\sigma(a) \preceq^{\varphi}_{\pi(\sigma(K))} \sigma(a')$.*

Our second axiom is specific to bipartite tournaments, and expresses a *duality* between the two sides A and B: given the two sets of conceptually disjoint entities participating in a bipartite tournament, it should not matter which one we label A and which one we label B. We need the notion of a *dual tournament*.

Definition 5.4.1. *The* dual tournament *of K is $\overline{K} = \mathbf{1} - K^{\top}$, where $\mathbf{1}$ denotes the matrix consisting entirely of 1s.*

$\overline{K}$ is essentially the same tournament as K, but with the roles of A and B swapped. In particular, $A_K = B_{\overline{K}}$, $B_K = A_{\overline{K}}$ and $K_{ab} = 1$ iff $\overline{K}_{ba} = 0$. Also note that $\overline{\overline{K}} = K$. The duality axiom states that the ranking of the Bs in K is the same as the As in $\overline{K}$.

Axiom 5.4.2 (dual). *$b \sqsubseteq^{\varphi}_{K} b'$ iff $b \preceq^{\varphi}_{\overline{K}} b'$.*

Whilst **dual** is not necessarily a universally desirable property – one can imagine situations where A and B are not fully abstract and should not be treated symmetrically – it is important to consider in any study of bipartite tournaments. Note that **dual** implies $a \preceq^{\varphi}_{K} a'$ iff $a \sqsubseteq^{\varphi}_{\overline{K}} a'$, so that a **dual**-operator can be defined by giving the ranking for one of A or B only, and defining the other by duality. This explains our choice to define **anon** (and subsequent axioms) solely in terms of the A ranking: the analogous anonymity constraint for the B ranking follows from **anon** together with **dual**.

An Independence Property. *Independence axioms* play a crucial role in social choice. We present a bipartite adaptation of a classic axiom introduced in [85], which has subsequently been called *Independence of Irrelevant Matches* [47].

Axiom 5.4.3 (IIM). *If K_1, K_2 are tournaments of the same size with identical a-th and a'-th rows, then $a \preceq^{\varphi}_{K_1} a'$ iff $a \preceq^{\varphi}_{K_2} a'$.*

IIM is a strong property, which says the relative ranking of a and a' does not depend on the results of any match not involving a or a'. This axiom has been questioned for generalised tournaments [47], and a similar argument can be made against it here: although each player in A faces the same opponents, we may wish to take the *strength* of opponents into account, e.g. by rewarding victories against highly-ranked players in B. Consequently we do not view **IIM** as an essential requirement, but rather introduce it to facilitate comparison with our work and the existing tournament literature.

Monotonicity Properties. Our final axioms are monotonicity properties, which express the idea that *more victories are better*. The first axiom follows our original intuition for constructing the natural ranking associated with a chain graph; namely that $K(a) \subseteq K(a')$ indicates a' has performed at least as well as a.

Axiom 5.4.4 (mon)**.** *If* $K(a) \subseteq K(a')$ *then* $a \preceq_K^{\varphi} a'$.

Note that **mon** simply says $\preceq_K^{\varphi}$ extends the (in general, partial) preorder $\leqslant_K^{A}$. Yet another standard axiom is *positive responsiveness*.

Axiom 5.4.5 (pos-resp)**.** *If* $a \preceq_K^{\varphi} a'$ *and* $K_{a',b} = 0$ *for some* $b \in B$, *then* $a \prec_{K+\mathbf{1}_{a',b}}^{\varphi} a'$, *where* $\mathbf{1}_{a',b}$ *is the matrix with 1 in position* (a', b) *and zeros elsewhere.*

That is, adding an extra victory for a should only improve its ranking, with ties now broken in its favour. This version of positive responsiveness was again introduced in [85], where together with **anon** and **IIM** it characterises the *points system* ranking method for round-robin tournaments, which simply ranks players according to the number of victories. The analogous operator in our framework is φ_{count}, and it can be shown that φ_{count} is uniquely characterised by **anon**, **IIM**, **pos-resp** and **dual**. Finally, note that **pos-resp** also acts as a kind of *strategyproofness*: a cannot improve its ranking by deliberately losing a match. Specifically, if $K_{ab} = 1$ and $a \preceq_K^{\varphi} a'$, then **pos-resp** implies $a \prec_{K-\mathbf{1}_{ab}}^{\varphi} a'$.

5.4.2 Axiom Compatibility with chain-min

We come to analysing the compatibility of **chain-min** with the axioms. First, the negative results.

Theorem 5.4.1. *There is no operator satisfying **chain-min** and any of **anon**, **IIM** or **pos-resp**.*

Proof. We take each axiom in turn. Let φ be any operator satisfying **chain-min**.

anon: Consider $K = \left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$, and define permutations $\sigma = \pi = (1\ 2)$, i.e. the permutations which simply swap 1 and 2. It is easily seen that $\pi(\sigma(K)) = K$. Supposing φ satisfied **anon**, we would get $1 \preceq_K^{\varphi} 2$ iff $\sigma(1) \preceq_{\pi(\sigma(K))}^{\varphi} \sigma(2)$ iff $2 \preceq_K^{\varphi} 1$, which implies $1 \approx_K^{\varphi} 2$. On the other hand, we have

$$\mathcal{M}(K) = \{\left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right], \left[\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right], \left[\begin{smallmatrix}1 & 0\\ 0 & 0\end{smallmatrix}\right], \left[\begin{smallmatrix}0 & 0\\ 0 & 1\end{smallmatrix}\right]\}$$

Since φ satisfies **chain-min** and $1, 2 \in A$ rank equally in $\preceq^{\varphi}_{K}$, there must be $K' \in \mathcal{M}(K)$ such that 1 and 2 rank equally in $\leqslant^{\mathcal{A}}_{K'}$, i.e. $K'(1) = K'(2)$. But clearly there is no such K'; all tournaments in $\mathcal{M}(K)$ have distinct first and second rows. Hence φ cannot satisfy **anon**.

IIM: Suppose φ satisfies **chain-min** and **IIM**. Write

$$K_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Note that the first and second rows of K_1 and K_2 are identical, so by **IIM** we have $1 \preceq^{\varphi}_{K_1} 2$ iff $1 \preceq^{\varphi}_{K_2} 2$. Both tournaments have a unique closest chain tournament requiring changes to only a single entry:

$$\mathcal{M}(K_1) = \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \right\}, \quad \mathcal{M}(K_2) = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right\}$$

Write ${K_1}'$ and ${K_2}'$ for these nearest chain tournaments respectively. By **chain-min**, we must have $\varphi(K_i) = (\leqslant^{\mathcal{A}}_{{K_i}'}, \leqslant^{\mathcal{B}}_{{K_i}'})$. In particular, $1 \prec^{\varphi}_{K_1} 2$ and $2 \prec^{\varphi}_{K_2} 1$. But this contradicts **IIM**, and we are done.

pos-resp: Suppose φ satisfies **chain-min** and **pos-resp**, and consider

$$K = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

K has a unique closest chain tournament K':

$$\mathcal{M}(K) = \{K'\} = \left\{ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

chain-min therefore implies $\varphi(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$. Note that $K'(1) = K'(2)$, so we have $1 \approx^{\varphi}_{K} 2$. In particular, $1 \preceq^{\varphi}_{K} 2$. Since $K_{23} = 0$, we may apply **pos-resp** to get $1 \prec^{\varphi}_{K+\mathbf{1}_{23}} 2$. But $K + \mathbf{1}_{23}$ is just K'. Since the chain property already holds for K', we have $\mathcal{M}(K') = \{K'\}$ and consequently

$$\varphi(K + \mathbf{1}_{23}) = \varphi(K') = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'}) = \varphi(K)$$

so in fact $1 \approx^{\varphi}_{K+\mathbf{1}_{23}} 2$, contradicting **pos-resp**. □

Note that the counterexample for **anon** is particularly simple: we take $K = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Swapping the rows and columns brings us back to K, so **anon** implies $1, 2 \in A$ rank equally. However, we saw that no chain tournament in $\mathcal{M}(K)$ yields this ranking.

The MLE results of Section 5.3.2 provides informal explanation for this result. For K above to arise in the noise model of Definition 5.3.3 there must have been two 'mistakes' (false positives or false negatives). This is less likely than a single mistake from just one of $1, 2 \in A$, but the likelihood maximisation forces us to choose one or the other. A similar argument explains the **pos-resp** failure.

It is also worth noting that **anon** only fails at the last step of chain editing, where a single element of $\mathcal{M}(K)$ is chosen. Indeed, the set $\mathcal{M}(K)$ itself *does* exhibit the kind of symmetry one might expect: we have $\mathcal{M}(\pi(\sigma(K))) = \{\pi(\sigma(K')) \mid K' \in \mathcal{M}(K)\}$. This means that an operator which aggregates the rankings from *all* $K' \in \mathcal{M}(K)$ – e.g. any anonymous social welfare function – would satisfy **anon**. The other axioms are compatible with **chain-min**.

Theorem 5.4.2. *For each of **dual** and **mon**, there exists an operator satisfying **chain-min** and the stated property.*

Despite the simplicity of **mon**, Theorem 5.4.2 is deceptively difficult to prove, and we devote the rest of this section to its proof. We describe operators satisfying **chain-def** and **dual** or **mon** non-constructively by first taking an *arbitrary* chain-minimal operator φ, and using properties of the set $\mathcal{M}(K)$ to produce another operator φ' satisfying **dual** or **mon**. Note also that we have not yet constructed an operator satisfying **dual**, **mon** and **chain-min** simultaneously, although we conjecture that such operators do exist.

First we show compatibility of **chain-min** and **dual**. We need a preliminary result.

Lemma 5.4.1. *Let K be a tournament. Then*

1. $\leqslant^{\mathcal{B}}_{K} = \leqslant^{\mathcal{A}}_{\overline{K}}$
2. $K' \in \mathcal{M}(K)$ *if and only if* $\overline{K'} \in \mathcal{M}\left(\overline{K}\right)$

Proof. Fix an $m \times n$ tournament K.

1. Note that for any $b \in B$, we have $K^{-1}(b) = A \setminus \overline{K}(b)$. Indeed, for any $a \in A = A_K = B_{\overline{K}}$,

$$\begin{aligned} a \in K^{-1}(b) &\iff K_{ab} = 1 \\ &\iff 1 - K_{ab} = 0 \\ &\iff \overline{K}_{ba} = 0 \\ &\iff a \notin \overline{K}(b) \end{aligned}$$

This means that for any $b, b' \in B$,

$$\begin{aligned} b \leqslant^{\mathcal{B}}_{K} b' &\iff K^{-1}(b) \supseteq K^{-1}(b') \\ &\iff A \setminus \overline{K}(b) \supseteq A \setminus \overline{K}(b') \\ &\iff \overline{K}(b) \subseteq \overline{K}(b') \\ &\iff b \leqslant^{\mathcal{A}}_{\overline{K}} b' \end{aligned}$$

so $\leqslant^{\mathcal{B}}_{K} = \leqslant^{\mathcal{A}}_{\overline{K}}$.

2. 'only if": Suppose $K' \in \mathcal{M}(K)$. First we show that $\overline{K'}$ has the chain property. It is sufficient to show that $\leqslant^{\mathcal{B}}_{K'}$ is a total preorder,[11] since part (1) then implies $\leqslant^{\mathcal{A}}_{\overline{K'}}$ is a total preorder and $\overline{K'}$ has the chain property by definition.

   Since $\leqslant^{\mathcal{B}}_{K'}$ always has reflexivity and transitivity, we only need to show the totality property. Let $b, b' \in B$ and suppose $b \not\leqslant^{\mathcal{B}}_{K'} b'$. We must show $b' \leqslant^{\mathcal{B}}_{K'} b$, i.e. $(K')^{-1}(b') \supseteq (K')^{-1}(b)$. To that end, let $a \in (K')^{-1}(b)$.

   Since $(K')^{-1}(b) \not\supseteq (K')^{-1}(b')$, there is some $\hat{a} \in (K')^{-1}(b')$ with $\hat{a} \notin (K')^{-1}(b)$. That is, $b' \in K'(\hat{a})$ but $b \notin K'(\hat{a})$. Since $b \in K'(a)$, we have $K'(a) \not\subseteq K'(\hat{a})$. By the chain property for K', we get $K'(\hat{a}) \subset K'(a)$. Finally, this means $b' \in K'(\hat{a}) \subseteq K'(a)$, i.e $a \in (K')^{-1}(b')$. This shows $b' \leqslant^{\mathcal{B}}_{K'} b$ as required.

   It remains to show that $d(\overline{K}, \overline{K'})$ is minimal. Since every tournament is the dual of its dual, any $n \times m$ chain tournament is of the form $\overline{K''}$ for an $m \times n$ tournament K''. The above argument shows that the chain property is preserved by taking the dual, so that K'' has the chain property also. Since $K' \in \mathcal{M}(K)$, we have $d(K, K'') \geq d(K, K')$. It is easily verified that the Hamming distance is also preserved under duals, so

   $$d(\overline{K}, \overline{K'}) = d(K, K') \leq d(K, K'') = d(\overline{K}, \overline{K''})$$

   We have shown that $\overline{K'}$ is as close to $\overline{K}$ as any other $n \times m$ tournament with the chain property, which shows $\overline{K'} \in \mathcal{M}\left(\overline{K}\right)$ as required.

   'if": Suppose $\overline{K'} \in \mathcal{M}\left(\overline{K}\right)$. By the 'only if' statement above, we have $\overline{\overline{K'}} \in \mathcal{M}\left(\overline{\overline{K}}\right)$. But $\overline{\overline{K}} = K$ and $\overline{\overline{K'}} = K'$, so $K' \in \mathcal{M}(K)$ as required.

∎

We can now find an operator with both **chain-min** and **dual**.

Proposition 5.4.1. *There exists an operator φ satisfying **chain-min** and **dual**.*

Proof. Let φ be an arbitrary operator satisfying **chain-min**. Then there is a function $\alpha : \mathcal{K} \to \mathcal{K}$ such that $\varphi(K) = (\leqslant^{\mathcal{A}}_{\alpha(K)}, \leqslant^{\mathcal{B}}_{\alpha(K)})$ and $\alpha(K) \in \mathcal{M}(K)$ for all tournaments K. We will construct a new function α', based on α, such that $\alpha'(\overline{K}) = \overline{\alpha'(K)}$.

Let $\ll$ be a total order on the set of all tournaments $\mathcal{K}$.[12] Write

$$T = \{K \in \mathcal{K} \mid K \ll \overline{K}\}$$

Note that since $K \neq \overline{K}$ for all K, exactly one of K and $\overline{K}$ lies in T. Informally, we view the tournaments in T as somehow 'canonical', and those in $\mathcal{K} \setminus T$ as the dual

[11] Note that we claim this holds for any K' with the chain property, but this has not yet been proven.

of a canonical tournament. We use this notion to define α':

$$\alpha'(K) = \begin{cases} \alpha(K), & K \in T \\ \overline{\alpha(\overline{K})}, & K \notin T \end{cases}$$

First we claim $\alpha'(K) \in \mathcal{M}(K)$ for all K. Indeed, if $K \in T$ then $\alpha'(K) = \alpha(K) \in \mathcal{M}(K)$ by the assumption on α. Otherwise, $\alpha(\overline{K}) \in \mathcal{M}\left(\overline{K}\right)$, so Lemma 5.4.1 part (2) implies $\alpha'(K) = \overline{\alpha(\overline{K})} \in \mathcal{M}\left(\overline{\overline{K}}\right) = \mathcal{M}(K)$.

Next we show $\overline{\alpha'(K)} = \alpha'(\overline{K})$. First suppose $K \in T$. Then $\alpha'(K) = \alpha(K)$ and $\overline{K} \notin T$, so $\alpha'(\overline{K}) = \overline{\alpha(\overline{\overline{K}})} = \overline{\alpha(K)} = \overline{\alpha'(K)}$ as required. Similarly, if $K \notin T$ then $\overline{K} \in T$, so $\alpha'(\overline{K}) = \alpha(\overline{K})$, and $\alpha'(K) = \overline{\alpha(\overline{K})} = \overline{\alpha'(\overline{K})}$. Taking the dual of both sides, we get $\overline{\alpha'(K)} = \alpha'(\overline{K})$.

Finally, define a new operator φ' by $\varphi'(K) = (\leqslant^{\mathcal{A}}_{\alpha'(K)}, \leqslant^{\mathcal{B}}_{\alpha'(K)})$. Since $\alpha'(K) \in \mathcal{M}(K)$ for all K, φ' satisfies **chain-min**. Moreover, using Lemma 5.4.1 part (1) and the fact that $\overline{\alpha'(K)} = \alpha'(\overline{K})$, for any tournament K and $b, b' \in B$ we have

$$\begin{aligned} b \sqsubseteq^{\varphi'}_{K} b' &\iff b \leqslant^{\mathcal{B}}_{\alpha'(K)} b' \\ &\iff b \leqslant^{\mathcal{A}}_{\overline{\alpha'(K)}} b' \\ &\iff b \leqslant^{\mathcal{A}}_{\alpha'(\overline{K})} b' \\ &\iff b \sqsubseteq^{\varphi'}_{\overline{K}} b' \end{aligned}$$

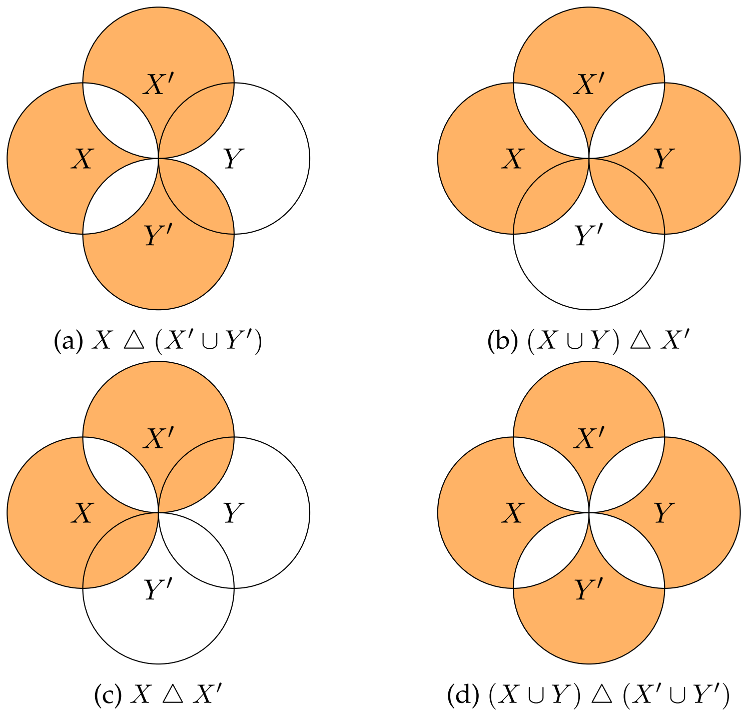
which shows φ' also satisfies **dual**. □

To find an operator satisfying **chain-min** and **mon**, we proceed in three stages. First, Lemma 5.4.2 shows that if $K(a_1) \subseteq K(a_2)$ and $K' \in \mathcal{M}(K)$ is some closest chain tournament with the reverse inclusion $K'(a_2) \subseteq K'(a_1)$, then swapping a_1 and a_2 in K' yields obtain another closest chain tournament $K'' \in \mathcal{M}(K)$. Next, we show in Lemma 5.4.3 that by performing successive swaps in this way, we can find $K' \in \mathcal{M}(K)$ such that $K'(a_1) \subseteq K'(a_2)$ whenever $K(a_1) \subset K(a_2)$ (note the strict inclusion). Finally, we modify this K' in Lemma 5.4.4 to additionally satisfy $K'(a_1) = K'(a_2)$ whenever $K(a_1) = K(a_2)$. This shows that there always exist an element of $\mathcal{M}(K)$ extending the neighbourhood-subset relation $\leqslant^{\mathcal{A}}_{K}$, and consequently it is possible to satisfy **chain-min** and **mon** simultaneously.

Definition 5.4.2. *Let K be a tournament and $a_1, a_2 \in A$. We denote by* $\mathsf{swap}(K; a_1, a_2)$ *the tournament obtained by swapping the a_1 and a_2-th rows of K, i.e.*

$$[\mathsf{swap}(K; a_1, a_2)]_{ab} = \begin{cases} K_{a_1,b}, & a = a_2 \\ K_{a_2,b}, & a = a_1 \\ K_{a,b}, & a \notin \{a_1, a_2\} \end{cases}$$

[12] Note that $\mathcal{K}$ is countable, so such an order can be easily constructed. Alternatively, one could use the axiom of choice and appeal to the well-ordering theorem to obtain $\ll$.



(a) $X \triangle (X' \cup Y')$

(b) $(X \cup Y) \triangle X'$

(c) $X \triangle X'$

(d) $(X \cup Y) \triangle (X' \cup Y')$

Figure 5.2: Depictions of the sets in Equation (5.5)

Lemma 5.4.2. *Suppose* $K(a_1) \subseteq K(a_2)$ *and* $K' \in \mathcal{M}(K)$ *is such that* $K'(a_2) \subseteq K'(a_1)$. *Then* $\mathsf{swap}(K'; a_1, a_2) \in \mathcal{M}(K)$.

Proof. Write $K'' = \mathsf{swap}(K'; a_1, a_2)$. It is clear that K'' has the chain property since K' does. Since $K' \in \mathcal{M}(K)$, we have $d(K, K'') \geq d(K, K')$. We will show that $d(K, K'') \leq d(K, K')$ also, which implies $d(K, K'') = d(K, K') = m(K)$ and thus $K'' \in \mathcal{M}(K)$.

To that end, observe that for any tournament $\hat{K}$,

$$d(K, \hat{K}) = \sum_{a \in A} |K(a) \triangle \hat{K}(a)|$$

Noting that $K'(a) = K''(a)$ for $a \notin \{a_1, a_2\}$ and $K''(a_1) = K'(a_2)$, $K''(a_2) = K'(a_1)$, we have

$$\begin{aligned} d(K, K') - d(K, K'') &= \sum_{i \in \{1,2\}} \big(|K(a_i) \triangle K'(a_i)| - |K(a_i) \triangle K''(a_i)|\big) \\ &= |K(a_1) \triangle K'(a_1)| - |K(a_1) \triangle K'(a_2)| \\ &\quad + |K(a_2) \triangle K'(a_2)| - |K(a_2) \triangle K'(a_1)| \end{aligned}$$

To simplify notation, write $X = K(a_1)$, $X' = K'(a_2)$, $Y = K(a_2) \setminus K(a_1)$ and $Y' = K'(a_1) \setminus K'(a_2)$. Since $K(a_1) \subseteq K(a_2)$ and $K'(a_2) \subseteq K'(a_1)$ by hypothesis, we have

$$\begin{aligned} K(a_1) &= X; \qquad & K(a_2) &= X \cup Y \\ K'(a_1) &= X' \cup Y'; \qquad & K'(a_2) &= X' \end{aligned}$$

and $X \cap Y = X' \cap Y' = \emptyset$. Rewriting the above we have

$$\begin{aligned} d(K, K') - d(K, K'') &= |K(a_1) \bigtriangleup K'(a_1)| + |K(a_2) \bigtriangleup K'(a_2)| \\ &\quad - |K(a_1) \bigtriangleup K'(a_2)| - |K(a_2) \bigtriangleup K'(a_1)| \\ &= |X \bigtriangleup (X' \cup Y')| + |(X \cup Y) \bigtriangleup X'| \\ &\quad - |X \bigtriangleup X'| - |(X \cup Y) \bigtriangleup (X' \cup Y')| \end{aligned} \tag{5.5}$$

Each of the symmetric differences in Eq. (5.5) are depicted in Figure 5.2. Note that each of these sets can be expressed as a union of the 8 disjoint subsets of $X \cup Y \cup X' \cup Y'$ shown in the figure. Expanding the symmetric differences in Eq. (5.5) and consulting Figure 5.2, it can be seen that most terms cancel out, and in fact we are left with

$$d(K, K') - d(K, K'') = 2|Y \cap Y'| \geq 0$$

This shows that $d(K, K'') \leq d(K, K')$, and the proof is complete. □

Notation. *For a relation R on a set X and $x \in X$, write*

$$U(x, R) = \{y \in X \mid x \; R \; y\}$$

$$L(x, R) = \{y \in X \mid y \; R \; x\}$$

for the upper- and lower-sets of x respectively.

Lemma 5.4.3. *For any tournament K there is $K' \in \mathcal{M}(K)$ such that for all $a \in A$:*

$$U(a, <_K^{\mathcal{A}}) \subseteq U(a, \leqslant_{K'}^{\mathcal{A}})$$

That is, $K(a) \subset K(a')$ implies $K'(a) \subseteq K'(a')$ for all $a, a' \in A$.

Proof. Write $A = \{a_1, \ldots, a_m\}$, ordered such that $|L(a_1, \leqslant_K^{\mathcal{A}})| \leq \cdots \leq |L(a_m, \leqslant_K^{\mathcal{A}})|$. We will show by induction that for each $0 \leq i \leq m$ there is $K_i \in \mathcal{M}(K)$ such that:

$$1 \leq j \leq i \implies U(a_j, <_K^{\mathcal{A}}) \subseteq U(a_j, \leqslant_{K_i}^{\mathcal{A}}) \tag{$*$}$$

The result follows by taking $K' = K_m$.

The case $i = 0$ is vacuously true, and we may take K_0 to be an arbitrary member of $\mathcal{M}(K)$. For the inductive step, suppose ($*$) holds for some $0 \leq i < m$. If $U(a_{i+1}, <_K^{\mathcal{A}}) = \emptyset$ then we may set $K_{i+1} = K_i$, so assume that $U(a_{i+1}, <_K^{\mathcal{A}})$ is non-empty. Take some $\hat{a} \in \min(U(a_{i+1}, <_K^{\mathcal{A}}), \leqslant_{K_i}^{\mathcal{A}})$. Then $\hat{a}$ has (one of) the smallest neighbourhoods in K_i amongst those in A with a strictly larger neighbourhood than a_{i+1} in K.

If $K_i(a_{i+1}) \subseteq K_i(\hat{a})$ then we claim (∗) holds with $K_{i+1} = K_i$. Indeed, for $j < i+1$ the inclusion in (∗) holds since it does for K_i. For $j = i+1$, let $a \in U(a_{i+1}, <_K^{\mathcal{A}})$. The definition of $\hat{a}$ implies $K_i(a) \not\subset K_i(\hat{a})$; since K_i has the chain property this means $K_i(\hat{a}) \subseteq K_i(a)$. Consequently $K_i(a_{i+1}) \subseteq K_i(\hat{a}) \subseteq K_i(a)$, i.e. $a \in U(a_{i+1}, \leqslant_{K_i}^{\mathcal{A}}) = U(a_{i+1}, \leqslant_{K_{i+1}}^{\mathcal{A}})$ as required.

For the remainder of the proof we therefore suppose $K_i(a_{i+1}) \not\subseteq K_i(\hat{a})$. The chain property for K_i gives $K_i(\hat{a}) \subset K_i(a_{i+1})$. Since $K_i \in \mathcal{M}(K)$ and $K(a_{i+1}) \subset K(\hat{a})$, we may apply Lemma 5.4.2. Set $K_{i+1} = \mathsf{swap}(K_i; a_{i+1}, \hat{a}) \in \mathcal{M}(K)$. The inclusion in (∗) is easy to show for $j = i+1$: if $a \in U(a_{i+1}, <_K^{\mathcal{A}})$ then either $a = \hat{a}$ – in which case $K_{i+1}(a_{i+1}) \subset K_{i+1}(a)$ by construction – or $a \neq \hat{a}$ and $K_{i+1}(a_{i+1}) = K_i(\hat{a}) \subseteq K_i(a) = K_{i+1}(a)$. In either case $a \in U(a_{i+1}, \leqslant_{K_{i+1}}^{\mathcal{A}})$ as required.

Now suppose $1 \leq j < i+1$. First note that due to our assumption on the ordering of $\{a_1, \ldots, a_m\}$, we have $a_j \neq \hat{a}$ (indeed, if $a_j = \hat{a}$ then $K(a_{i+1}) \subset K(a_j)$ and $|L(a_j, <_K^{\mathcal{A}})| > |L(a_{i+1}, <_K^{\mathcal{A}})|$). Since $a_j \neq a_{i+1}$ also, a_j was not involved in the swapping in the construction of K_{i+1}, and consequently $K_{i+1}(a_j) = K_i(a_j)$. Let $a \in U(a_j, <_K^{\mathcal{A}})$. We must show that $K_{i+1}(a_j) \subseteq K_{i+1}(a)$. We consider cases.

Case 1: $a = \hat{a}$. Using the fact that (∗) holds for K_i we have

$$K_{i+1}(a_j) = K_i(a_j) \subseteq K_i(\hat{a}) \subset K_i(a_{i+1}) = K_{i+1}(\hat{a})$$

Case 2: $a = a_{i+1}$. Here $K(a_j) \subset K(a_{i+1}) \subset K(\hat{a})$, i.e. $\hat{a} \in U(a_j, <_K^{\mathcal{A}})$. Applying the inductive hypothesis again we have

$$K_{i+1}(a_j) = K_i(a_j) \subseteq K_i(\hat{a}) = K_{i+1}(a_{i+1})$$

Case 3: $a \notin \{\hat{a}, a_{i+1}\}$. Here neither a_j nor a were involved in the swap, so $K_{i+1}(a_j) = K_i(a_j) \subseteq K_i(a) = K_{i+1}(a)$.

By induction, the proof is complete. □

Lemma 5.4.4. *Let K be a tournament and suppose $K' \in \mathcal{M}(K)$ is such that $U(a, <_K^{\mathcal{A}}) \subseteq U(a, \leqslant_{K'}^{\mathcal{A}})$ for all $a \in A$. Then there is $K'' \in \mathcal{M}(K)$ such that $\leqslant_K^{\mathcal{A}} \subseteq \leqslant_{K''}^{\mathcal{A}}$.*

Proof. Let $A_1, \ldots, A_t \subseteq A$ be the equivalence classes of $\approx_K^{\mathcal{A}}$, the symmetric part of $\leqslant_K^{\mathcal{A}}$. Note that $a \approx_K^{\mathcal{A}} a'$ iff $K(a) = K(a')$, so we can associate each A_i with a neighbourhood $B_i \subseteq B$ such that $K(a) = B_i$ whenever $a \in A_i$.

Our aim is to select a single element from each equivalence class A_i, which we denote by $f(A_i)$, and modify K' to set the neighbourhood of each $a \in A_i$ to $K'(f(A_i))$. To that end, construct a function $f : \{A_1, \ldots, A_t\} \to A$ such that

$$f(A_i) \in \operatorname*{arg\,min}_{a \in A_i} |B_i \triangle K'(a)| \in A_i$$

Define K'' by $K''_{ab} = K'_{f([a]),b}$, where $[a]$ denotes the equivalence class of a. Then $K''(a) = K'(f([a]))$ for all a.

Next we show that $K'' \in \mathcal{M}(K)$. Note that K'' has the chain property, since $a_1 \leqslant^{\mathcal{A}}_{K''} a_2$ iff $f([a_1]) \leqslant^{\mathcal{A}}_{K'} f([a_2])$, and $f([a_1]), f([a_2])$ are guaranteed to be comparable with respect to $\leqslant^{\mathcal{A}}_{K'}$ since K' has the chain property. To show $d(K, K'')$ is minimal, observe that

$$\begin{aligned} d(K, K'') &= \sum_{a \in A} |K(a) \triangle K''(a)| \\ &= \sum_{i=1}^{t} \sum_{a \in A_i} |B_i \triangle K'(f(A_i))| \end{aligned}$$

By definition of f, we have $|B_i \triangle K'(f(A_i))| \leq |B_i \triangle K'(a)|$ for all $a \in A_i$. Consequently

$$\begin{aligned} d(K, K'') &\leq \sum_{i=1}^{t} \sum_{a \in A_i} |B_i \triangle K'(a)| \\ &= d(K, K') \\ &= m(K) \end{aligned}$$

which implies $K'' \in \mathcal{M}(K)$.

We are now ready to prove the result. Suppose $a \leqslant^{\mathcal{A}}_{K} a'$ i.e. $K(a) \subseteq K(a')$. If $K(a) = K(a')$ then $[a] = [a']$, so

$$K''(a) = K'(f([a])) = K'(f([a'])) = K''(a')$$

and in particular $K''(a) \subseteq K''(a')$. If instead $K(a) \subset K(a')$, then $K(f([a])) = K(a) \subset K(a') = K(f([a']))$, i.e. $f([a]) <^{\mathcal{A}}_{K} f([a'])$. By the assumption on K' in the statement of the lemma, this means $f([a]) \leqslant^{\mathcal{A}}_{K'} f([a'])$, and so

$$K''(a) = K'(f([a])) \subseteq K'(f([a'])) = K''(a')$$

In either case $K''(a) \subseteq K''(a')$, i.e. $a \leqslant^{\mathcal{A}}_{K''} a'$. Since a, a' were arbitrary, this shows that $\leqslant^{\mathcal{A}}_{K} \subseteq \leqslant^{\mathcal{A}}_{K''}$ as required. $\square$

The pieces are now in place to prove the following.

Proposition 5.4.2. *There exists an operator φ satisfying **chain-min** and **mon**.*

Proof. For any tournament K, write

$$\mathcal{M}_{\mathsf{mon}}(K) = \{K' \in \mathcal{M}(K) \mid \leqslant^{\mathcal{A}}_{K} \subseteq \leqslant^{\mathcal{A}}_{K'}\}$$

By Lemma 5.4.3 and Lemma 5.4.4, $\mathcal{M}_{\mathsf{mon}}(K)$ is non-empty. Let $\ll$ be any total order on the set $\mathcal{K}$ of all tournaments. Define a function $\alpha : \mathcal{K} \to \mathcal{K}$ by

$$\alpha(K) = \min(\mathcal{M}_{\mathsf{mon}}(K), \ll) \in \mathcal{M}_{\mathsf{mon}}(K)$$

Note that the minimum is unique since $\ll$ is a total order. Defining an operator φ by $\varphi(K) = (\leqslant^{\mathcal{A}}_{\alpha(K)}, \leqslant^{\mathcal{B}}_{\alpha(K)})$, we see that φ satisfies **chain-min** and **mon**, as required. $\square$

Theorem 5.4.2 now follows from Proposition 5.4.1 and Proposition 5.4.2.

5.5 Match-preference operators

The counterexample for **chain-min** and **anon** suggests that chain-minimal operators require some form of tie-breaking mechanism when the tournaments in $\mathcal{M}(K)$ cannot be distinguished while respecting anonymity. While this limits the use of chain-minimal operators as general purpose ranking methods, it is not such a problem if additional information is available to guide the tie-breaking. In this section we introduce a new class of operators for this case.

The core idea is to single out a unique chain tournament close to K by paying attention to not only the *number* of entries in K that need to be changed to produce a chain tournament, *which* entries. Specifically, we assume the availability of a total order on the set of matrix indices $\mathbb{N} \times \mathbb{N}$ (the *matches*) which indicates our willingness to change an entry in K: the higher up (a, b) is in the ranking, the more acceptable it is to change K_{ab} during chain editing.

This total order – called the *match-preference relation* – is fixed for all tournaments K; this means we are dealing with extra information about how tournaments are *constructed in matrix form*, not extra information about any specific tournament K.

One possible motivation for such a ranking comes from cases where matches occur at distinct points in time. In this case the matches occurring more recently are (presumably) more representative of the players' *current* abilities, and we should therefore prefer to modify the outcome of old matches where possible.

For the formal definition we need notation for the *vectorisation* of a tournament K: for a total order $\trianglelefteq$ on $\mathbb{N} \times \mathbb{N}$ and an $m \times n$ tournament K, we write $\mathrm{vec}_{\trianglelefteq}(K)$ for the vector in $\{0, 1\}^{mn}$ obtained by collecting the entries of K in the order given by $\trianglelefteq \restriction (A \times B)$,[13] starting with the minimal entry. That is, $\mathrm{vec}_{\trianglelefteq}(K) = (K_{a_1,b_1}, \ldots, K_{a_{mn},b_{mn}})$, where $(a_1, b_1), \ldots, (a_{mn}, b_{mn})$ is the unique enumeration of $A \times B$ such that $(a_i, b_i) \trianglelefteq (a_{i+1}, b_{i+1})$ for each i.

The operator corresponding to $\trianglelefteq$ is defined using the notion of a *choice function*: a function α which maps any tournament K to an element of $\mathcal{M}(K)$. Any such function defines a chain-minimal operator φ by setting $\varphi(K) = (\leqslant^{\mathcal{A}}_{\alpha(K)}, \leqslant^{\mathcal{B}}_{\alpha(K)})$.

Definition 5.5.1. *Let $\trianglelefteq$ be a total order on $\mathbb{N} \times \mathbb{N}$. Define an operator $\varphi_{\trianglelefteq}$ according to the choice function*

$$\alpha_{\trianglelefteq}(K) = \underset{K' \in \mathcal{M}(K)}{\arg\min}\ \mathrm{vec}_{\trianglelefteq}(K \oplus K') \tag{5.6}$$

[13] This denotes the restriction of $\trianglelefteq$ to $A \times B$, i.e. $\trianglelefteq \cap ((A \times B) \times (A \times B))$.

where $[K \oplus K']_{ab} = |K_{ab} - K'_{ab}|$, *and the minimum is taken w.r.t the lexicographic ordering on* $\{0,1\}^{|A|\cdot|B|}$.[14] *Operators generated in this way will be called* match-preference operators.

Example 5.5.1. *Let* $\trianglelefteq$ *be the lexicographic order*[15] *on* $\mathbb{N} \times \mathbb{N}$ *so that* $\mathrm{vec}_{\trianglelefteq}(K \oplus K')$ *is obtained by collecting the entries of* $K \oplus K'$ *row-by-row, from top to bottom and left to right. Take* K *from Example 5.3.1. Writing* $K_1, \dots, K_4$ *for the elements of* $\mathcal{M}(K)$ *in the order that they appear in Example 5.3.1 and setting* $v_i = \mathrm{vec}_{\trianglelefteq}(K \oplus K_i)$, *we have*

$$v_1 = (0100\ 0000\ 10000); \qquad v_2 = (0010\ 0000\ 10000)$$
$$v_3 = (0000\ 0100\ 10000); \qquad v_4 = (0000\ 0010\ 10000)$$

The lexicographic minimum is the one with the 1 entries as far right as possible, which in this case is v_4. *Consequently* $\varphi_{\trianglelefteq}$ *ranks* K *according to* K_4, *i.e.* $1 \prec_K^{\varphi_{\trianglelefteq}} 2 \prec_K^{\varphi_{\trianglelefteq}} 3$ *and* $1 \approx_K^{\varphi_{\trianglelefteq}} 3 \sqsubset_K^{\varphi_{\trianglelefteq}} 2 \sqsubset_K^{\varphi_{\trianglelefteq}} 4$.

To conclude the discussion of match-preference operators, we note that one can compute $\alpha_{\trianglelefteq}(K)$ as the unique closest chain tournament to K w.r.t a *weighted* Hamming distance, and thereby avoid the need to enumerate $\mathcal{M}(K)$ in full as per Eq. (5.6). First, a technical result is required.

Lemma 5.5.1. *Let* k *and* l *be integers with* $1 \le k \le l$. *Then*

$$\sum_{i=k}^{l} 2^{-i} < 2^{-(k-1)}.$$

Proof. This follows from the formula for the sum of a finite geometric series:

$$\sum_{i=0}^{n-1} r^i = \frac{1-r^n}{1-r}$$

which holds for all $r \neq 1$. In this case we have

$$\begin{aligned}
\sum_{i=k}^{l} 2^{-i} &= \sum_{i=0}^{l} 2^{-i} - \sum_{i=0}^{k-1} 2^{-i} \\
&= \sum_{i=0}^{l} \left(\frac{1}{2}\right)^i - \sum_{i=0}^{k-1} \left(\frac{1}{2}\right)^i \\
&= \frac{1-\left(\frac{1}{2}\right)^{l+1}}{1-\left(\frac{1}{2}\right)} - \frac{1-\left(\frac{1}{2}\right)^{k}}{1-\left(\frac{1}{2}\right)} \\
&= 2\left(2^{-k} - 2^{-(l+1)}\right) \\
&= 2^{-(k-1)} - \underbrace{2^{-l}}_{>0} \\
&< 2^{-(k-1)}
\end{aligned}$$

[14] Note that $K \oplus K'$ is 1 in exactly the entries where K and K' differ.

[15] That is, $(a,b) \trianglelefteq (a',b')$ iff $a < a'$ or ($a = a'$ and $b \le b'$).

as required. □

The characterisation in terms of weighted Hamming distances is as follows

Theorem 5.5.1. *Let $\trianglelefteq$ be a total order on $\mathbb{N} \times \mathbb{N}$. Then for any $m, n \in \mathbb{N}$ there exists a function $w : [m] \times [n] \to \mathbb{R}_{\geq 0}$ such that for all $m \times n$ tournaments K:*

$$\underset{K' \in \mathcal{C}_{m,n}}{\arg\min}\, d_w(K, K') = \{\alpha_{\trianglelefteq}(K)\} \tag{5.7}$$

where $d_w(K, K') = \sum_{(a,b) \in [m] \times [n]} w(a, b) \cdot |K_{ab} - K'_{ab}|$.

Proof. Let $\trianglelefteq$ be a total order on $\mathbb{N} \times \mathbb{N}$ and let $m, n \in \mathbb{N}$. For $a \in [m]$ and $b \in [n]$, write

$$p(a, b) = 1 + |\{(a', b') \in [m] \times [n] : (a', b') \lhd (a, b)\}|$$

for the 'position' of (a, b) in $\trianglelefteq \restriction ([m] \times [n])$ (where 1 corresponds to the minimal pair). Define w by

$$w(a, b) = 1 + 2^{-p(a,b)}$$

If we abuse notation slightly and view w as an $m \times n$ matrix, we have, by construction, $\mathrm{vec}_{\trianglelefteq}(w) = (1 + 2^{-1}, \ldots, 1 + 2^{-mn})$. Noting that $|K_{ab} - K'_{ab}| = [K \oplus K']_{ab}$ for any tournaments K, K', and letting $\bullet$ denote the dot product, it is easy to see that

$$\begin{aligned} d_w(K, K') &= \mathrm{vec}_{\trianglelefteq}(w) \bullet \mathrm{vec}_{\trianglelefteq}(K \oplus K') \\ &= (1 + 2^{-1}, \ldots, 1 + 2^{-mn}) \bullet \mathrm{vec}_{\trianglelefteq}(K \oplus K') \\ &= d(K, K') + \boldsymbol{x} \bullet \mathrm{vec}_{\trianglelefteq}(K \oplus K') \end{aligned}$$

where $\boldsymbol{x} = (2^{-1}, \ldots, 2^{-mn})$ and $d(K, K')$ is the unweighted Hamming distance. In particular, since $\boldsymbol{x}$ and $\mathrm{vec}_{\trianglelefteq}(K \oplus K')$ are non-negative, we have $d_w(K, K') \geq d(K, K')$.

Now, we will show that for any $m \times n$ tournament K and $K' \in \mathcal{C}_{m,n}$ with $K' \neq \alpha_{\trianglelefteq}(K)$ we have $d_w(K, \alpha_{\trianglelefteq}(K)) < d_w(K, K')$. Since $\alpha_{\trianglelefteq}(K) \in \mathcal{M}(K) \subseteq \mathcal{C}_{m,n}$ by definition, this will show that $\alpha_{\trianglelefteq}(K)$ is the unique minimum in Equation (5.7), as required.

So, let K be an $m \times n$ tournament and $K' \in \mathcal{C}_{m,n}$. To ease notation, write $v = \mathrm{vec}_{\trianglelefteq}(K \oplus \alpha_{\trianglelefteq}(K))$ and $v' = \mathrm{vec}_{\trianglelefteq}(K \oplus K')$. There are two cases.

Case 1: $K' \notin \mathcal{M}(K)$. In this case we have $d(K, K') \geq m(K) + 1$, and

$$\begin{aligned}
d_w(K, \alpha_{\trianglelefteq}(K)) &= \underbrace{d(K, \alpha_{\trianglelefteq}(K))}_{=m(K)} + \boldsymbol{x} \bullet v \\
&= m(K) + \sum_{i=1}^{mn} 2^{-i} \cdot \underbrace{v_i}_{\leq 1} \\
&\leq m(K) + \underbrace{\sum_{i=1}^{mn} 2^{-i}}_{<2^{-0}=1} \\
&< m(K) + 1 \\
&\leq d(K, K') \\
&\leq d_w(K, K')
\end{aligned}$$

where Lemma 5.5.1 was applied in the 4th step. This shows $d_w(K, \alpha_{\trianglelefteq}(K)) < d_w(K, K')$, as required.

Case 2: $K \in \mathcal{M}(K)$. In this case we have

$$\begin{aligned}
d(K, \alpha_{\trianglelefteq}(K)) - d(K, K') &= (m(K) + \boldsymbol{x} \bullet v) - (m(K) + \boldsymbol{x} \bullet v') \\
&= \boldsymbol{x} \bullet (v - v')
\end{aligned}$$

Now, since $K' \in \mathcal{M}(K)$, v' appears as one of the vectors over which the $\arg\min$ is taken in Equation (5.6). By definition of $\alpha_{\trianglelefteq}$ we therefore know that v strictly precedes v' with respect to the lexicographic order on $\{0, 1\}^{mn}$. Consequently there is $j \geq 1$ such that $v_i = v'_i$ for $i < j$ and $v_j < v'_j$. That is, $v_j = 0$ and $v'_j = 1$. This means

$$\begin{aligned}
d(K, \alpha_{\trianglelefteq}(K)) - d(K, K') &= \boldsymbol{x} \bullet (v - v') \\
&= \sum_{i=1}^{mn} 2^{-i}(v_i - v'_i) \\
&= \sum_{i=1}^{j-1} 2^{-i} \underbrace{(v_i - v'_i)}_{=0} + \sum_{i=j}^{mn} 2^{-i}(v_i - v'_i) \\
&= 2^{-j} \underbrace{(v_j - v'_j)}_{=-1} + \sum_{i=j+1}^{mn} 2^{-i} \underbrace{(v_i - v'_i)}_{\leq 1} \\
&\leq -2^{-j} + \sum_{i=j+1}^{mn} 2^{-i} \\
&< -2^{-j} + 2^{-j} \\
&= 0
\end{aligned}$$

where Lemma 5.5.1 was applied in the second to last step. Again, this shows $d_w(K, \alpha_{\trianglelefteq}(K)) < d_w(K, K')$, and the proof is complete. □

For example, the weights corresponding to $\trianglelefteq$ from Example 5.5.1 and $m = 2$, $n = 3$ are $w = \left[\begin{smallmatrix} 1.5 & 1.25 & 1.125 \\ 1.0625 & 1.03125 & 1.015625 \end{smallmatrix}\right]$.

5.6 Relaxing chain-min

Having studied chain-minimal operators in some detail, we turn to two remaining problems: **chain-min** is incompatible with **anon**, and computing a chain-minimal operator is NP-hard. In this section we obtain both anonymity and tractability by relaxing the **chain-min** requirement to a property we call *chain-definability*. We go on to characterise the class of operators with this weaker property via a greedy approximation algorithm, single out a particularly intuitive instance, revisit the axioms of Section 5.4, and present new axioms which characterise this intuitive instance.

5.6.1 Chain-definability

The source of the difficulties with **chain-min** lies in the minimisation aspect of chain editing. A natural way to retain the spirit of **chain-min** without the complications is to require that $\varphi(K)$ corresponds to *some* chain tournament, not necessarily one closest to K. We call this property *chain-definability*.

Axiom 5.6.1 (chain-def). *For every $m \times n$ tournament K there is $K' \in \mathcal{C}_{m,n}$ such that $\varphi(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$.*

Clearly **chain-min** implies **chain-def**. 'Chain-definable' operators can also be cast in the MLE framework of Section 5.3.2 as those whose rankings correspond to *some* (not necessarily MLE) state θ.

At first glance it may seem difficult to determine whether a given pair of rankings correspond to a chain tournament, since the number of such tournaments grows rapidly with m and n. Fortunately, **chain-def** can be characterised without reference to chain tournaments by considering the number of *ranks* of $\preceq^{\varphi}_{K}$ and $\sqsubseteq^{\varphi}_{K}$. In what follows $\mathsf{ranks}(\preceq)$ denotes the number of ranks of a total preorder $\preceq$, i.e. the number of equivalence classes of its symmetric part.

Theorem 5.6.1. *φ satisfies **chain-def** if and only if $|\mathsf{ranks}(\preceq^{\varphi}_{K}) - \mathsf{ranks}(\sqsubseteq^{\varphi}_{K})| \leq 1$ for every tournament K.*

Proof. First we set up some notation. For a total preorder $\preceq$ on a set Z and $z \in Z$, write $[z]_{\preceq}$ for the rank of $\preceq$ containing z, i.e. the equivalence class of z in the symmetric closure of $\preceq$:

$$[z]_{\preceq} = \{z' \in Z \mid z \preceq z' \text{ and } z' \preceq z\}.$$

Also note that $\preceq$ can be extended to a total order on the ranks by setting $[z]_\preceq \leq [z']_\preceq$ iff $z \preceq z'$.

We start with the 'only if' statement of the theorem. Suppose φ satisfies **chain-def**, and let K be a tournament. We need to show that $|\mathsf{ranks}(\preceq_K^\varphi) - \mathsf{ranks}(\sqsubseteq_K^\varphi)| \leq 1$.

By chain-definability, there is K' with the chain property such that $a \preceq_K^\varphi a'$ iff $K'(a) \subseteq K'(a')$ and $b \sqsubseteq_K^\varphi b'$ iff $(K')^{-1}(b) \supseteq (K')^{-1}(b')$. Write

$$\mathcal{X} = \{[a]_{\preceq_K^\varphi} \mid a \in A, K'(a) \neq \emptyset\}$$

$$\mathcal{Y} = \{[b]_{\sqsubseteq_K^\varphi} \mid b \in B, (K')^{-1}(b) \neq \emptyset\}$$

for the set of ranks in each of the two orders, excluding those who have empty neighbourhoods in K'. Note that $[a]_{\preceq_K^\varphi} = [a']_{\preceq_K^\varphi}$ if and only if $K'(a) = K'(a')$ (and similar for B).

We will show that $|\mathcal{X}| = |\mathcal{Y}|$. Enumerate $\mathcal{X} = \{X_1, \ldots, X_s\}$ and $\mathcal{Y} = \{Y_1, \ldots, Y_t\}$, ordered such that $X_1 < \cdots < X_s$ and $Y_1 < \cdots < Y_t$. First we show $|\mathcal{X}| \leq |\mathcal{Y}|$.

For each $1 \leq i \leq s$, the a_i be an arbitrary element of X_i. Then $a_1 \prec_K^\varphi \cdots \prec_K^\varphi a_s$, so $\emptyset \subset K'(a_1) \subset \cdots \subset K'(a_s)$. Since these inclusions are strict, we can choose $b_1, \ldots, b_s \in B$ such that $b_1 \in K'(a_1)$ and $b_{i+1} \in K'(a_{i+1}) \setminus K'(a_i)$ for $1 \leq i < s$.

It follows that $a_i \in (K')^{-1}(b_i) \setminus (K')^{-1}(b_{i+1})$, and thus $(K')^{-1}(b_i) \not\subseteq (K')^{-1}(b_{i+1})$. Since K' has the chain property, this means $(K')^{-1}(b_{i+1}) \subset (K')^{-1}(b_i)$, i.e. $b_i \sqsubset_K^\varphi b_{i+1}$.

We now have $b_1 \sqsubset_K^\varphi \cdots \sqsubset_K^\varphi b_s$; a chain of s strict inequalities in $\sqsubseteq_K^\varphi$. The corresponding ranks $[b_1], \ldots, [b_s]$ are all distinct and lie inside $\mathcal{Y}$. But now we have found $s = |\mathcal{X}|$ distinct elements of $\mathcal{Y}$, so $|\mathcal{X}| \leq |\mathcal{Y}|$ as promised.

Repeating this argument with the roles of $\mathcal{X}$ and $\mathcal{Y}$ interchanged, we find that $|\mathcal{Y}| \leq |\mathcal{X}|$ also, and therefore $|\mathcal{X}| = |\mathcal{Y}|$.

To conclude, note that $\mathsf{ranks}(\preceq_K^\varphi) \in \{|\mathcal{X}|, |\mathcal{X}| + 1\}$, since there can exist at most one rank which was excluded from $\mathcal{X}$ (namely, those $a \in A$ with $K'(a) = \emptyset$). For identical reasons, $\mathsf{ranks}(\sqsubseteq_K^\varphi) \in \{|\mathcal{Y}|, |\mathcal{Y}| + 1\}$. Since $|\mathcal{X}| = |\mathcal{Y}|$, it is clear that $\mathsf{ranks}(\preceq_K^\varphi)$ and $\mathsf{ranks}(\sqsubseteq_K^\varphi)$ can differ by at most one, as required.

Now we prove the 'if' statement. Let K be a tournament. We have $|\mathsf{ranks}(\preceq_K^\varphi) - \mathsf{ranks}(\sqsubseteq_K^\varphi)| \leq 1$, and must show there is tournament K' with the chain property such that $\varphi(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$.

Let $X_1 < \cdots < X_s$ and $Y_1 < \cdots < Y_t$ be the ranks of $\preceq_K^\varphi$ and $\sqsubseteq_K^\varphi$ respectively. By hypothesis $|s - t| \leq 1$. Define $g : \{1, \ldots, s\} \to \{0, \ldots, t\}$ by

$$g(i) = \begin{cases} i, & s \in \{t-1, t\} \\ i - 1, & s = t + 1. \end{cases}$$

Not that the two cases above cover all possibilities, since $|s-t| \leq 1$. For $i \in [s]$, write

$$N_i = \bigcup_{0 \leq j \leq g(i)} Y_j,$$

where $Y_0 := \emptyset$. Note that $g(i+1) = g(i)+1$, and consequently

$$N_{i+1} = \bigcup_{j \leq g(i)+1} Y_j = N_i \cup Y_{g(i)+1} = N_i \cup Y_{g(i+1)}.$$

Since $g(i+1) > 0$ we have $Y_{g(i+1)} \neq \emptyset$, and thus $N_{i+1} \supset N_i$ for all $i < s$.

Now, for any $a \in A$, let $p(a) \in [s]$ be the unique integer such that $a \in X_{p(a)}$; such $p(a)$ always exists since $\{X_1, \dots, X_s\}$ is a partition of A. Note that due to the assumption on the ordering of the X_i, we have $a \preceq_K^{\varphi} a'$ if and only if $p(a) \leq p(a')$.

Let K' be the unique tournament such that $K'(a) = N_{p(a)}$ for each $a \in A$. Since $N_1 \subset \cdots \subset N_p$, we have

$$\begin{aligned} a \preceq_K^{\varphi} a' &\iff p(a) \leq p(a') \\ &\iff N_{p(a)} \subseteq N_{p(a')} \\ &\iff K'(a) \subseteq K'(a') \\ &\iff a \leqslant_{K'}^{\mathcal{A}} a', \end{aligned} \tag{5.8}$$

i.e. $\preceq_K^{\varphi} = \leqslant_{K'}^{\mathcal{A}}$. Since $\preceq_K^{\varphi}$ is a total preorder, this shows that K' has the chain property.

It only remains to show that $\sqsubseteq_K^{\varphi} = \leqslant_{K'}^{\mathcal{B}}$. First note that if $a \in X_i$ and $b \in Y_j$, the fact that $\{Y_1, \dots, Y_t\}$ are disjoint implies

$$\begin{aligned} a \in (K')^{-1}(b) &\iff b \in K'(a) = N_i = \bigcup_{0 \leq k \leq g(i)} Y_k \\ &\iff j \leq g(i). \end{aligned}$$

Hence $(K')^{-1}(b)$ only depends on j: every $b \in Y_j$ shares the same neighbourhood M_j, given by

$$M_j = \bigcup_{i \in [s]:\ g(i) \geq j} X_i.$$

Note that if $1 \leq j < t$,

$$\begin{aligned} M_j &= \bigcup_{i \in [s]:\ g(i) \geq j} X_i \\ &= \left(\bigcup_{i \in [s]:\ g(i) \geq j+1} X_i \right) \cup \left(\bigcup_{i \in g^{-1}(j)} X_i \right) \\ &= M_{j+1} \cup \bigcup_{i \in g^{-1}(j)} X_i. \end{aligned}$$

Since $1 \leq j < t$ we have

$$g^{-1}(j) = \begin{cases} \{j\}, & s \in \{t-1, t\} \\ \{j+1\}, & s = t+1. \end{cases}$$

In particular $g^{-1}(j) \neq \emptyset$, which means $\bigcup_{i \in g^{-1}(j)} X_i \neq \emptyset$ and thus $M_j \supset M_{j+1}$ for all $1 \leq j < t$.

Finally, since $(K')^{-1}(b) = M_j$ for $b \in Y_j$ and $M_1 \supset \cdots \supset M_t$, an argument almost identical to (5.8) shows that $\sqsubseteq_K^{\varphi} = \leqslant_{K'}^{\mathcal{B}}$.

We have shown that $\varphi(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$ and that K' has the chain property, and the proof is therefore complete. □

5.6.2 Interleaving Operators

According to Theorem 5.6.1, to construct a chain-definable operator it is enough to ensure that the number of ranks of $\preceq_K^{\varphi}$ and $\sqsubseteq_K^{\varphi}$ differ by at most one. A simple way to achieve this is to iteratively select and remove the top-ranked players of A and B simultaneously, until one of A or B is exhausted. We call such operators *interleaving operators*. Closely related ranking methods have been previously introduced for non-bipartite tournaments by Bouyssou [14].

Formally, our procedure is defined by two functions f and g which select the next top ranks given a tournament K and subsets $A' \subseteq A$, $B' \subseteq B$ of the remaining players.

Definition 5.6.1. *An $\mathcal{A}$-*selection function *is a mapping $f : \mathcal{K} \times 2^{\mathbb{N}} \times 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ such that for any tournament K, $A' \subseteq A$ and $B' \subseteq B$:*

1. *$f(K, A', B') \subseteq A'$;*
2. *If $A' \neq \emptyset$ then $f(K, A', B') \neq \emptyset$;*
3. *$f(K, A', \emptyset) = A'$*

*Similarly, a $\mathcal{B}$-*selection function *is a mapping $g : \mathcal{K} \times 2^{\mathbb{N}} \times 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ such that*

1. *$g(K, A', B') \subseteq B'$;*
2. *If $B' \neq \emptyset$ then $g(K, A', B') \neq \emptyset$;*
3. *$g(K, \emptyset, B') = B'$*

The corresponding interleaving operator ranks players according to how soon they are selected in this way; the earlier the better.

Definition 5.6.2. *Let f and g be selection functions and K a tournament. Write $A_0 = A$, $B_0 = B$, and for $i \geq 0$:*

$$A_{i+1} = A_i \setminus f(K, A_i, B_i); \quad B_{i+1} = B_i \setminus g(K, A_i, B_i)$$

For $a \in A$ and $b \in B$, write $r(a) = \max\{i \mid a \in A_i\}$ and $s(b) = \max\{i \mid b \in B_i\}$. We define the corresponding interleaving operator $\varphi = \varphi^{\mathsf{int}}_{f,g}$ *by $a \preceq^{\varphi}_K a'$ iff $r(a) \geq r(a')$ and $b \sqsubseteq^{\varphi}_K b'$ iff $s(b) \geq s(b')$.*

Note that A_i and B_i are the players left remaining after i applications of f and g, i.e. after removing the top i ranks from both sides. Taking the maximum index in the definition of r and s is justified by the following result, which shows the interleaving process eventually terminates with $A_i = B_i = \emptyset$. Since $A_{i+1} \subseteq A_i$ and $B_{i+1} \subseteq B_i$, this shows r and s are well-defined.

Proposition 5.6.1. *Let f and g be selection functions. Fix a tournament K and let A_i, B_i ($i \geq 0$) be as in Definition 5.6.2. Then there are $j, j' \geq 1$ such that $A_j = \emptyset$ and $B_{j'} = \emptyset$. Moreover, there is $t \geq 1$ such that both $A_t = B_t = \emptyset$.*

Proof. Suppose $i \geq 0$ and $A_i \neq \emptyset$. Then properties 1 and 2 for f in Definition 5.6.1 imply that $\emptyset \subset f(K, A_i, B_i) \subseteq A_i$, and consequently $A_{i+1} = A_i \setminus f(K, A_i, B_i) \subset A_i$.

Supposing that $A_j \neq \emptyset$ for all $j \geq 0$, we would have $A_0 \supset A_1 \supset A_2 \supset \cdots$ which clearly cannot be the case since each A_j lies inside A which is a finite set. Hence there is $j \geq 1$ such that $A_j = \emptyset$. Moreover, since $A_j \supseteq A_{j+1} \supseteq A_{j+2} \supseteq \cdots$, we have $A_k = \emptyset$ for all $k \geq j$.

An identical argument with g shows that there is $j' \geq 1$ such that $B_{j'} = \emptyset$ and $B_k = \emptyset$ for all $k \geq j'$.

Taking $t = \max\{j, j'\}$, we have $A_t = B_t = \emptyset$ as required. □

Before giving a concrete example of an interleaving operator, we note that interleaving is not just *one* way to satisfying **chain-def**, it is the *only* way.

Theorem 5.6.2. *An operator φ satisfies **chain-def** if and only if $\varphi = \varphi^{\mathsf{int}}_{f,g}$ for some selection functions f, g.*

Theorem 5.6.2 justifies our study of interleaving operators, and provides a different perspective on chain-definability via the selection functions f and g.

Proof. Throughout the proof we will refer to a pair of total preorders $(\preceq, \sqsubseteq)$ as 'chain-definable' if there is a chain tournament K such that $\preceq = \leqslant^{\mathcal{A}}_K$ and $\sqsubseteq = \leqslant^{\mathcal{B}}_K$.

First we prove the 'if' direction. Let $\varphi = \varphi^{\mathsf{int}}_{f,g}$ be an interleaving operator with selection functions f, g, and fix a tournament K. We will show that $\varphi(K)$ is chain-definable.

As per Proposition 5.6.1, let $j, j' \geq 1$ be the minimal integers such that $A_j = \emptyset$ and $B_{j'} = \emptyset$. Then we have $A_0 \supset \cdots \supset A_{j-1} \supset A_j = \emptyset$ and $B_0 \supset \cdots \supset B_{j'-1} \supset B_{j'} = \emptyset$.

Recall that, for $a \in A$, we have by definition $r(a) = \max\{i \mid a \in A_i\}$, which is the unique integer such that $a \in A_{r(a)} \setminus A_{r(a)+1}$. Since $a \preceq_K^\varphi a'$ iff $r(a) \geq r(a')$, it follows that the non-empty sets $A_0 \setminus A_1, \dots, A_{j-1} \setminus A_j$ form the ranks of the total preorder $\preceq_K^\varphi$ (that is, the equivalence classes of the symmetric closure $\approx_K^\varphi$). Thus, $\preceq_K^\varphi$ has j ranks. An identical argument shows that $\sqsubseteq_K^\varphi$ has j' ranks.

It follows from Theorem 5.6.1 that $\varphi(K)$ is chain-definable if and only if $|j - j'| \leq 1$. If $j = j'$ this is clear. Suppose $j < j'$. Then $A_j = \emptyset$ and $B_j \neq \emptyset$. By property 3 for g in Definition 5.6.1, we have $g(K, A_j, B_j) = g(K, \emptyset, B_j) = B_j$. But this means $B_{j+1} = B_j \setminus g(K, A_j, B_j) = B_j \setminus B_j = \emptyset$. Consequently $j' = j + 1$, and $|j - j'| = |-1| = 1$

If instead $j > j'$, then a similar argument using property 3 for f in Definition 5.6.1 shows that $j = j' + 1$, and we have $|j - j'| = |1| = 1$.

Hence $|j - j'| \leq 1$ in all cases, and $\varphi(K)$ is chain-definable as required.

Now for the 'only if' direction. Suppose φ satisfies **chain-def**. We will define f, g such that $\varphi = \varphi_{f,g}^{\mathsf{int}}$. The idea behind the construction is straightforward: since f and g pick off the next-top-ranked As and Bs at each iteration, simply define $f(K, A_i, B_i)$ as the maximal elements of A_i with respect to the existing ordering $\preceq_K^\varphi$ (g will be defined similarly). The interleaving algorithm will then select the ranks of $\preceq_K^\varphi$ and $\sqsubseteq_K^\varphi$ one-by-one; the fact that $\varphi(K)$ is chain-definable ensures that we select *all* the ranks before the iterative procedure ends. The formal details follow.

Fix a tournament K. By Theorem 5.6.1, $|\mathsf{ranks}(\preceq_K^\varphi) - \mathsf{ranks}(\sqsubseteq_K^\varphi)| \leq 1$. Taking $t = \max\{\mathsf{ranks}(\preceq_K^\varphi), \mathsf{ranks}(\sqsubseteq_K^\varphi)\}$, we can write $X_1, \dots, X_t \subseteq A$ and $Y_1, \dots, Y_t \subseteq B$ for the ranks of $\preceq_K^\varphi$ and $\sqsubseteq_K^\varphi$ respectively, possibly with $X_1 = \emptyset$ if $\mathsf{ranks}(\sqsubseteq_K^\varphi) = 1 + \mathsf{ranks}(\preceq_K^\varphi)$ or $Y_1 = \emptyset$ if $\mathsf{ranks}(\preceq_K^\varphi) = 1 + \mathsf{ranks}(\sqsubseteq_K^\varphi)$. Note that $X_i, Y_i \neq \emptyset$ for $i > 1$. Assume these sets are ordered such that $a \preceq_K^\varphi a'$ iff $i \leq j$ whenever $a \in X_i$ and $a' \in X_j$ (and similar for the Y_i). Also note that the $X_i \cap X_j = \emptyset$ for $i \neq j$ (and similar for the Y_i).

Now set[16]

$$f(K, A', B') = \begin{cases} \max(A', \preceq_K^\varphi), & B' \neq \emptyset \\ A', & B' = \emptyset \end{cases}$$

$$g(K, A', B') = \begin{cases} \max(B', \sqsubseteq_K^\varphi), & A' \neq \emptyset \\ B', & A' = \emptyset \end{cases}$$

It is not difficult to see that f and g satisfy the conditions of Definition 5.6.1 for selection functions. We claim that for with A_i, B_i denoting the interleaving sets for

[16] Here $\max(Z, \preceq) = \{z \in Z \mid \nexists z' \in Z : z \prec z'\}$, for any set Z and a total preorder $\preceq$ on Z (with strict part $\prec$).

K and f, g, for all $0 \leq i \leq t$ we have

$$A_i = \bigcup_{j=1}^{t-i} X_j, \quad B_i = \bigcup_{j=1}^{t-i} Y_j \tag{5.9}$$

For $i = 0$ this is clear: since $X_1, \ldots, X_t$ contains all ranks of $\preceq_K^{\varphi}$ we have $\bigcup_{j=1}^{t-0} = X_1 \cup \cdots \cup X_t = A = A_0$ (and similar for B).

Now suppose (5.9) holds for some $0 \leq i < t$. We will show that $f(K, A_i, B_i) = X_{t-i}$ by considering three possible cases, at least one of which must hold.

Case 1: ($A_i \neq \emptyset$, $B_i \neq \emptyset$). Here we have

$$\begin{aligned} f(K, A_i, B_i) &= \max(A_i, \preceq_K^{\varphi}) \\ &= \max(X_1 \cup \cdots \cup X_{t-i}, \preceq_K^{\varphi}) \\ &= X_{t-i} \end{aligned}$$

since the X_j form (disjoint) ranks of $\preceq_K^{\varphi}$ with $X_j \prec X_k$ for $j < k$.

Case 2: ($B_i = \emptyset$). Here we have $\bigcup_{j=1}^{t-i} Y_j = \emptyset$. Since $t - i \geq 1$ and $Y_j \neq \emptyset$ for $j > 1$, it must be the case that $t - i = 1$ and $B_i = Y_1 = \emptyset$. Consequently by the induction hypothesis we have $A_i = \bigcup_{j=1}^{1} X_j = X_1$, and thus

$$\begin{aligned} f(K, A_i, B_i) &= f(K, A_i, \emptyset) \\ &= A_i \\ &= X_1 \\ &= X_{t-i} \end{aligned}$$

Case 3: ($A_i = \emptyset$). By a similar argument as in case 2, we must have $t - i = 1$ and $A_i = X_1 = \emptyset$. Using the fact that $f(K, A_i, B_i) \subseteq A_i$ we get

$$\begin{aligned} f(K, A_i, B_i) &= \underbrace{f(K, \emptyset, B_i)}_{\subseteq \emptyset} \\ &= \emptyset \\ &= X_1 \\ &= X_{t-i} \end{aligned}$$

We have now covered all cases, and have shown that $f(K, A_i, B_i) = X_{t-i}$ must hold. Consequently, using again the fact that the X_j are disjoint,

$$\begin{aligned} A_{i+1} &= A_i \setminus f(K, A_i, B_i) \\ &= \left(\bigcup_{j=1}^{t-i} X_j \right) \setminus X_{t-i} \\ &= \bigcup_{j=1}^{t-(i+1)} X_j \end{aligned}$$

as required. By almost identical arguments we can show that $g(K, A_i, B_i) = Y_{t-i}$, and thus $B_{i+1} = \bigcup_{j=1}^{t-(i+1)} Y_j$ also. By induction, (5.9) holds for all $0 \leq i \leq t$.

It remains to show that $a \preceq_K^{\varphi} a'$ iff $a \preceq_K^{\varphi_{f,g}^{\mathsf{int}}} a'$ and that $b \sqsubseteq_K^{\varphi} b'$ iff $b \sqsubseteq_K^{\varphi_{f,g}^{\mathsf{int}}} b'$.

For $a \in A$, let $p(a)$ be the unique integer such that $a \in X_{p(a)}$, i.e. $p(a)$ is the index of the rank of a in the ordering $\preceq_K^{\varphi}$. Note that we have

$$a \in A_i = X_1 \cup \cdots \cup X_{t-i} \iff t - i \geq p(a)$$

and therefore

$$r(a) = \max\{i \mid a \in A_i\} = \max\{i \mid t - i \geq p(a)\} = t - p(a)$$

Using the fact that $X_i \prec X_j$ for $i < j$, we get

$$\begin{aligned} a \preceq_K^{\varphi_{f,g}^{\mathsf{int}}} a' &\iff r(a) \geq r(a') \\ &\iff t - p(a) \geq t - p(a') \\ &\iff p(a) \leq p(a') \\ &\iff a \preceq_K^{\varphi} a' \end{aligned}$$

A similar argument shows that $b \sqsubseteq_K^{\varphi} b'$ iff $b \sqsubseteq_K^{\varphi_{f,g}^{\mathsf{int}}} b'$ for any $b, b' \in B$. Since K was arbitrary, we have shown that $\varphi = \varphi_{f,g}^{\mathsf{int}}$ as required. □

We now come to an important example.

Example 5.6.1. *Define the* cardinality-based interleaving operator $\varphi_{\mathsf{CI}} = \varphi_{f,g}^{\mathsf{int}}$ *where* $f(K, A', B') = \arg\max_{a \in A'} |K(a) \cap B'|$ *and* $g(K, A', B') = \arg\min_{b \in B'} |K^{-1}(b) \cap A'|$*, so that the 'winners' at each iteration are the As with the most wins, and the Bs with the least losses, when restricting to A' and B' only. We take the* $\arg\min/\arg\max$ *to be the emptyset whenever A' or B' is empty.*

Table 5.1 shows the iteration of the algorithm for a 4×5 tournament K. In each row i we show K with the rows and columns of $A \setminus A_i$ and $B \setminus B_i$ greyed out, so as to make it more clear how the f and g values are calculated.[17] *For brevity we also write f and g in place of $f(K, A_i, B_i)$ and $g(K, A_i, B_i)$ respectively.*

The r and s values can be read off as 0, 2, 1, 3 for A and 0, 3, 1, 1, 2 for B, giving the ranking on A as $4 \prec 2 \prec 3 \prec 1$, and the ranking on B as $2 \sqsubset 5 \sqsubset 3 \approx 4 \sqsubset 1$. Note also that each $f(K, A_i, B_i)$ is a rank of $\preceq_K^{\varphi}$ (and similar for $g(K, A_i, B_i)$), so the rankings can in fact be read off by looking at the f and g columns of Table 5.1.

The interleaving algorithm can also be seen as a greedy algorithm for converting K into a chain graph directly. Indeed, by setting the neighbourhood of each $a \in$

[17] Note that while f and g for φ_{CI} are independent of the greyed out entries, we do not require this property for selection functions in general.

Table 5.1: Iteration of the interleaving algorithm for φ_{CI}

i	K	A_i	B_i	f	g	K'_i
0	$\begin{bmatrix} 1&1&1&1&0 \\ 0&1&0&0&1 \\ 0&1&0&1&1 \\ 0&1&1&0&0 \end{bmatrix}$	$\{1,2,3,4\}$	$\{1,2,3,4,5\}$	$\{1\}$	$\{1\}$	$\begin{bmatrix} 1&1&1&1&1 \\ 0&1&0&0&1 \\ 0&1&0&1&1 \\ 0&1&1&0&0 \end{bmatrix}$
1	$\begin{bmatrix} 1&1&1&1&0 \\ 0&1&0&0&1 \\ 0&1&0&1&1 \\ 0&1&1&0&0 \end{bmatrix}$	$\{2,3,4\}$	$\{2,3,4,5\}$	$\{3\}$	$\{3,4\}$	$\begin{bmatrix} 1&1&1&1&1 \\ 0&1&0&0&1 \\ 0&1&1&1&1 \\ 0&1&0&0&0 \end{bmatrix}$
2	$\begin{bmatrix} 1&1&1&1&0 \\ 0&1&0&0&1 \\ 0&1&0&1&1 \\ 0&1&1&0&0 \end{bmatrix}$	$\{2,4\}$	$\{2,5\}$	$\{2\}$	$\{5\}$	-
3	$\begin{bmatrix} 1&1&1&1&0 \\ 0&1&0&0&1 \\ 0&1&0&1&1 \\ 0&1&1&0&0 \end{bmatrix}$	$\{4\}$	$\{2\}$	$\{4\}$	$\{2\}$	-
4	-	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	-

$f(K, A_i, B_i)$ to B_i, and removing each $b \in g(K, A_i, B_i)$ from the neighbourhoods of all $a \in A_{i+1}$, we eventually obtain a chain graph. We show this process in the K'_i column of Table 5.1, where only three entries need to be changed.[18] The selection functions f and g can therefore be seen as *heuristics* with the goal of finding a chain graph 'close' to K.

The operator φ_{CI} from Example 5.6.1 uses simple cardinality-based heuristics, and can be seen as a chain-definable version of φ_{count} (which is not chain-definable). It is also the bipartite counterpart to repeated applications of Copeland's rule [14]. Note that $f(K, A_i, B_i)$ and $g(K, A_i, B_i)$ can be computed in $O(N^2)$ time at each iteration i, where $N = |A| + |B|$. Since there cannot be more than N iterations, it follows that the rankings of φ_{CI} can be computed in $O(N^3)$ time.

5.6.3 Axiom Compatibility

We now revisit the axioms of Section 5.4 in relation to chain-definable operators in general and φ_{CI} specifically. Firstly, the weakening of **chain-min** pays off: **chain-def** is compatible with all our axioms.

Theorem 5.6.3. *For each of **anon**, **dual**, **IIM**, **mon** and **pos-resp**, there exists an operator satisfying **chain-def** and the stated property.*

Proof. Since **chain-min** implies **chain-def**, Theorem 5.4.2 implies the existence of an operator with **chain-def** and **dual**, and an operator with **chain-def** and **mon**. Moreover, the trivial operator which ranks all As and Bs equally satisfies **chain-def**, **anon** and **IIM**. It only remains to show that there is an operator satisfying both **chain-def** and **pos-resp**.

To that end, for any tournament K, define K' by

$$K'_{ab} = \begin{cases} 1, & b \leq |K(a)| \\ 0, & b > |K(a)| \end{cases}$$

[18] In this example $\mathcal{M}(K)$ contains a single tournament a distance of 2 from K, so φ_{CI} makes one more change than necessary.

Note that $K'(a) = \{1, \ldots, |K(a)|\}$ for $|K(a)| > 0$. Consequently $K'(a) \subseteq K'(a')$ iff $|K(a)| \leq |K(a)|$. We see that K' has the chain property, and the operator φ defined by $\varphi(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$ satisfies **chain-def**. In particular, $a \preceq^\varphi_K a'$ iff $|K(a)| \leq |K(a')|$.

To show **pos-resp**, suppose $a \preceq^\varphi_K a'$ and $K_{a',b} = 0$ for some $a, a' \in A$ and $b \in B$. Write $\hat{K} = K + \mathbf{1}_{a',b}$.

Since $a \preceq^\varphi_K a'$ implies $|K(a)| \leq |K(a')|$, we have $|\hat{K}(a')| = 1 + |K(a')| > |K(a)| = |\hat{K}(a)|$, and therefore $a \prec^\varphi_{\hat{K}} a'$ as required for **pos-resp**. □

Unfortunately, these cannot all hold at the same time. Indeed, taking $K = \left[\begin{smallmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{smallmatrix}\right]^\top$ and assuming **anon** and **pos-resp**, the ranking on A is fully determined as $1 \prec 2 \approx 3 \prec 4$, and $\mathsf{ranks}(\preceq^\varphi_K) = 3$. However, **anon** with **dual** implies the ranking of B is flat, i.e. $\mathsf{ranks}(\sqsubseteq^\varphi_K) = 1$. This contradicts **chain-def** by Theorem 5.6.1, yielding the following impossibility result.

Theorem 5.6.4. *There is no operator satisfying **chain-def**, **anon**, **dual** and **pos-resp**.*

Proof. For contradiction, suppose there is an operator φ satisfying the stated axioms. Consider

$$K = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

and two tournaments obtained by removing a single 1 entry:

$$K_1 = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{0} \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & \mathbf{0} \end{bmatrix}.$$

Now, **anon** in K_1 gives $1 \approx^\varphi_{K_1} 2$ (e.g. take $\sigma = (1\ 2)$, $\pi = \mathrm{id}_B$). In particular, $1 \preceq^\varphi_{K_1} 2$, so **pos-resp** implies $1 \prec^\varphi_K 2$. A similar argument with K_2 shows that $3 \approx^\varphi_{K_2} 4$ and $3 \prec^\varphi_K 4$.

On the other hand, applying **anon** to K directly with $\sigma = (2\ 3)$ and $\pi = (1\ 2)$, we see that $2 \approx^\varphi_K 3$. The ranking of A is thus fully determined as $1 \prec 2 \approx 3 \prec 4$. In particular, $\mathsf{ranks}(\preceq^\varphi_K) = 3$.

But now considering the dual tournament $\overline{K} = \left[\begin{smallmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{smallmatrix}\right]$ and applying permutations $\sigma = (1\ 2)$ and $\pi = (2\ 3)$, we obtain $1 \approx^\varphi_{\overline{K}} 2$ by **anon**, i.e. the A ranking in $\overline{K}$ is flat. By **dual** this implies the B ranking in K is flat, i.e. $\mathsf{ranks}(\sqsubseteq^\varphi_K) = 1$. We see that $\mathsf{ranks}(\preceq^\varphi_K)$ and $\mathsf{ranks}(\sqsubseteq^\varphi_K)$ differ by 2, contradicting **chain-def** according to Theorem 5.6.1. □

For interleaving operators, we have the following sufficient conditions for $\varphi^{\mathsf{int}}_{f,g}$ to satisfy various axioms.

Lemma 5.6.1. *Let $\varphi = \varphi^{int}_{f,g}$ be an interleaving operator.*

1. *If for any tournament K, $A' \subseteq A$, $B' \subseteq B$ and for any pair of permutations $\sigma : A \to A$ and $\pi : B \to B$ we have*

$$f(\pi(\sigma(K)), \sigma(A'), \pi(B')) = \sigma(f(K, A', B'))$$
$$g(\pi(\sigma(K)), \sigma(A'), \pi(B')) = \pi(g(K, A', B'))$$

then φ satisfies ***anon****.*

2. *If for any tournament K and $A' \subseteq A$, $B \subseteq B$ we have*

$$g(K, A', B') = f(\overline{K}, B', A')$$

then φ satisfies ***dual****.*

3. *If for any tournament K, $A' \subseteq A$, $B' \subseteq B$ and $a, a' \in A'$ we have*

$$K(a) \subseteq K(a') \implies a \notin f(K, A', B') \text{ or } a' \in f(K, A', B')$$

then φ satisfies ***mon****.*

Proof. We take each statement in turn.

1. Let K be a tournament. For brevity, write $K' = \pi(\sigma(K))$. Let us write A_i, B_i and A'_i, B'_i ($i \geq 0$) for the sets defined in Definition 5.6.2 for K and K' respectively. We claim that for all $i \geq 0$:

$$A'_i = \sigma(A_i), \quad B'_i = \pi(B_i) \tag{5.10}$$

For $i = 0$ this is trivial since $A'_0 = A = \sigma(A) = \sigma(A_0)$ since σ is a bijection. The fact that $B'_0 = \pi(B_0)$ is shown similarly.

Suppose that (5.10) holds for some $i \geq 0$. Then applying our assumption on f:

$$\begin{aligned} A'_{i+1} &= A'_i \setminus f(K', A'_i, B'_i) \\ &= \sigma(A_i) \setminus f(K', \sigma(A_i), \pi(B_i)) \\ &= \sigma(A_i) \setminus \sigma(f(K, A_i, B_i)) \\ &= \sigma(A_i \setminus f(K, A_i, B_i)) \\ &= \sigma(A_{i+1}) \end{aligned}$$

(note that $\sigma(X) \setminus \sigma(Y) = \sigma(X \setminus Y)$ holds for any sets X, Y due to injectivity of σ). Using the assumption on g we can show that $B'_{i+1} = \pi(B_{i+1})$ in a similar manner. Therefore, by induction, (5.10) holds for all $i \geq 0$. This means that for any $a \in A$ we have

$$\sigma(a) \in A'_i \iff \sigma(a) \in \sigma(A_i) \iff a \in A_i$$

and therefore, with r_K and $r_{K'}$ denoting the functions $A \to \mathbb{N}_0$ defined in Definition 5.6.2 for K and K' respectively,

$$\begin{aligned} r_{K'}(\sigma(a)) &= \max\{i \mid \sigma(a) \in A'_i\} \\ &= \max\{i \mid a \in A_i\} \\ &= r_K(a) \end{aligned}$$

From this it easily follows that $a \preceq^{\varphi}_K a'$ iff $\sigma(a) \preceq^{\varphi}_{K'} \sigma(a')$, i.e. φ satisfies **anon**.

2. Once again, fix a tournament K and let A_i, B_i and A'_i, B'_i denote the sets from Definition 5.6.2 for K and $\overline{K}$ respectively. It is easy to show by induction that the assumption on f and g implies $A'_i = B_i$ and $B'_i = A_i$ for all $i \geq 0$. This means that for any $b \in B_K$:

$$\begin{aligned} s_K(b) &= \max\{i \mid b \in B_i\} \\ &= \max\{i \mid b \in A'_i\} \\ &= r_{\overline{K}}(b) \end{aligned}$$

which implies $b \sqsubseteq^{\varphi}_K b'$ iff $b \preceq^{\varphi}_{\overline{K}} b'$, as required for **dual**.

3. Let K be a tournament and $a, a' \in A$ such that $K(a) \subseteq K(a')$. We must show that $a \preceq^{\varphi}_K a'$.

   Suppose otherwise, i.e. $a' \prec^{\varphi}_K a$. Then $r(a') > r(a)$. Note that by definition of r, we have $a \in A_{r(a)} \setminus A_{r(a)+1} = f(K, A_{r(a)}, B_{r(a)})$. Since $r(a') \geq r(a) + 1$ and $A_{r(a)} \supseteq A_{r(a)+1} \supseteq A_{r(a)+2} \supseteq \cdots$, we get $a' \in A_{r(a)+1} \subseteq A_{r(a)}$. In particular, $a' \notin f(K, A_{r(a)}, B_{r(a)})$.

   Piecing this all together, we have $a, a' \in A_{r(a)}$, $K(a) \subseteq K(a')$, $a \in f(K, A_{r(a)}, B_{r(a)})$ and $a' \notin f(K, A_{r(a)}, B_{r(a)})$. But this directly contradicts our assumption on f, so we are done.

□

For the specific operator φ_{CI}, Lemma 5.6.1 yields the following.

Theorem 5.6.5. *φ_{CI} satisfies **chain-def**, **anon**, **dual** and **mon**, and does not satisfy **IIM** or **pos-resp**.*

Proof. We take each axiom in turn. Let f and g be the selection functions corresponding to φ_{CI} from Example 5.6.1.

chain-def. Since φ_{CI} is an interleaving operator, **chain-def** follows from Theorem 5.6.2.

anon. Let K be a tournament and let $\sigma : A \to A$ and $\pi : B \to B$ be bijective mappings. Write $K' = \pi(\sigma(K))$. We will show that the conditions on f and g in Lemma 5.6.1 part (1) are satisfied.

Let $A' \subseteq A$ and $B' \subseteq B$. We have

$$\begin{aligned} f(K', \sigma(A'), \pi(B')) &= \underset{\hat{a} \in \sigma(A')}{\arg\max} |K'(\hat{a}) \cap \pi(B')| \\ &= \sigma(\underset{a \in A'}{\arg\max} |K'(\sigma(a)) \cap \pi(B')|) \end{aligned}$$

where we make the 'substitution' $a = \sigma^{-1}(\hat{a})$. Using the defintion of $K' = \pi(\sigma(K))$ it is easily seen that $K'(\sigma(a)) = \pi(K(a))$. Also, since π is a bijection we have $\pi(X) \cap \pi(Y) = \pi(X \cap Y)$ for any sets X and Y, and $|\pi(X)| = |X|$. Thus

$$\begin{aligned} f(K', \sigma(A'), \pi(B')) &= \sigma(\underset{a \in A'}{\arg\max} |K'(\sigma(a)) \cap \pi(B')|) \\ &= \sigma(\underset{a \in A'}{\arg\max} |\pi(K(a)) \cap \pi(B')|) \\ &= \sigma(\underset{a \in A'}{\arg\max} |\pi(K(a) \cap B')|) \\ &= \sigma(\underset{a \in A'}{\arg\max} |K(a) \cap B'|) \\ &= \sigma(f(K, A', B')) \end{aligned}$$

as required. The result for g follows by a near-identical argument. Thus φ_{CI} satisfies **anon** by Lemma 5.6.1 part (1).

dual. Fix a tournament K and let $A' \subseteq A$, $B' \subseteq B$. Note that for $b \in B'$ we have

$$\begin{aligned} |K^{-1}(b) \cap A'| &= |(A \setminus \overline{K}(b)) \cap A'| \\ &= |A' \setminus \overline{K}(b)| \\ &= |A'| - |\overline{K}(b) \cap A'| \end{aligned}$$

Consequently

$$\begin{aligned} g(K, A', B') &= \underset{b \in B'}{\arg\min} |K^{-1}(b) \cap A'| \\ &= \underset{b \in B'}{\arg\min} \left(|A'| - |\overline{K}(b) \cap A'|\right) \\ &= \underset{b \in B'}{\arg\max} |\overline{K}(b) \cap A'| \\ &= f(\overline{K}, B', A') \end{aligned}$$

and, by Lemma 5.6.1 part (2), φ_{CI} satisfies **dual**.

mon. Once again, we use Lemma 5.6.1. Let K be a tournament and $A' \subseteq A$, $B' \subseteq B$. Suppose $a, a' \in A'$ with $K(a) \subseteq K(a')$. We need to show that either $a \notin f(K, A', B')$ or $a' \in f(K, A', B')$

Suppose $a \in f(K, A', B')$. Then $a \in \arg\max_{\hat{a} \in A'} |K(\hat{a}) \cap B'|$, so $|K(a) \cap B'| \geq |K(a') \cap B'|$. On the other hand $K(a) \cap B' \subseteq K(a') \cap B'$, so $|K(a) \cap B'| \leq |K(a') \cap B'|$. Consequently $|K(a) \cap B'| = |K(a') \cap B'|$, and so $a' \in f(K, A', B')$. This shows the property required by Lemma 5.6.1 part (3) is satisfied, and thus φ_{CI} satisfies **mon**.

pos-resp. We have show that φ_{CI} satisfies **chain-def**, **anon** and **dual**; due to impossibility result of Theorem 5.6.4, φ_{CI} cannot satisfy **pos-resp**.

IIM. Write

$$K_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Note that the first and second rows of each tournament are identical, so **IIM** would imply $1 \preceq^{\varphi_{\mathsf{CI}}}_{K_1} 2$ iff $1 \preceq^{\varphi_{\mathsf{CI}}}_{K_2} 2$. However, it is easily verified that $1 \prec^{\varphi_{\mathsf{CI}}}_{K_1} 2$ whereas $2 \prec^{\varphi_{\mathsf{CI}}}_{K_2} 1$. Therefore φ_{CI} does not satisfy **IIM**. □

Note that **anon** *is* satisfied. This makes φ_{CI} an important example of a well-motivated, tractable, chain-definable and anonymous operator, meeting the criteria outlined at the start of this section.

5.6.4 Axiomatic Characterisation of φ_{CI}

In Theorem 5.6.5 we saw which of the axioms from Section 5.4 hold for φ_{CI}. We now characterise φ_{CI} axiomatically by introducing two new axioms. The first, which we call **rank-removal**, is a technical axiom obtained via a related property specific to interleaving operators. The second, called **argmax**, says that the maximum rank in $\preceq^{\varphi}_{K}$ should coincide with that of $\preceq^{\varphi_{\mathsf{CI}}}_{K}$. Together with **dual** and **chain-def**, these will characterise φ_{CI}.

Unlike the axioms introduced so far, which were straightforward, general properties for ranking methods, the new axioms are geared specifically towards characterising φ_{CI}. Thus, this section takes a *descriptive* perspective as opposed to the *normative* perspective of Section 5.4.

Towards the characterisation result, we first note that φ_{CI} satisfies a kind of independence property for interleaving operators: $f(K, A', B')$ and $g(K, A', B')$ only depends on the sub-matrix of K with rows and columns corresponding to A' and B'. In graphical terms, the greyed out rows and columns in Table 5.1 do not affect the output of f and g. In general this does not hold for interleaving operators. We express this formally as an axiom for interleaving operators $\varphi^{\mathsf{int}}_{f,g}$, called *sub-matrix independence*.

Axiom 5.6.2 (SMI). *Let K be a tournament and A_i, B_i a pair of non-empty sets arising in the interleaving algorithm for f, g and K in Definition 5.6.2. Write $A_i = \{a_1, \ldots, a_{m'}\}$ and $B_i = \{b_1, \ldots, b_{n'}\}$, ordered such that $a_p < a_{p+1}$ and $b_q < b_{q+1}$. Let K^- be the* *corresponding $m' \times n'$ sub-matrix of K, where $K^-_{pq} = K_{a_p,b_q}$. Then for all $a_p \in A_i$ and $b_q \in B_i$,*

$$\begin{aligned} a_p \in f(K, A_i, B_i) &\iff p \in f(K^-, [m'], [n']), \\ b_q \in g(K, A_i, B_i) &\iff q \in g(K^-, [m'], [n']). \end{aligned}$$

Note that **SMI** is a property of the selection functions f and g. In principle, it is possible that an interleaving operator φ admits two pairs of selection functions (f, g) and (f', g') such that **SMI** holds for one pair but not the other. However, it will follow from later results (Proposition 5.6.4) that this is not possible: **SMI** either holds for φ or does not, independently of the choice of selection functions f and g. Nevertheless, to avoid circularity we will say φ satisfies **SMI** if *there exist* f, g with the **SMI** property such that $\varphi = \varphi^{\mathsf{int}}_{f,g}$.

First, φ_{CI} does indeed satisfy **SMI**.

Proposition 5.6.2. *φ_{CI} satisfies SMI.*

Proof. Let f and g denote the selection functions for φ_{CI}. Let A_i, B_i and K^- be as in the statement of **SMI**. Note that for any $a_p \in A_i$,

$$\begin{aligned} K(a_p) \cap B_i &= \{b \in B_i \mid K_{a_p,b} = 1\} \\ &= \{b_q \mid q \in [n'], K_{a_p,b_q} = 1\} \\ &= \{b_q \mid q \in K^-(p)\}, \end{aligned}$$

so $|K(a_p) \cap B_i| = |K^-(p)| = |K^-(p) \cap [n']|$. Consequently,

$$\begin{aligned} a_p \in f(K, A_i, B_i) &\iff a_p \in \arg\max_{a \in A_i} |K(a) \cap B_i| \\ &\iff p \in \arg\max_{p' \in [m']} |K(a_{p'}) \cap B_i| \\ &\iff p \in \arg\max_{p' \in [m']} |K^-(p)| \\ &\iff p \in f(K^-, [m'], [n']) \end{aligned}$$

as required. An identical argument shows the desired property for g. Hence φ_{CI} satisfies **SMI**. □

Note that in the statement of **SMI**, $[m'] = A_{K^-}$ and $[n'] = B_{K^-}$. Consequently, **SMI** implies that the ranks of A and B are fully determined by the *maximal* ranks of successively smaller sub-tournaments. This is expressed in the following result, which shows that two **SMI** operators agreeing on maximal ranks for all K must in fact be equal.

Proposition 5.6.3. *Let φ and φ' be interleaving operators satisfying **SMI**. Suppose that for all tournaments K,*

$$\max(A, \preceq_K^{\varphi}) = \max(A, \preceq_K^{\varphi'})$$
$$\max(B, \sqsubseteq_K^{\varphi}) = \max(B, \sqsubseteq_K^{\varphi'}).$$

Then $\varphi = \varphi'$.

Proof. Let f, g and f', g' be selection functions corresponding to φ and φ' respectively. Take a tournament K. To show $\varphi(K) = \varphi(K')$ it is sufficient to show $A_i = A_i'$ and $B_i = B_i'$ for all $i \geq 0$, where A_i, B_i and A_i', B_i' are the interleaving sets from Definition 5.6.2 for φ and φ' respectively.

We proceed by induction on i. For $i = 0$ this is clear, since $A_0 = A_0' = A$ and $B_0 = B_0' = B$ by definition. Suppose $A_i = A_i'$ and $B_i = B_i'$. If $A_i = A_i' = \emptyset$ then $A_{i+1} = A_{i+1}' = \emptyset$ (since $A_{i+1} \subseteq A_i$), and similarly $B_i = B_i' = \emptyset$ implies $B_{i+1} = B_{i+1}' = \emptyset$. Hence we may assume without loss of generality that $A_i, B_i \neq \emptyset$.

Write $A_i = A_i' = \{a_1, \ldots, a_{m'}\}$ and $B_i = B_i' = \{b_1, \ldots, b_{n'}\}$, with $a_p < a_{p+1}$ and $b_q < b_{q+1}$. Let K^- be the associated sub-matrix, as in the statement of **SMI**. From property 1 from Definition 5.6.1 for f and **SMI**, we have

$$\begin{aligned} f(K, A_i, B_i) &= \{a_p \mid p \in f(K^-, [m'], [n'])\} \\ f'(K, A_i', B_i') &= \{a_p \mid p \in f'(K^-, [m'], [n'])\}. \end{aligned} \tag{5.11}$$

But $[m']$ and $[n']$ are the full set of players in K^-, so

$$\begin{aligned} f(K^-, [m'], [n']) &= \max(A_{K^-}, \preceq_{K^-}^{\varphi}) \\ &= \max(A_{K^-}, \preceq_{K^-}^{\varphi'}) \\ &= f'(K^-, [m'], [n']), \end{aligned}$$

where our assumption on the maximal ranks for φ and φ' is employed in the second step. Consulting (5.11) we see $f(K, A_i, B_i) = f'(K, A_i', B_i')$. An identical argument shows $g(K, A_i, B_i) = g'(K, A_i', B_i')$. Together with the induction hypothesis, we get

$$A_{i+1} = A_i \setminus f(K, A_i, B_i) = A_i' \setminus f'(K, A_i', B_i') = A_{i+1}'$$

and similarly, $B_{i+1} = B_{i+1}'$. By induction, $A_i = A_i'$ and $B_i = B_i'$ for all $i \geq 0$, and we are done. $\square$

This result simplifies the task of characterising φ_{CI} among the interleaving operators with **SMI**, since we only need to consider the maximal ranks of A and B. In fact, given that φ_{CI} satisfies **dual** we only need to consider the A ranking. The following axiom says that maximally-ranked players in A are exactly those for whom $|K(a)|$ is maximal; this will clearly capture the maximal ranks of φ_{CI} together with **dual**.

Axiom 5.6.3 (argmax). $\max(A, \preceq_K^\varphi) = \arg\max_{a \in A} |K(a)|$.

Note that **argmax** does *not* require $\preceq_K^\varphi$ to reduce to φ_{count}, since we only consider a such that $|K(a)|$ is *maximal*.

Now, since **chain-def** characterises interleaving operators, to obtain a characterisation of φ_{CI} among *all* operators it suffices to find an alternative version of **SMI** which can be applied to any operator. This is the role of the following axiom, which says that removing the maximally ranked players from each side preserves the ordering among the rest of A and B.

Axiom 5.6.4 (rank-removal). *Suppose* $\max(A, \preceq_K^\varphi) \neq A$ *and* $\max(B, \sqsubseteq_K^\varphi) \neq B$*. Write* $A \setminus \max(A, \preceq_K^\varphi) = \{a_1, \dots, a_{m'}\}$ *and* $B \setminus \max(B, \sqsubseteq_K^\varphi) = \{b_1, \dots, b_{m'}\}$*, ordered such that* $a_p < a_{p+1}$ *and* $b_q < b_{q+1}$*. Let* K^- *be the corresponding* $m' \times n'$ *sub-matrix of* K*. Then for all* p, p' *and* q, q'*,*

$$a_p \preceq_K^\varphi a_{p'} \iff p \preceq_{K^-}^\varphi p'$$
$$b_q \sqsubseteq_K^\varphi b_{q'} \iff q \sqsubseteq_{K^-}^\varphi q'.$$

In what sense does **rank-removal** capture **SMI**? In the following we show it is *equivalent* to **SMI**, when taken with **chain-def**. We need a preliminary lemma.

Lemma 5.6.2. *Let* f, g *be selection functions,* K *a tournament, and* A_i, B_i *sets arising in the interleaving algorithm for* f, g *and* K*. Suppose* $A_i \neq \emptyset$ *and* $B_i \neq \emptyset$*. Then*

$$f(K, A_i, B_i) = \max(A_i, \preceq_K^{\varphi_{f,g}^{\mathsf{int}}})$$
$$g(K, A_i, B_i) = \max(B_i, \sqsubseteq_K^{\varphi_{f,g}^{\mathsf{int}}}).$$

Proof. We show the first statement; the second follows by an identical argument. For brevity, write φ for $\varphi_{f,g}^{\mathsf{int}}$. For the left-to-right inclusion, suppose $a \in f(K, A_i, B_i)$. Then $a \in A_i$. Take any $a' \in A_i$. We need to show $a' \preceq_K^\varphi a$. Indeed, we have $a \in A_i \cap f(K, A_i, B_i)$, so $a \notin A_{i+1}$. Consequently $r(a) = \max\{j \mid a \in A_j\} = i$. Since $a' \in A_i$, $r(a') \geq i = r(a)$. Hence $a' \preceq_K^\varphi a$.

For the right-to-left we show the contrapositive. Suppose $a \notin f(K, A_i, B_i)$. If $a \notin A_i$ then clearly $a \notin \max(A_i, \preceq_K^\varphi)$. So suppose $a \in A_i$. Then $a \in A_i \setminus f(K, A_i, B_i) = A_{i+1}$. Hence $r(a) \geq i + 1 > i$. On the other hand, since $A_i \neq \emptyset$ we have by properties of the selection function f that $f(K, A_i, B_i) \neq \emptyset$. Thus there is some $a' \in A_i \cap f(K, A_i, B_i)$. We see that $r(a') = i < r(a)$, so $a \prec_K^\varphi a'$. Hence $a \notin \max(A_i, \preceq_K^\varphi)$, as required. □

Proposition 5.6.4. φ *satisfies **chain-def** and **rank-removal** if and only if* φ *is an interleaving operator satisfying **SMI**.*

Proof. For the 'if" direction, suppose $\varphi = \varphi^{\mathsf{int}}_{f,g}$ is an interleaving operator satisfying **SMI**. Then φ satisfies **chain-def** by Theorem 5.6.2. We show **rank-removal**. Let $\{a_1, \dots, a_{m'}\}$, $\{b_1, \dots, b_{n'}\}$ and K^- be as in the statement of **rank-removal**.

Let A_i, B_i denote the interleaving sets for φ applied to K, and A_i^-, B_i^- those for φ applied to K^-. First we claim that for all $p \in [m']$, $q \in [n']$ and $i \geq 0$,

$$\begin{aligned} p \in A_i^- &\iff a_p \in A_{i+1}, \\ q \in B_i^- &\iff b_q \in B_{i+1}. \end{aligned} \tag{5.12}$$

We prove (5.12) by induction on i. Take $i = 0$. By definition, $A_0^- = A_{K^-} = [m']$, so $p \in A_0^-$ always holds. Recall that $\{a_1, \dots, a_{m'}\} = A \setminus \max(A, \preceq_K^{\varphi})$. It is easily seen that $\max(A, \preceq_K^{\varphi}) = f(K, A_0, B_0)$, so in fact $\{a_1, \dots, a_{m'}\} = A_1$. Hence $a_p \in A_1$ always holds. The argument for the Bs is identical. This proves (5.12) for $i = 0$.

For the inductive step, suppose (5.12) holds for some $i \geq 0$. We consider cases. First, suppose $A_i^- = \emptyset$. Then $A_{i+1}^- \subseteq A_i^-$ means $A_{i+1}^- = \emptyset$. On the other hand the inductive hypothesis gives $A_{i+1} = \emptyset$, so we get $A_{i+2} = \emptyset$. In particular, the first part of (5.12) holds for $i+1$. For the second part, using property 3 from Definition 5.6.1 for the selection function g we find

$$B_{i+1}^- = B_i^- \setminus g(K^-, \emptyset, B_i^-) = B_i^- \setminus B_i^- = \emptyset.$$

By the inductive hypothesis again we have $A_{i+1} = \emptyset$, so the same line of reasoning gives $B_{i+1} = \emptyset$. Thus (5.12) holds. The case $B_i^- = \emptyset$ is identical, using properties of f instead of g.

Now suppose both A_i^- and B_i^- are non-empty. Recall $A_i^- \subseteq A_0^- = [m']$ and $B_i^- \subseteq B_0^- = [n']$. Write $A_i^- = \{p_1, \dots, p_{m''}\}$ and $B_i^- = \{q_1, \dots, q_{n''}\}$ in increasing order. Let K^{--} be the $m'' \times n''$ sub-matrix of K^- formed by A_i^- and B_i^-, i.e. $K_{st}^{--} = K_{p_s,q_t}^-$. Since $\varphi = \varphi^{\mathsf{int}}_{f,g}$ satisfies **SMI**, we get that for all $s \in [m'']$ and $t \in [n'']$,

$$p_s \in f(K^-, A^-i, B_i^-) \iff s \in f(K^{--}, [m''], [n'']).$$

Now, recall that $K_{pq}^- = K_{a_p,b_q}$. Hence

$$K_{st}^{--} = K_{p_s,q_t}^- = K_{a_{p_s},b_{q_t}}.$$

By the inductive hypothesis,

$$\begin{aligned} A_{i+1} &= \{a_p \mid p \in A_i^-\} = \{a_{p_s} \mid s \in [m'']\}, \\ B_{i+1} &= \{b_q \mid q \in B_i^-\} = \{b_{q_t} \mid t \in [n'']\}. \end{aligned}$$

We see that K^{--} can also be viewed as the $m'' \times n''$ sub-matrix of K formed by A_{i+1} and B_{i+1}. Applying **SMI** in this instance, we find

$$a_{p_s} \in f(K, A_{i+1}, B_{i+1}) \iff s \in f(K^{--}, [m''], [n'']).$$

Putting things together,

$$a_{p_s} \in f(K, A_{i+1}, B_{i+1}) \iff p_s \in f(K^-, A_i^-, B_i^-).$$

Consequently, for any $p \in [m']$ we get

$$\begin{aligned}
p \in A_{i+1}^- &\iff p \in A_i^- \text{ and } p \notin f(K^-, A_i^-, B_i^-) \\
&\iff \exists s \in [m''] : \; p = p_s \text{ and } p_s \notin f(K^-, A_i^-, B_i^-) \\
&\iff \exists s \in [m''] : \; p = p_s \text{ and } a_{p_s} \notin f(K, A_{i+1}, B_{i+1}) \\
&\iff a_p \in A_{i+1} \text{ and } a_p \notin f(K, A_{i+1}, B_{i+1}) \\
&\iff a_p \in A_{i+2}.
\end{aligned}$$

An identical argument shows $q \in B^-i+1 \iff b_q \in B_{i+2}$. By induction, this completes the proof of (5.12).

Finally, for any $p \in [m']$ we get

$$\begin{aligned}
r_{K^-}(p) &= \max\{i \mid p \in A_i^-\} \\
&= \max\{i \mid a_p \in A_{i+1}\} \\
&= \max\{i-1 \mid a_p \in A_i\} \\
&= -1 + r_K(a_p),
\end{aligned}$$

where we use (5.12) and the fact that $a_p \in A_1$. Since the additive constant of -1 does not affect the ranking, we see that the ranking of $[m']$ in $\preceq^{\varphi}_{K^-}$ corresponds exactly to that of $\{a_1, \ldots, a_{m'}\}$ in $\preceq^{\varphi}_{K}$, as required for **rank-removal**. The case for the B ranking is identical, and we are done.

Now for the 'only if" direction, suppose φ satisfies **chain-def** and **rank-removal**. By Theorem 5.6.2 there are selection functions f and g such that $\varphi = \varphi^{\mathsf{int}}_{f,g}$. We need to show that **SMI** holds. Let K, $A_i = \{a_1, \ldots, a_{m'}\}, B_i = \{b_1, \ldots, b_{n'}\}$ and K^- be as in the statement of **SMI**. Without loss of generality, $i > 0$. A simple induction shows that

$$A_j = A \setminus \bigcup_{k<j} f(K, A_k, B_k)$$

for all $j \geq 0$. From Lemma 5.6.2 we see that A_i is the result of removing the top $i-1$ ranks of players in A, according to the ranking $\preceq^{\varphi}_{K}$. Applying **rank-removal** $i-1$ times, we get

$$a_p \preceq^{\varphi}_{K} a_{p'} \iff p \preceq^{\varphi}_{K^-} p'$$

for all $p, p' \in [m']$. Consequently, using Lemma 5.6.2 again,

$$\begin{aligned} a_p \in f(K, A_i, B_i) &\iff a_p \in \max(A_i, \preceq_K^{\varphi}) \\ &\iff \forall p' \in [m'] : \; a_{p'} \preceq_K^{\varphi} a_p \\ &\iff \forall p' \in [m'] : \; p' \preceq_{K^-}^{\varphi} p \\ &\iff p \in \max([m'], \preceq_{K^-}^{\varphi}) \\ &\iff p \in f(K^-, [m'], [n']) \end{aligned}$$

as required for **SMI**. One can show $b_q \in g(K, A_i, B_i) \iff q \in g(K^-, [m'], [n'])$ by an identical argument. $\square$

Finally, we can state the axiomatic characterisation of φ_{CI}.

Theorem 5.6.6. *φ_{CI} is the unique operator satisfying **dual**, **chain-def**, **rank-removal** and **argmax**.*

Proof. We have already seen in Theorem 5.6.5 that φ_{CI} satisfies **dual** and **chain-def**. For **argmax**, note that for any tournament K,

$$\begin{aligned} \max(A, \preceq_K^{\varphi_{\mathsf{CI}}}) &= f(K, A, B) \\ &= \arg\max_{a \in A} |K(a) \cap B| \\ &= \arg\max_{a \in A} |K(a)| \end{aligned}$$

directly from the definition of the selection function f for φ_{CI}. Finally, **rank-removal** follows from Proposition 5.6.4 since φ_{CI} is an interleaving operator with **SMI** (by Proposition 5.6.2).

For uniqueness, suppose some operator φ also satisfies the stated axioms. By Proposition 5.6.4, φ is an interleaving operator satisfying **SMI**. To show $\varphi = \varphi_{\mathsf{CI}}$ it is sufficient by Proposition 5.6.3 to show that φ and φ_{CI} agree on maximal ranks for all tournaments K. Clearly this is the case for the ranking of A, since **argmax** completely prescribes $\max(A, \preceq_K^{\varphi})$ solely in terms of K. Moreover, by **dual** we have $\sqsubseteq_K^{\varphi} = \preceq_{\overline{K}}^{\varphi}$, so

$$\begin{aligned} \max(B_K, \sqsubseteq_K^{\varphi}) &= \max(A_{\overline{K}}, \preceq_{\overline{K}}^{\varphi}) \\ &= \arg\max_{b \in A_{\overline{K}}} |\overline{K}(b)| \\ &= \max(A_{\overline{K}}, \preceq_{\overline{K}}^{\varphi_{\mathsf{CI}}}) \\ &= \max(B, \sqsubseteq_K^{\varphi_{\mathsf{CI}}}) \end{aligned}$$

where we apply **argmax** for φ and φ_{CI} to the dual tournament $\overline{K}$. This completes the proof. $\square$

5.7 Related Work

On chain graphs. Chain graphs were originally introduced by Yannakakis [106], who proved that *chain completion* – finding the minimum number of edges that when added to a bipartite graph form a chain graph – is NP-complete. Hardness results have subsequently been obtained for chain *deletion* [78] (where only edge deletions are allowed) and chain *editing* [33] (where both additions and deletions are allowed). We refer the reader to the work of Jiao, Ravi, and Gatterbauer [58] and Drange et al. [33] for a more detailed account of this literature. Outside of complexity theory, chain graphs have been studied for their spectral properties in [4, 44], and the more general notion of a *nested colouring* was introduced in [22].

On tournaments in social choice. Tournaments have important applications in the design of voting rules, where an alternative x beats y in a pairwise comparison if a majority of voters prefer x to y. Various *tournament solutions* have been proposed, which select a set of 'winners' from a given tournament.[19] Of particular relevance to our work are the *Slater set* and *Kemeney's rule* [18], which find minimal sets of edges to invert in the tournament graph such that the beating relation becomes a total order.[20] These methods are intuitively similar to chain editing: both involve making minimal changes to the tournament until some property is satisfied. A rough analogue to the Slater set in our framework is the union of the top-ranked players from each $K' \in \mathcal{M}(K)$. Solutions based on the covering relation – such as the *uncovered* and *Banks* set [18] – also bear similarity to chain editing.

Finally, note that directed versions of chain graphs (obtained by orienting edges from A to B and adding missing edges from B to A) correspond to *acyclic tournaments*, and a topological sort of A becomes a linearisation of the chain ranking $\leqslant_K^A$. This suggests a connection between chain deletion and the standard *feedback arc set* problem for removing cycles and obtaining a ranking.

On generalised tournaments. A *generalised* tournament [47] is a pair (X, T), where $X = [t]$ for some $t \in \mathbb{N}$ and $T \in \mathbb{R}_{\geq 0}^{t \times t}$ is a non-negative $t \times t$ matrix with $T_{ii} = 0$ for all $i \in X$. In this formalism each encounter between a pair of players i and j is represented by *two* numbers: T_{ij} and T_{ji}. This allows one to model both intensities of victories and losses (including draws) via the difference $T_{ij} - T_{ji}$, and the case where a comparison is not available (where $T_{ij} = T_{ji} = 0$).

Any $m \times n$ bipartite tournament K has a natural generalised tournament representation via the $(m+n) \times (m+n)$ *anti-diagonal block matrix* $T = \left[\begin{smallmatrix} 0 & K \\ K & 0 \end{smallmatrix}\right]$, where the

[19]Note that a ranking, such as we consider in this chapter, induces a set of winners by taking the maximally ranked players.

[20] Note that like chain editing, Kemeny's rule also admits a maximum likelihood characterisation [35].

top-left and bottom-right blocks are the $m \times m$ and $n \times n$ zero matrices respectively. However, such anti-diagonal block matrices are often excluded in the generalised tournament literature due to an assumption of *irreducibility*, which requires that the directed graph corresponding to T is strongly connected. This is not the case in general for T constructed as above, which means not all existing tournament operators (and tournament axioms) are well-defined for bipartite inputs.[21] Consequently, bipartite tournaments are a special case of generalised tournaments *in principle*, but not in practise.

5.8 Conclusion

Summary. In this chapter we studied chain editing, an interesting problem from computational complexity theory, as a ranking mechanism for bipartite tournaments. We analysed such mechanisms from a probabilistic viewpoint via the MLE characterisation, and in axiomatic terms. To resolve both the failure of an important anonymity axiom and NP-hardness, we weakened the chain editing requirement to one of *chain definability*, and characterised the resulting class of operators by the intuitive interleaving algorithm. Moreover, we characterised the particular interleaving instance φ_{CI} by way of new axioms.

Limitations and future work. The hardness of chain editing remains a limitation of our approach. A possible remedy is to look to one of the numerous variant problems that are polynomial-time solvable [58]; determining their applicability to ranking is an interesting topic for future work. One could develop approximation algorithms for chain editing, possibly based on existing approximations of chain completion [77]. The interleaving operators of Section 5.6.2 go in this direction, but we did not yet obtain any theoretical or experimental bounds on the approximation ratio.

A second limitation of our work lies in the assumptions of the probabilistic model; namely that the true state of the world can be reduced to vectors of numerical skill levels which totally describe the tournament participants. This assumption may be violated when the competitive element of a tournament is *multi-faceted*, since a single number cannot represent multiple orthogonal components of a player's capabilities. Nevertheless, if skill levels are taken as *aggregations* of these components, chain editing may prove to be a useful, albeit simplified, model.

Finally, there is room for more detailed axiomatic investigation. In this chapter we have stuck with fairly standard social choice axioms and performed preliminary

[21] We note that Slutzki and Volij [91] side-step the reducibility issue by decomposing T into irreducible components and ranking each separately, although their methods may give only *partial* orders.

analysis. However, the indirect nature of the comparisons in a bipartite tournament presents unique challenges; new axioms may need to be formulated to properly evaluate bipartite ranking methods in a normative sense.

Part III

Logic-based Perspectives

6 Introduction

- ...

7 Expertise and Knowledge

7.1 Introduction

In order to properly assess incoming information, it is important to consider the expertise of the reporting source. We should generally believe statements within the domain of expertise of the source, but ignore (or otherwise discount) statements about which the source has no expertise. This applies even when dealing with honest sources: a well-meaning but non-expert source may make false claims due to lack of expertise on the relevant facts. The situation may be further complicated if a source comments on multiple topics at once: we must *filter out* the parts of the statement within their domain of expertise.

Expertise has been well-studied, with perspectives from behavioural and cognitive science [19, 37], sociology [21], and philosophy [60, 101, 46], among other fields. In this work we study the *logical* content of expertise, and its relation to truthfulness of information.

Specifically, we generalise the *modal logic* setting of Singleton [87]. The two core notions of the framework are *expertise* and *soundness of information*. Intuitively, a source has expertise on φ if they are able to correctly refute φ in any situation where it is false.[1] Thus, our notion of expertise *does not depend on the "actual" state of affairs*, but only on the source's epistemic state.

It is *sound* for a source to report φ if φ is true *up to lack of expertise*: if φ is logically weakened to a proposition ψ on which the source has expertise, then ψ must be true. That is, the consequences of φ on which the source has expertise are true. This formalises the idea of "filtering out" parts of a statement within a source's expertise. For example, suppose $\varphi = p \wedge q$, and the source has expertise on p but not q. Supposing p is true but q is false, φ is false. However, if we discard information by ignoring q (on which the source has no expertise), we obtain the weaker formula

[1] Note that we could instead consider the dual case: expertise means being able to *verify* when a proposition is true.

p, which is true. Thus $p \wedge q$ is *false*, but *sound* for the source to report.

In terms of refutation, φ is sound if the source cannot refute $\neg\varphi$. That is, either φ is in fact true, or the source does not possess sufficient expertise to rule out φ.

This informal picture of expertise already suggests a close connection between expertise, soundness and *knowledge*. Indeed, we will see that, under certain conditions, expertise can be equivalently interpreted in terms of *S4 or S5 knowledge*, familiar from epistemic logic.

Beyond the individual expertise of a single source, one can also consider the *collective expertise* of a group. For example, consider a government composed of ministers and officials. While it would be unreasonable to expect each individual to have expertise on all areas of policy, one hopes that there is sufficient broadness of expertise among the government so that this is the case *collectively* (i.e. if expertise is "pooled together"). Such collective expertise may even go beyond the sum of its parts: if A is an expert on policy area X and B is an expert on how X affects some other area Y, then together A and B will have expertise on Y.

Towards defining collective expertise we will again turn to (multi-agent) epistemic logic, borrowing from the well-known notions of *distributed* and *common knowledge* [40]. Just as individual expertise (and soundness) can be expressed in terms of knowledge, we will see that collective expertise can be expressed in terms of collective knowledge.

[**TODO:** mention dynamic stuff.]

Contributions. On the conceptual side, we extend the modal framework of expertise of Singleton [87] to reason about the expertise of sources and soundness of information. We generalise this framework by working with a more general semantics, introducing collective expertise among multiple sources, and considering how expertise may evolve via learning and announcements. On the technical side we obtain axiomatisations for the more general semantics, and axiomatise several new sub-classes of models with additional axioms.

Chapter outline. In Section 7.2 we give a motivating example and define the syntax and semantics. Section 7.3 looks at how expertise may be closed under certain operations (e.g. conjunction, negation). The core connection with epistemic logic is given in Section 7.4. We turn to axiomatics in Section 7.5, and give sound and complete logics for various classes of expertise models. In Section 7.6 we generalise to multiple sources. Section 7.7 introduces the dynamic extension of the logic, and we conclude in Section 7.8. Several of the main proofs have also been formalised with the Lean theorem prover.[2] [**TODO:** reference appendix instead.]

[2] https://github.com/anonymous-logician/expertise

7.2 Expertise and Soundness

Before the formal definitions we give an example to illustrate the notions of *expertise* and *soundness*, which are central to the framework.

Example 7.2.1. *Consider an economist reporting on the possible impact of a novel virus which has recently been detected. The virus may or may not be highly infectious (i) and go on to cause a high death toll (d), and there may or may not be economic prosperity in the near future (p). The economist reports that despite the virus, the economy will prosper and there will not be mass deaths ($p \land \neg d$). Assume the economist is an expert on matters relating to the economy ($\mathsf{E}p$, $\mathsf{E}\neg p$), but not on matters of public health ($\neg\mathsf{E}d$, $\neg\mathsf{E}\neg d$). For the sake of the example, suppose the virus will in fact cause a high death toll, but the economy will nonetheless prosper. Then while the report of $p \land \neg d$ is false, it is true if one* ignores the parts on which the economist has no expertise *(namely, $\neg d$); in doing so we obtain p, which is true. The report therefore carries* some *true information, even though it is false. We say $p \land \neg d$ is* sound *for the economist in this case.*

Syntax Let Prop be a countable set of atomic propositions. To start with, we consider a single information source. Our language $\mathcal{L}$ includes modal operators to express expertise and soundness statements for this source, and is defined by the following grammar:

$$\varphi ::= p \mid \varphi \land \varphi \mid \neg\varphi \mid \mathsf{E}\varphi \mid \mathsf{S}\varphi \mid \mathsf{A}\varphi$$

for $p \in \mathsf{Prop}$. We read $\mathsf{E}\varphi$ as "the source has expertise on φ, and $\mathsf{S}\varphi$ has "φ is sound for the source to report". We include the universal modality A [48] for technical convenience; $\mathsf{A}\varphi$ is read as "φ holds in all states". Other logical connectives ($\lor$, $\to$, $\leftrightarrow$) and constants ($\top$, $\bot$) are introduced as abbreviations.

Semantics On the semantic side, we use the notion of an *expertise model*.

Definition 7.2.1. *An* expertise model *(hereafter, just* model*) is a triple $M = (X, P, V)$, where X is a set of states, $P \subseteq 2^X$ is a collection of subsets of X, and $V : \mathsf{Prop} \to 2^X$ is a valuation function. An* expertise frame *is a pair $F = (X, P)$. The class of all models is denoted by $\mathbb{M}$.*

The sets in P are termed *expertise sets*, and represent the propositions on which the source has expertise. Given the earlier informal description of expertise as refutation, we interpret $A \in P$ as saying that whenever the "actual" state is outside A, the source knows so.

For an expertise model $M = (X, P, V)$, the satisfaction relation between states $x \in X$ and formulas $\varphi \in \mathcal{L}$ is defined recursively as follows:

$$\begin{array}{lll}
M, x & \models p & \iff x \in V(p) \\
M, x & \models \varphi \wedge \psi & \iff M, x \models \varphi \text{ and } M, x \models \psi \\
M, x & \models \neg\varphi & \iff M, x \not\models \varphi \\
M, x & \models \mathsf{E}\varphi & \iff \|\varphi\|_M \in P \\
M, x & \models \mathsf{S}\varphi & \iff \forall A \in P : \|\varphi\|_M \subseteq A \implies x \in A \\
M, x & \models \mathsf{A}\varphi & \iff \forall y \in X : M, y \models \varphi
\end{array}$$

where $\|\varphi\|_M = \{x \in X \mid M, x \models \varphi\}$ is the truth set of φ. For an expertise frame $F = (X, P)$, write $F \models \varphi$ iff $M, x \models \varphi$ for all models M based on F and all $x \in X$. Write $M \models \varphi$ iff $M, x \models \varphi$ for all $x \in X$, and $\models \varphi$ iff $M \models \varphi$ for all models M; we say φ is *valid* in this case. Write $\varphi \equiv \psi$ iff $\varphi \leftrightarrow \psi$ is valid. For a set $\Gamma \subseteq \mathcal{L}$, write $\Gamma \models \varphi$ iff for all models M and states x, if $M, x \models \psi$ for all $\psi \in \Gamma$ then $M, x \models \varphi$.

The clauses for atomic propositions and propositional connectives are standard. For expertise formulas, we have that $\mathsf{E}\varphi$ holds exactly when the set of states where φ is true is an element of P. Expertise is thus a special case of the *neighbourhood semantics* [81], where each point $x \in X$ has the same set of neighbourhoods. The clause for soundness reflects the intuition that φ is sound exactly when all logically weaker formulas on which the source has expertise must be true: if $A \in P$ (i.e. the source has expertise on A) and A contains all φ states, then $x \in A$. In terms of refutation, $\mathsf{S}\varphi$ holds iff there is no expertise set A, false at the actual state x, which allows the source to rule out φ.

Our truth conditions for expertise and soundness also have topological interpretations, if one views P as the collection of closed sets of a topology on X:[3] $\mathsf{E}\varphi$ holds iff $\|\varphi\|_M$ is closed, and $\mathsf{S}\varphi$ holds at x iff x lies in the *closure* of $\|\varphi\|_M$.[4] In this case we can view the closure operation as *expanding* the set $\|\varphi\|_M$ along the lines of the source's expertise; φ is sound if the "actual" state x is included in this expansion. Finally, the clause for the universal modality A states that $\mathsf{A}\varphi$ holds iff φ holds at all states $y \in X$.

Example 7.2.2. *To formalise Example 7.2.1, consider the model $M = (X, P, V)$ shown in Fig. 7.1, where $X = 2^{\{i,p,d\}}$, $P = \{\emptyset, X, \{ipd, pd, ip, p\}, \{id, d, i, \emptyset\}\}$ (indicated by the solid rectangles; sets in X are written as strings for brevity), and $V(q) = \{S \mid q \in S\}$. Then we have $M \models \mathsf{E}p$ but $M \not\models \mathsf{E}d$. The economist's report of $p \wedge \neg d$ is represented by the dashed region. We see that while $M, ipd \not\models p \wedge \neg d$, all expertise sets containing the dashed region also contain ipd, so $M, ipd \models \mathsf{S}(p \wedge \neg d)$. That is, the economist's report is false but*

[3] For this to be the case, P must be closed under intersections and finite unions, and contain both the empty set and X itself. We will turn to these closure properties in Section 7.3.

[4] Our semantics for soundness is therefore dual to the *interior semantics* for modal logic, where $\Box\varphi$ is true at x iff x lies in the interior of $\|\varphi\|$.

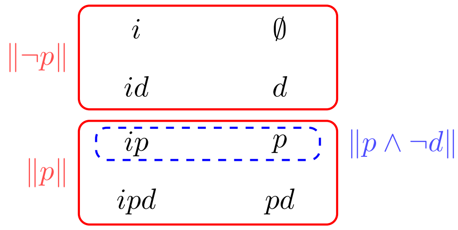


Figure 7.1: Expertise model from Example 7.2.2, which formalises the situation described in Example 7.2.1. Note that $\emptyset, X \in P$, but for clarity this is not indicated in the diagram.

sound if the "actual" state of the world were ipd. This act of "expanding" $\|p \wedge \neg d\|$ until we reach an expertise set corresponds to ignoring the parts of the report on which the economist has no expertise, as in Example 7.2.1.

We further illustrate the semantics by listing some valid formulas.

Proposition 7.2.1. *The following formulas are valid:*

1. $\varphi \to \mathsf{S}\varphi$
2. $\mathsf{E}\varphi \leftrightarrow \mathsf{AE}\varphi$
3. $\mathsf{A}(\varphi \to \psi) \to (\mathsf{S}\varphi \wedge \mathsf{E}\psi \to \psi)$
4. $\mathsf{E}\varphi \to \mathsf{A}(\mathsf{S}\varphi \to \varphi)$

Proof. Let $M = (X, P, V)$ be a model and $x \in X$. (1) and (2) are clear. For (3), suppose $M, x \models \mathsf{A}(\varphi \to \psi)$. Then $\|\varphi\|_M \subseteq \|\psi\|_M$. Further, suppose $M, x \models \mathsf{S}\varphi \wedge \mathsf{E}\psi$. Then $\|\varphi\|_M \subseteq \|\psi\|_M \in P$; taking $A = \|\psi\|_M$ in the definition of the semantics for S, we get by $M, x \models \mathsf{S}\varphi$ that $x \in \|\psi\|_M$, i.e. $M, x \models \psi$. Finally, (4) follows from (2) and (3) by taking $\psi = \varphi$. □

Here (1) says that all truths are sound. (2) says that expertise is global. (3) says that if the source has expertise on ψ, and ψ is logically weaker than some sound formula φ, then ψ is in fact true. This formalises the idea that if φ is true *up to lack of expertise*, then weakening φ until expertise holds (i.e. discarding parts of φ on which the source does not have expertise) results in something true. (4) says that if the source has expertise on φ, then whenever φ is sound it is also true.

7.3 Closure Properties

So far we have not imposed any constraints on the collection of expertise sets P. But given our interpretation of P, it may be natural to require that P is closed under certain set-theoretic operations. Say a frame $F = (X, P)$ is

- *closed under intersections* if $\{A_i\}_{i \in I} \subseteq P$ implies $\bigcap_{i \in I} A_i \in P$
- *closed under unions* if $\{A_i\}_{i \in I} \subseteq P$ implies $\bigcup_{i \in I} A_i \in P$
- *closed under finite unions* if $A, B \in P$ implies $A \cup B \in P$
- *closed under complements* if $A \in P$ implies $X \setminus A \in P$

In the first two cases we allow the empty collection $\emptyset \subseteq P$, and employ the nullary intersection convention $\bigcap \emptyset = X$. Consequently, closure under intersections implies $X \in P$, and closure under unions implies $\emptyset \in P$.

Say a model has any of the above properties if the underlying frame does. Write $\mathbb{M}_{\mathsf{int}}$, $\mathbb{M}_{\mathsf{unions}}$, $\mathbb{M}_{\mathsf{unions}}$, $\mathbb{M}_{\mathsf{finite-unions}}$ and $\mathbb{M}_{\mathsf{compl}}$ for the classes of models closed under intersections, unions, finite unions and complements respectively.

What are the intuitive interpretations of these closure conditions? Consider again our interpretation of $A \in P$: whenever the actual state is not in A, the source knows so. With this in mind, closure under intersections is a natural property: if $x \notin \bigcap_{i \in I} A_i$ then there is some $i \in I$ such that $x \notin A_i$; the source can then use this to refute A_i and therefore know that the actual state x does not lie in the intersection $\bigcap_{i \in I} A_i$. A similar argument can be made for finite unions: if $x \notin A \cup B$ then the source can use $x \notin A$ and $x \notin B$ to refute both A and B. Closure under *arbitrary* unions is less clear cut; determining that $x \notin \bigcup_{i \in I} A_i$ requires the source to refute (potentially) infinitely many propositions A_i. This is more demanding from a computational and cognitive perspective, and we therefore view closure under (arbitrary) unions as an optional property which may or may not be appropriate depending on the situation one wishes to model. Finally, closure under complements removes the distinction between refutation and *verification*: if the agent can refute A whenever A is false, they can also verify A whenever A is true. We view this as another optional property, which is appropriate in situations where *symmetric* expertise is desirable (i.e. when expertise on φ and $\neg\varphi$ should be considered equivalent).

Several of these properties can be formally captured in our language at the level of frames.

Proposition 7.3.1. *Let $F = (X, P)$ be a non-empty frame. Then*

1. *F is closed under intersections iff $F \models \mathsf{A}(\mathsf{S}\varphi \to \varphi) \to \mathsf{E}\varphi$ for all $\varphi \in \mathcal{L}$*
2. *F is closed under finite unions iff $F \models \mathsf{E}\varphi \land \mathsf{E}\psi \to \mathsf{E}(\varphi \lor \psi)$ for all $\varphi \in \mathcal{L}$*
3. *F is closed under complements iff $F \models \mathsf{E}\varphi \leftrightarrow \mathsf{E}\neg\varphi$ for all $\varphi \in \mathcal{L}$*

Proof. We prove only the first claim; the others are straightforward.

"if": We show the contrapositive. Suppose F is not closed under intersections. Then there is a collection $\{A_i\}_{i\in I} \subseteq P$ such that $B := \bigcap_{i\in I} A_i \notin P$. Let p be an arbitrary atomic proposition, and define a valuation V by $V(p) = B$ and $V(q) = \emptyset$ for $q \neq p$. Let $M = (X, P, V)$ be the corresponding model. Since X is assumed to be non-empty, we may take some $x \in X$.

We claim that $M, x \models \mathsf{A}(\mathsf{S}p \to p)$ but $M, x \not\models \mathsf{E}p$. Clearly $M, x \not\models \mathsf{E}p$ since $\|p\|_M = B \notin P$. For $M, x \models \mathsf{A}(\mathsf{S}p \to p)$, suppose $y \in X$ and $M, y \models \mathsf{S}p$. Let $j \in I$. Then $A_j \in P$, and

$$\|p\|_M = B = \bigcap_{i\in I} A_i \subseteq A_j$$

so by $M, y \models \mathsf{S}p$ we have $y \in A_j$. Hence $y \in \bigcap_{j\in I} A_j = B = \|p\|_M$, so $M, y \models p$. This shows that any $y \in X$ has $M, y \models \mathsf{S}p \to p$, and thus $M, x \models \mathsf{A}(\mathsf{S}p \to p)$. Hence $F \not\models \mathsf{A}(\mathsf{S}p \to p) \to \mathsf{E}p$.

"only if": Suppose F is closed under intersections. Let M be a model based on F and take $x \in X$. Let $\varphi \in \mathcal{L}$. Suppose $M, x \models \mathsf{A}(\mathsf{S}\varphi \to \varphi)$. Then $\|\mathsf{S}\varphi\|_M \subseteq \|\varphi\|_M$. But since $\models \varphi \to \mathsf{S}\varphi$, we have $\|\varphi\|_M \subseteq \|\mathsf{S}\varphi\|_M$ too. Hence $\|\varphi\|_M = \|\mathsf{S}\varphi\|_M$, i.e.

$$\|\varphi\|_M = \|\mathsf{S}\varphi\|_M = \bigcap\{A \in P \mid \|\varphi\|_M \subseteq A\} \in P$$

where we use the fact that P is closed under intersections in the final step. Hence $\|\varphi\|_M \in P$, so $M, x \models \mathsf{E}\varphi$. □

The question of whether closure under (arbitrary) unions can be expressed in the language is still open. By Proposition 7.3.1 (1) and Proposition 7.2.1 (4), the language fragment $\mathcal{L}_{\mathsf{SA}}$ containing only the S and A modalities is equally expressive as the full language $\mathcal{L}$ with respect to $\mathbb{M}_{\mathsf{int}}$, since $\mathsf{E}\varphi$ is equivalent to $\mathsf{A}(\mathsf{S}\varphi \to \varphi)$ in such models. In general $\mathcal{L}_{\mathsf{SA}}$ is strictly less expressive, since $\mathcal{L}_{\mathsf{SA}}$ cannot distinguish between a model and its closure under intersections.

Lemma 7.3.1. *Let $M = (X, P, V)$ be a model, and $M' = (X, P', V)$ its closure under intersections, where $A \in P'$ iff $A = \bigcap_{i\in I} A_i$ for some $\{A_i\}_{i\in I} \subseteq P$. Then for all $\varphi \in \mathcal{L}_{\mathsf{SA}}$ and $x \in X$, we have $M, x \models \varphi$ iff $M', x \models \varphi$.*

Proof. By induction on $\mathcal{L}_{\mathsf{SA}}$ formulas. The cases for atomic propositions, propositional connectives and A are straightforward. We treat only the case for S. The "if" direction is clear using the induction hypothesis and the fact that $P \subseteq P'$. Suppose $M, x \models \mathsf{S}\varphi$. Take $A = \bigcap_{i\in I} A_i \in P'$, where each A_i is in P, such that $\|\varphi\|_{M'} \subseteq A$. By the induction hypothesis, $\|\varphi\|_M \subseteq A$. For any $i \in I$, $\|\varphi\|_M \subseteq A \subseteq A_i$ and $M, x \models \mathsf{S}\varphi$ gives $x \in A_i$. Hence $x \in \bigcap_{i\in I} A_i = A$. This shows $M', x \models \mathsf{S}\varphi$. □

It follows that $\mathcal{L}_{\mathsf{SA}}$ is strictly less expressive than $\mathcal{L}$.[5] To round off the discussion of closure properties, we note that within the class of frames closed under intersections, closure under finite unions is also captured by the well-known **K** axiom – $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$ – for the dual soundness operator $\hat{\mathsf{S}}\varphi := \neg\mathsf{S}\neg\varphi$:

Proposition 7.3.2. *Suppose $F = (X, P)$ is non-empty and closed under intersections. Then F is closed under finite unions if and only if $F \models \hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi)$ for all $\varphi, \psi \in \mathcal{L}$.*

Proof. "if": We show the contrapositive. Suppose F is closed under intersections but not finite unions, so that there are $B_1, B_2 \in P$ with $B_1 \cup B_2 \notin P$. Set

$$C = \bigcap \{A \in P \mid B_1 \cup B_2 \subseteq A\}$$

By closure under intersections, $C \in P$. Clearly $B_1 \cup B_2 \subseteq C$. Since $C \in P$ but $B_1 \cup B_2 \notin P$, $B_1 \cup B_2 \subset C$. Hence there is $x \in C \setminus (B_1 \cup B_2)$.

Now pick distinct atomic propositions p and q, and let V be any valuation with $V(p) = B_1 \cup B_2$ and $V(q) = B_1$. Let $M = (X, P, V)$ be the corresponding model. We make three claims:

- $M, x \models \mathsf{S}p$: Take $A \in P$ such that $\|p\|_M \subseteq A$. Then $B_1 \cup B_2 \subseteq A$, so $C \subseteq A$. Since $x \in C$, we have $x \in A$ as required.
- $M, x \not\models \mathsf{S}q$: This is clear since $B_1 \in P$, $\|q\|_M \subseteq B_1$, but $x \notin B_1$.
- $M, x \not\models \mathsf{S}(p \land \neg q)$: Note that $\|p \land \neg q\|_M = V(p) \setminus V(q) = B_2 \setminus B_1$. Therefore we have $B_2 \in P$ and $\|p \land \neg q\|_M \subseteq B_2$, but $x \notin B_2$.

Now set $\varphi = \neg q$ and $\psi = \neg p$. We have

$$\hat{\mathsf{S}}(\varphi \to \psi) = \neg\mathsf{S}\neg(\varphi \to \psi) \equiv \neg\mathsf{S}(\varphi \land \neg\psi) \equiv \neg\mathsf{S}(p \land \neg q)$$

$$\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi = \neg\mathsf{S}\neg\varphi \to \neg\mathsf{S}\neg\psi \equiv \neg\mathsf{S}q \to \neg\mathsf{S}p \equiv \mathsf{S}p \to \mathsf{S}q$$

From the claims above we see that $M, x \models \hat{\mathsf{S}}(\varphi \to \psi)$ but $M, x \not\models \hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi$. Since M is a model based on F, we are done.

"only if": Suppose F is closed under intersections and finite unions. Let M be a model based on F and x a state in M. Suppose $M, x \models \hat{\mathsf{S}}(\varphi \to \psi)$ and $M, x \models \hat{\mathsf{S}}\varphi$. Then $M, x \not\models \mathsf{S}\neg(\varphi \to \psi)$ and $M, x \not\models \mathsf{S}\neg\varphi$. Hence there is $A \in P$ such that $\|\neg(\varphi \to \psi)\|_M \subseteq A$ but $x \notin A$, and $B \in P$ such that $\|\neg\varphi\|_M \subseteq B$ but $x \notin B$. Note

$$\|\neg\psi\|_M \subseteq \|\varphi \land \neg\psi\|_M \cup \|\neg\varphi\|_M = \|\neg(\varphi \to \psi)\|_M \cup \|\neg\varphi\|_M \subseteq A \cup B.$$

Since $x \notin A \cup B$ and $A \cup B \in P$ by closure under finite unions, this shows $M, x \not\models \mathsf{S}\neg\psi$, i.e. $M, x \models \hat{\mathsf{S}}\psi$. This completes the proof of $F \models \hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi)$. □

[5] Indeed, consider $M = (X, P, V)$, where $X = \{1, 2, 3\}$, $P = \{\{1, 2\}, \{2, 3\}\}$ and $V(p) = \{1, 2\}$, $V(q) = \{2, 3\}$ for some fixed $p, q \in \mathsf{Prop}$. Let M' be as in Lemma 7.3.1. Then $M', 1 \models \mathsf{E}(p \land q)$ and $M, 1 \not\models \mathsf{E}(p \land q)$, but M and M' agree on $\mathcal{L}_{\mathsf{SA}}$ formulas. Hence $\mathsf{E}(p \land q)$ is not equivalent to any $\mathcal{L}_{\mathsf{SA}}$ formula.

7.4 Connection with Epistemic Logic

In this section we explore the connection between our logic and *epistemic logic*, for certain classes of expertise models. In particular, we show a one-to-one mapping between classes of expertise models and *S4 and S5 relational models*, and a translation from $\mathcal{L}$ to the modal language with knowledge operator K which allows expertise and soundness to be expressed in terms of *knowledge*.

First, we introduce the syntax and (relational) semantics of epistemic logic. Let $\mathcal{L}_{\mathsf{KA}}$ be the language formed from Prop with modal operators K and A. We read $\mathsf{K}\varphi$ as *the source knows* φ.

Definition 7.4.1. *A* relational model *is a triple $M^* = (X, R, V)$, where X is a set of states, $R \subseteq X \times X$ is a binary relation on X, and $V : \mathsf{Prop} \to 2^X$ is a valuation function. The class of all relational models is denoted by $\mathbb{M}^*$.*

The satisfaction relation for $\mathcal{L}_{\mathsf{KA}}$ is defined recursively: the clauses for atomic propositions, propositional connectives and A are the same as for expertise models, and

$$M^*, x \models \mathsf{K}\varphi \iff \forall y \in X : xRy \implies M^*, y \models \varphi.$$

As is standard, R is interpreted as an *epistemic accessibility relation*: xRy means that the sources considers y possible if the "actual" state of the world is x. We will be interested in the logics of S4 and S5, which are axiomatised by **KT4** and **KT5**, respectively:

- **K**: $\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)$
- **T**: $\mathsf{K}\varphi \to \varphi$
- **4**: $\mathsf{K}\varphi \to \mathsf{K}\mathsf{K}\varphi$
- **5**: $\neg\mathsf{K}\varphi \to \mathsf{K}\neg\mathsf{K}\varphi$

T says that all knowledge is true, **4** expresses *positive introspection* of knowledge, and **5** expresses *negative introspection*.

It is well known that S4 is sound and complete for the class of relational models where R is reflexive and transitive, and that S5 is sound and complete for the class of relational models where R is an equivalence relation. Accordingly, we write $\mathbb{M}^*_{\mathsf{S4}}$ for the class of all M^* where R is reflexive and transitive, and $\mathbb{M}^*_{\mathsf{S5}}$ for M^* where R is an equivalence relation.

Our first result connecting expertise and knowledge is on the semantic side: we show there is a bijection between expertise models closed under intersections and unions and S4 models. Moreover, there is a close connection between the collection of expertise sets P and the corresponding relation R. Since expertise models closed under intersections and unions are *Alexandrov topological spaces* (where P is the set of closed sets), this is essentially a reformulation of a known result linking relational semantics over S4 frames and topological interior semantics over Alexandrov spaces [10, 80].[6] To be self-contained, we prove it for our setting here. First, we show how to map a collection of sets P to a binary relation.

Definition 7.4.2. *For a set X and $P \subseteq 2^X$, let R_P be the binary relation on X given by*

$$xR_Py \iff \forall A \in P : y \in A \implies x \in A$$

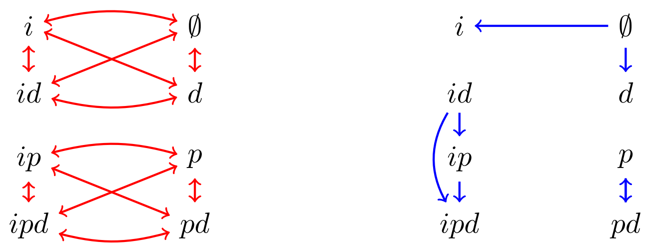


Figure 7.2: Left: the relation R_P corresponding to X and P from Example 7.2.2 (with reflexive edges omitted). Note that R_P is an equivalence relation, with equivalence classes $\|p\|$ and $\|\neg p\|$. Right: an example of a non-symmetric relation R_P, corresponding to $P = \{\emptyset, X, \{id, ip, ipd\}, \{id, ip\}, \{id\}, \{i, \emptyset\}, \{\emptyset, d\}, \{p, pd\}\}$.

In the case where P is the collection of closed sets of a topology on X, R_P is the *specialisation preorder*. Fig. 7.2 shows an example of R_P for X and P from Example 7.2.2. In what follows, say a set $A \subseteq X$ is *downwards closed* with respect to a relation R if xRy and $y \in A$ implies $x \in A$.

Lemma 7.4.1. *Let X be a set and R, S reflexive and transitive relations on X. Then if R and S share the same downwards closed sets, $R = S$.*

Proof. Suppose xRy. Set $A = \{z \in X \mid zSy\}$. By transitivity of S, A is downwards closed wrt S. By assumption, A must also be downwards closed wrt R. By reflexivity of S, $y \in A$. Hence xRy implies $x \in A$, i.e. xSy. This shows $R \subseteq S$, and the reverse inclusion holds by a symmetrical argument. Hence $R = S$. □

Lemma 7.4.2. *Let X be a set.*

1. *For any $P \subseteq 2^X$, R_P is reflexive and transitive.*
2. *If $P \subseteq 2^X$ is closed under unions and intersections, then for all $A \subseteq X$:*

$$A \in P \iff A \text{ is downwards closed wrt } R_P.$$

[6] In fact, the interior semantics has an intrinsic epistemic interpretation (without appeal to any link with relational semantics) if one views open sets as *evidence* [80, pp. 24].

3. *If R is a reflexive and transitive relation on X, there is $P \subseteq 2^X$ closed under unions and intersections such that $R_P = R$.*

Proof.

1. Straightforward by the definition of R_P.

2. Suppose P is closed under unions and intersections and let $A \subseteq X$. First suppose $A \in P$. Then A is downwards closed with respect to R_P: if $y \in A$ and xR_Py then, by definition of R_P, we have $x \in A$.

   Next suppose A is downwards closed with respect to R_P. We claim

$$A = \bigcup_{y \in A} \bigcap \{B \in P \mid y \in B\}$$

   Since P is closed under intersections and unions, this will show $A \in P$. The left-to-right inclusion is clear, since any $y \in A$ lies in the term of the union corresponding to y. For the right-to-left inclusion, take any x in the set on the RHS. Then there is $y \in A$ such that $x \in \bigcap\{B \in P \mid y \in B\}$. But this is just a rephrasing of xR_Py. Since A is downwards closed, we get $x \in A$ as required.

3. Take any reflexive and transitive relation R. Set

$$P = \{A \subseteq X \mid A \text{ is downwards closed wrt } R\}.$$

   It is easily seen that P is closed under unions and intersections. We need to show that $R_P = R$. By (1), R_P is reflexive and transitive. By Lemma 7.4.1, it is sufficient to show that R_P and R share the same downwards closed sets. Indeed, for any $A \subseteq X$ we get by (2) and the definition of P that

$$\begin{aligned} A \text{ is downwards closed wrt } R_P &\iff A \in P \\ &\iff A \text{ is downwards closed wrt } R. \end{aligned}$$

   Hence $R = R_P$.

□

We can now state the correspondence between expertise models and S4 relational models.

Theorem 7.4.1. *The mapping $f : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}} \to \mathbb{M}^*_{\mathsf{S4}}$ given by $(X, P, V) \mapsto (X, R_P, V)$ is bijective.*

Proof. Lemma 7.4.2 (1) shows that f is well-defined, i.e. that $f(M)$ does indeed lie in $\mathbb{M}^*_{\mathsf{S4}}$ for any expertise model M. Injectivity follows from Lemma 7.4.2 (2), since P is fully determined by R_P for expertise models closed under unions and intersections. Finally, Lemma 7.4.2 (3) shows that f is surjective. □

If we consider closure under complements together with intersections, an analogous result holds with S5 taking the place of S4.

Theorem 7.4.2. *The mapping* $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$ *given by* $(X, P, V) \mapsto (X, R_P, V)$ *is bijective.*

Proof. First, note that $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \subseteq \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$, since any union of sets in P can be written as a complement of intersection of complements of sets in P. Therefore g is simply the restriction of f from Theorem 7.4.1 to $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$.

To show g is well-defined, we need to show that R_P is an equivalence relation whenever P is closed under intersections and complements. Reflexivity and transitivity were already shown in Lemma 7.4.2 (1). We show R_P is symmetric. Suppose xR_Py. Let $A \in P$ such that $x \in A$. Write $B = X \setminus A$. Then since P is closed under complements, $B \in P$. Since xR_Py and $x \notin B$, we cannot have $y \in B$. Thus $y \notin B = X \setminus A$, i.e. $y \in A$. This shows yR_Px. Hence R_P is an equivalence relation.

Injectivity of g is inherited from injectivity of f from Theorem 7.4.1. For surjectivity, it suffices to show that $f^{-1}(M^*)$ is closed under complements when $M^* = (X, R, V) \in \mathbb{M}^*_{\mathsf{S5}}$. Recall, from Lemma 7.4.2 (3), that $f^{-1}(M^*) = (X, P, V)$, where $A \in P$ iff A is downwards closed with respect to R. Suppose $A \in P$, i.e. A is downwards closed. To show $X \setminus A$ is downwards closed, suppose $y \in X \setminus A$ and xRy. By symmetry of R, yRx. If $x \in A$, then downwards closure of A would give $y \in A$, but this is false. Hence $x \notin A$, i.e. $x \in X \setminus A$. Thus $X \setminus A$ is downwards closed, so P is closed under complements. This completes the proof. □

The mappings between expertise models and relational models also preserve the truth value of formulas, via the following translation $t : \mathcal{L} \to \mathcal{L}_{\mathsf{KA}}$, which expresses expertise and soundness in terms of knowledge:

$$\begin{aligned}
t(p) &= p \\
t(\varphi \wedge \psi) &= t(\varphi) \wedge t(\psi) \\
t(\neg\varphi) &= \neg t(\varphi) \\
t(\mathsf{E}\varphi) &= \mathsf{A}(\neg t(\varphi) \to \mathsf{K}\neg t(\varphi)) \\
t(\mathsf{S}\varphi) &= \neg\mathsf{K}\neg t(\varphi) \\
t(\mathsf{A}\varphi) &= \mathsf{A}t(\varphi).
\end{aligned}$$

The only interesting cases are for $\mathsf{E}\varphi$ and $\mathsf{S}\varphi$. The translation of $\mathsf{E}\varphi$ corresponds directly to the intuition of expertise as refutation: in all possible scenarios, if φ is false the source knows so. The translation of $\mathsf{S}\varphi$ says that soundness is just the dual of knowledge: φ is sound if the source does not *know* that φ is false.

Theorem 7.4.3. *Let* $f : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}} \to \mathbb{M}^*_{\mathsf{S4}}$ *be the bijection from Theorem 7.4.1. Then for all* $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$, $x \in X$ *and* $\varphi \in \mathcal{L}$:

$$M, x \models \varphi \iff f(M), x \models t(\varphi) \tag{7.1}$$

Moreover, if $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$ is the bijection from Theorem 7.4.2, then for all $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$:

$$M, x \models \varphi \iff g(M), x \models t(\varphi) \tag{7.2}$$

Proof. Note that since g is defined as the restriction of f to $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$, (7.2) follows from (7.1). We show (7.1) only. Let $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$. Write $f(M) = (X, R, V)$. From the definition of f and Lemma 7.4.2 (2), we have

$$A \in P \iff A \text{ is downwards closed wrt R} \qquad (*)$$

We show (7.1) by induction. The only non-trivial cases are E and S formulas.

(E): Suppose $M, x \models \mathsf{E}\varphi$. Then $\|\varphi\|_M \in P$. By the induction hypothesis and $(*)$, this means $\|t(\varphi)\|_{f(M)}$ is downwards closed with respect to R. Now take $y \in X$ such that $f(M), y \models \neg t(\varphi)$. Then $y \notin \|t(\varphi)\|_{f(M)}$. Since this set is downwards closed, it cannot contain any R-successor of y. Hence $f(M), y \models \mathsf{K}\neg t(\varphi)$. This shows that $f(M), x \models \mathsf{A}(\neg t(\varphi) \to \mathsf{K}\neg t(\varphi))$, i.e. $f(M), x \models t(\mathsf{E}\varphi)$.

Now suppose $f(M), x \models t(\mathsf{E}\varphi)$, i.e. $f(M), x \models \mathsf{A}(\neg t(\varphi) \to \mathsf{K}\neg t(\varphi))$. We show $\|\varphi\|_M$ is downwards closed. Suppose yRz and $z \in \|\varphi\|_M$. By the induction hypothesis, $f(M), z \not\models \neg t(\varphi)$. Hence $f(M), y \not\models \mathsf{K}\neg t(\varphi)$. Since $\neg t(\varphi) \to \mathsf{K}\neg t(\varphi)$ holds everywhere in $f(M)$, this means $f(M), y \models t(\varphi)$; by the induction hypothesis again we get $M, y \models \varphi$ and thus $y \in \|\varphi\|_M$. This shows that $\|\varphi\|_M$ is downwards closed, and by $(*)$ we have $\|\varphi\|_M \in P$. Hence $M, x \models \mathsf{E}\varphi$.

(S): We show both directions by contraposition. Suppose $M, x \not\models \mathsf{S}\varphi$. Then there is $A \in P$ such that $\|\varphi\|_M \subseteq A$ and $x \notin A$. Since A is downwards closed (by $(*)$), this means xRy implies $y \notin A$ and hence $y \notin \|\varphi\|_M$, for any $y \in X$. By the induction hypothesis, we get that xRy implies $f(M), y \models \neg t(\varphi)$, i.e. $f(M), x \models \mathsf{K}\neg t(\varphi)$. Hence $f(M), x \not\models t(\mathsf{S}\varphi)$.

Finally, suppose $f(M), x \not\models t(\mathsf{S}\varphi)$, i.e. $f(M), x \models \mathsf{K}\neg t(\varphi)$. Let A be the R-downwards closure of $\|\varphi\|_M$, i.e.

$$A = \{y \in X \mid \exists z \in \|\varphi\|_M : yRz\}$$

Then $\|\varphi\|_M \subseteq A$ by reflexivity of R, and A is downwards closed by transitivity. Hence $A \in P$. But $x \notin A$, since for all z with xRz we have $f(M), z \models \neg t(\varphi)$, so $z \notin \|t(\varphi)\|_{f(M)} = \|\varphi\|_M$. Hence $M, x \not\models \mathsf{S}\varphi$. □

Taken together, the results of this section show that, when considering expertise models closed under intersections and unions, P *uniquely determines* an epistemic accessibility relation such that expertise and soundness have precise interpretations in terms of S4 knowledge. If we also impose closure under complements, the notion of knowledge is strengthened to S5. Moreover, every S4 and S5 model arises from some expertise model in this way.

Table 7.1: Axioms and inference rules for L.

$\mathsf{E}\varphi \leftrightarrow \mathsf{AE}\varphi$	(EA)
$\mathsf{A}(\varphi \leftrightarrow \psi) \to (\mathsf{E}\varphi \leftrightarrow \mathsf{E}\psi)$	(RE_{E})
$\mathsf{A}(\varphi \to \psi) \to (S\varphi \wedge E\psi \to \psi)$	(W_{E})
$\varphi \to \mathsf{S}\varphi$	(T_{S})
$\mathsf{SS}\varphi \to \mathsf{S}\varphi$	(4_{S})
$\mathsf{A}(\varphi \to \psi) \to (S\varphi \to \mathsf{S}\psi)$	(W_{S})
$\mathsf{A}(\varphi \to \psi) \to (\mathsf{A}\varphi \to \mathsf{A}\psi)$	(K_{A})
$\mathsf{A}\varphi \to \varphi$	(T_{A})
$\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi$	(5_{A})
From φ infer $\mathsf{A}\varphi$	($\mathrm{Nec}_{\mathsf{A}}$)
From $\varphi \to \psi$ and φ infer ψ	(MP)

7.5 Axiomatisation

In this section we give sound and complete logics with respect to various classes of expertise models. We start with the class of all expertise models $\mathbb{M}$, and show how adding more axioms captures the closure conditions of Section 7.3.

The General Case Let L be the extension of propositional logic generated by the axioms and inference rules shown in Table 7.1. Note that we treat A as a "box" and S as a "diamond" modality. Some of the axioms were already seen in Proposition 7.2.1; new ones include "replacement of equivalents" for expertise (RE_{E}), **4** for S (4_{S}), and (W_{S}), which says that if ψ is logically weaker than φ then the same holds for $\mathsf{S}\psi$ and $\mathsf{S}\varphi$. First, L is sound.

Lemma 7.5.1. L *is sound with respect to* $\mathbb{M}$.

Proof. The inference rules are clearly sound. All axioms were either shown to be sound in Proposition 7.2.1 or are straightforward to see, with the possible exception of (4_{S}) which we will show explicitly. Let $M = (X, P, V)$ be an expertise model and $x \in X$. Suppose $M, x \models \mathsf{SS}\varphi$. We need to show $M, x \models \mathsf{S}\varphi$. Take $A \in P$ such that $\|\varphi\|_M \subseteq A$. Now for any $y \in X$, if $M, y \models \mathsf{S}\varphi$ then clearly $y \in A$. Hence $\|\mathsf{S}\varphi\|_M \subseteq A$. But then $M, x \models \mathsf{SS}\varphi$ gives $x \in A$. Hence $M, x \models \mathsf{S}\varphi$. □

For completeness, we use a variation of the standard canonical model method. In taking this approach, one constructs a model whose states are maximally L-consistent sets of formulas, and aims to prove the *truth lemma*: that a set Γ satisfies φ in the canonical model if and only if $\varphi \in \Gamma$. However, the truth lemma poses some difficulties for our semantics. Roughly speaking, we find there is an obvious choice of P to ensure the truth lemma for $\mathsf{E}\varphi$ formulas, but that this may be too small for

$\mathsf{S}\varphi$ to be refuted when $\mathsf{S}\varphi \notin \Gamma$ (recall that $M, x \not\models \mathsf{S}\varphi$ iff *there exists* some $A \in P$ such that $\|\varphi\|_M \subseteq A$ and $x \notin A$). We therefore "enlargen" the set of states so we can add new expertise sets A – without affecting the truth value of expertise formulas – to obtain the truth lemma for soundness formulas.

First, some standard notation and terminology. Write $\vdash \varphi$ iff $\varphi \in \mathsf{L}$. For $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, write $\Gamma \vdash \varphi$ iff there are $\psi_0, \ldots, \psi_n \in \Gamma$, $n \geq 0$, such that $\vdash (\psi_0 \wedge \cdots \wedge \psi_n) \rightarrow \varphi$. Say Γ is *inconsistent* if $\Gamma \vdash \bot$, and *consistent* otherwise. Γ is *maximally consistent* iff Γ is consistent and $\Gamma \subset \Delta$ implies that Δ is inconsistent. We recall some standard facts about maximally consistent sets.

Lemma 7.5.2. *Let Γ be a maximally consistent set and $\varphi, \psi \in \mathcal{L}$. Then*

1. $\varphi \in \Gamma$ *iff* $\Gamma \vdash \varphi$
2. *If* $\varphi \rightarrow \psi \in \Gamma$ *and* $\varphi \in \Gamma$, *then* $\psi \in \Gamma$
3. $\neg\varphi \in \Gamma$ *iff* $\varphi \notin \Gamma$
4. $\varphi \wedge \psi \in \Gamma$ *iff* $\varphi \in \Gamma$ *and* $\psi \in \Gamma$

Proof.

1. First suppose $\varphi \in \Gamma$. Since $\varphi \rightarrow \varphi$ is an instance of the propositional tautology $p \rightarrow p$, we have $\vdash \varphi \rightarrow \varphi$. Since $\varphi \in \Gamma$, this gives $\Gamma \vdash \varphi$.

   Now suppose $\Gamma \vdash \varphi$. Set $\Delta = \Gamma \cup \{\varphi\}$. We claim Δ is consistent. If not, there are $\psi_0, \ldots, \psi_n \in \Delta$ such that $\vdash (\psi_0 \wedge \cdots \wedge \psi_n) \rightarrow \bot$. Since Γ is consistent, at least one of the ψ_i must be equal to φ. Without loss of generality, $\psi_0 = \varphi$ and $\psi_j \in \Gamma$ for $j > 0$. Hence, by propositional logic and (MP), $\vdash (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \neg\varphi$. Thus $\Gamma \vdash \neg\varphi$. But since $\Gamma \vdash \varphi$ also, it follows that $\Gamma \vdash \bot$, and thus Γ is inconsistent: contradiction. So Δ must be consistent after all. Clearly $\Gamma \subseteq \Delta$, and by maximal consistency of Γ, $\Gamma \not\subset \Delta$. Hence $\Delta = \Gamma$, so $\varphi \in \Gamma$ as required.

2. By propositional logic we have $\vdash ((\varphi \rightarrow \psi) \wedge \varphi) \rightarrow \psi$. Hence $\Gamma \vdash \psi$; by (1) we get $\psi \in \Gamma$.

3. If $\neg\varphi \in \Gamma$ then clearly $\varphi \notin \Gamma$, since otherwise Γ would be inconsistent. If $\varphi \notin \Gamma$ then $\Gamma \not\vdash \varphi$ by (1). Set $\Delta = \Gamma \cup \{\neg\varphi\}$. Then Δ is consistent (one can show that assuming Δ is inconsistent leads to $\Gamma \vdash \varphi$; a contradiction). Again, since $\Gamma \subseteq \Delta$ and Γ is maximally consistent, we must in fact have $\Gamma = \Delta$, so $\neg\varphi \in \Gamma$.

4. If $\varphi \wedge \psi \in \Gamma$ then both $\Gamma \vdash \varphi$ and $\Gamma \vdash \psi$, so $\varphi, \psi \in \Gamma$ by (1). Conversely, if $\varphi, \psi \in \Gamma$ then $\Gamma \vdash \varphi \wedge \psi$, so $\varphi \wedge \psi \in \Gamma$ by (1) again.

$\square$

Lemma 7.5.3 (Lindenbaum's Lemma)**.** *If $\Gamma \subseteq \mathcal{L}$ is consistent there is a maximally consistent set Δ such that $\Gamma \subseteq \Delta$.*

Let X_{L} denote the set of maximally consistent sets. Define a relation R by

$$\Gamma R \Delta \iff \forall \varphi \in \mathcal{L} : \mathsf{A}\varphi \in \Gamma \implies \varphi \in \Delta$$

The (T_{A}) and (5_{A}) axioms for A show that R is an equivalence relation; this is part of the standard proof that S5 is complete for equivalence relations.

Lemma 7.5.4. *R is an equivalence relation.*

Proof. We first show that R is reflexive and has the *Euclidean property* (xRy and xRz implies yRz). For reflexivity, let $\Gamma \in X_{\mathsf{L}}$. Suppose $\mathsf{A}\varphi \in \Gamma$. By ($\mathrm{T}_{\mathsf{A}}$) and closure of maximally consistent sets under modus ponens, $\varphi \in \Gamma$. Hence $\Gamma R \Gamma$.

For the Euclidean property, suppose $\Gamma R \Delta$ and $\Gamma R \Lambda$. We show $\Delta R \Lambda$ by contraposition. Suppose $\varphi \notin \Lambda$. Since $\Gamma R \Lambda$, this means $\mathsf{A}\varphi \notin \Gamma$. Hence $\neg\mathsf{A}\varphi \in \Gamma$, and by ($5_{\mathsf{A}}$) we get $\mathsf{A}\neg\mathsf{A}\varphi \in \Gamma$. Now $\Gamma R \Delta$ gives $\neg\mathsf{A}\varphi \in \Delta$, so $\mathsf{A}\varphi \notin \Delta$.

To conclude we need to show R is symmetric and transitive. For symmetry, suppose $\Gamma R \Delta$. By reflexivity, $\Gamma R \Gamma$. The Euclidean property therefore gives $\Delta R \Gamma$. For transitivity, suppose $\Gamma R \Delta$ and $\Delta R \Lambda$. By symmetry, $\Delta R \Gamma$. The Euclidean property again gives $\Gamma R \Lambda$. $\square$

For $\varphi \in \mathcal{L}$, let $|\varphi| = \{\Gamma \in X_{\mathsf{L}} \mid \varphi \in \Gamma\}$ be the *proof set* of φ. For $\Sigma \in X_{\mathsf{L}}$, let X_Σ be the equivalence class of Σ in R, and write $|\varphi|_\Sigma = |\varphi| \cap X_\Sigma$. Using what is essentially the standard proof of the truth lemma for the modal logic **K** with respect to relational semantics, (K_{A}) yields the following.

Lemma 7.5.5. *Let $\Sigma \in X_{\mathsf{L}}$. Then*

1. *For any $\varphi \in \mathcal{L}$, $\mathsf{A}\varphi \in \Sigma$ iff $|\varphi|_\Sigma = X_\Sigma$*
2. *For any $\varphi, \psi \in \mathcal{L}$, $\mathsf{A}(\varphi \to \psi) \in \Sigma$ iff $|\varphi|_\Sigma \subseteq |\psi|_\Sigma$*
3. *For any $\varphi, \psi \in \mathcal{L}$, $\mathsf{A}(\varphi \leftrightarrow \psi) \in \Sigma$ iff $|\varphi|_\Sigma = |\psi|_\Sigma$*

Proof.

1. For the left-to-right direction, suppose $\mathsf{A}\varphi \in \Sigma$. Let $\Gamma \in X_\Sigma$. Then $\Sigma R \Gamma$, so clearly $\varphi \in \Gamma$. Hence $|\varphi|_\Sigma = X_\Sigma$. For the other direction we show the contrapositive. Suppose $\mathsf{A}\varphi \notin \Sigma$. Set

$$\Gamma_0 = \{\psi \mid \mathsf{A}\psi \in \Gamma\} \cup \{\neg\varphi\}.$$

We claim Γ_0 is consistent. If not, without loss of generality there are $\psi_0, \dots, \psi_n \in \Gamma_0$ such that $\mathsf{A}\psi_i \in \Sigma$ for each i, and $\vdash \psi_0 \wedge \cdots \wedge \psi_n \to \varphi$. By propositional logic, we get $\vdash \psi_0 \to \cdots \to \psi_n \to \varphi$ (where the implication arrows associate to the right) and by (Nec_A), $\vdash \mathsf{A}(\psi_0 \to \cdots \to \psi_n \to \varphi)$. Since ($\mathrm{K}_\mathsf{A}$) together with (MP) says that A distributes over implications, repeated applications gives $\vdash \mathsf{A}\psi_0 \to \cdots \to \mathsf{A}\psi_n \to \mathsf{A}\varphi$ and propositional logic again gives $\vdash \mathsf{A}\psi_0 \wedge \cdots \wedge \mathsf{A}\psi_n \to \mathsf{A}\varphi$. But recall that $\mathsf{A}\psi_i \in \Sigma$. Hence $\Sigma \vdash \mathsf{A}\varphi$. Since Σ is maximally consistent, this means $\mathsf{A}\varphi \in \Sigma$: contradiction.

So Γ_0 is consistent. By Lindenbaum's lemma (Lemma 7.5.3), there is a maximally consistent set $\Gamma \supseteq \Gamma_0$. Clearly $\Sigma R \Gamma$, since if $\mathsf{A}\psi \in \Sigma$ then $\psi \in \Gamma_0 \subseteq \Gamma$. Moreover, $\neg\varphi \in \Gamma_0 \subseteq \Gamma$, so by consistency $\varphi \notin \Gamma$. Hence $\Gamma \in X_\Sigma \setminus |\varphi|_\Sigma$, and we are done.

2. Note that by (1) we have

$$\begin{aligned} \mathsf{A}(\varphi \to \psi) \in \Sigma &\iff |\varphi \to \psi|_\Sigma = X_\Sigma \\ &\iff \forall \Gamma \in X_\Sigma : \varphi \to \psi \in \Gamma \end{aligned}$$

Suppose $\mathsf{A}(\varphi \to \psi) \in \Sigma$. Take $\Gamma \in |\varphi|_\Sigma$. Then we have $\varphi, \varphi \to \psi \in \Gamma$, so $\psi \in \Gamma$. This shows $|\varphi|_\Sigma \subseteq |\psi|_\Sigma$. Conversely, suppose $|\varphi|_\Sigma \subseteq |\psi|_\Sigma$. Take $\Gamma \in X_\Sigma$. If $\varphi \notin \Gamma$ then $\neg\varphi \in \Gamma$, so $\neg\varphi \vee \psi \in \Gamma$ and thus $\varphi \to \psi \in \Gamma$. If $\varphi \in \Gamma$ then $\Gamma \in |\varphi|_\Sigma \subseteq |\psi|_\Sigma$, so $\psi \in \Gamma$. Thus $\varphi \to \psi \in \Gamma$ in this case too. Hence $\mathsf{A}(\varphi \to \psi) \in \Sigma$.

3. First note that $\mathsf{A}(\alpha \wedge \beta) \in \Sigma$ iff both $\mathsf{A}\alpha \in \Sigma$ and $\mathsf{A}\beta \in \Sigma$. This can be shown using (K_A), (MP) and instances of the propositional tautologies $(p \wedge q) \to p$ (for the left-to-right implication) and $p \to q \to (p \wedge q))$ (for the right-to-left implication). Recalling that $\varphi \leftrightarrow \psi$ is an abbreviation for $(\varphi \to \psi) \wedge (\psi \to \varphi)$, we get

$$\begin{aligned} \mathsf{A}(\varphi \leftrightarrow \psi) \in \Sigma &\iff \mathsf{A}(\varphi \to \psi) \in \Sigma \text{ and } \mathsf{A}(\psi \to \varphi) \in \Sigma \\ &\iff |\varphi|_\Sigma \subseteq |\psi|_\Sigma \text{ and } |\psi|_\Sigma \subseteq |\varphi|_\Sigma \\ &\iff |\varphi|_\Sigma = |\psi|_\Sigma \end{aligned}$$

as required.

□

Corollary 7.5.1. *Let $\Sigma \in X_\mathsf{L}$. For $\Gamma, \Delta \in X_\Sigma$ and $\varphi \in \mathcal{L}$, $\mathsf{A}\varphi \in \Gamma$ iff $\mathsf{A}\varphi \in \Delta$ and $\mathsf{E}\varphi \in \Gamma$ iff $\mathsf{E}\varphi \in \Delta$.*

Proof. For the first point, note that if $\mathsf{A}\varphi \in \Gamma$ then Lemma 7.5.5 gives $|\varphi|_\Gamma = X_\Gamma$. But since Γ and Δ are in the same equivalence class of R, $|\varphi|_\Gamma = |\varphi|_\Delta$ and $X_\Gamma = X_\Delta$. Hence $|\varphi|_\Delta = X_\Delta$, so $\mathsf{A}\varphi \in \Delta$ by Lemma 7.5.5. The converse holds by symmetry.

For the second point, if $\mathsf{E}\varphi \in \Gamma$ then $\mathsf{AE}\varphi \in \Gamma$ by (EA). Since $\Gamma R \Delta$, we get $\mathsf{E}\varphi \in \Delta$. Again, the converse holds by symmetry. □

We are ready to define the "canonical" model (for each Σ). Set $\widehat{X}_\Sigma = X_\Sigma \times \mathbb{R}$. This is the step described informally above: we enlargen X_Σ by considering uncountably many copies of each point (any uncountable set would do in place of $\mathbb{R}$). The valuation is straightforward: set $\widehat{V}_\Sigma(p) = |p|_\Sigma \times \mathbb{R}$. For the expertise component of the model, say $A \subseteq \widehat{X}_\Sigma$ is *S-closed* iff for all $\varphi \in \mathcal{L}$:

$$|\varphi|_\Sigma \times \mathbb{R} \subseteq A \implies |\mathsf{S}\varphi|_\Sigma \times \mathbb{R} \subseteq A.$$

Set $\widehat{P}_\Sigma = \widehat{P}^0_\Sigma \cup \widehat{P}^1_\Sigma$, where

$$\widehat{P}^0_\Sigma = \{|\varphi|_\Sigma \times \mathbb{R} \mid \mathsf{E}\varphi \in \Sigma\},$$
$$\widehat{P}^1_\Sigma = \{A \subseteq \widehat{X}_\Sigma \mid A \text{ is S-closed and } \forall \varphi \in \mathcal{L} : A \neq |\varphi|_\Sigma \times \mathbb{R}\}.$$

We have a version of the truth lemma for the model $\widehat{M}_\Sigma = (\widehat{X}_\Sigma, \widehat{P}_\Sigma, \widehat{V}_\Sigma)$.

Lemma 7.5.6. *For any* $(\Gamma, t) \in \widehat{X}_\Sigma$ *and* $\varphi \in \mathcal{L}$*,*

$$\widehat{M}_\Sigma, (\Gamma, t) \models \varphi \iff \varphi \in \Gamma,$$

i.e. $\|\varphi\|_{\widehat{M}_\Sigma} = |\varphi|_\Sigma \times \mathbb{R}$.

Proof. By induction. The cases for atomic propositions and the propositional connectives are straightforward by the definition of $\widehat{V}_\Sigma$ and properties of maximally consistent sets. The case for the universal modality A is also straightforward by Lemma 7.5.5 and Corollary 7.5.1. We treat the cases of E and S formulas.

(E): First suppose $\mathsf{E}\varphi \in \Gamma$. By Corollary 7.5.1, $\mathsf{E}\varphi \in \Sigma$. Hence $|\varphi|_\Sigma \times \mathbb{R} \in \widehat{P}^0_\Sigma$. By the induction hypothesis, $\|\varphi\|_{\widehat{M}_\Sigma} \in \widehat{P}^0_\Sigma$. Hence $\widehat{M}_\Sigma, (\Gamma, t) \models \mathsf{E}\varphi$.

Now suppose $\widehat{M}_\Sigma, (\Gamma, t) \models \mathsf{E}\varphi$. Then by the induction hypothesis, $|\varphi|_\Sigma \times \mathbb{R} \in \widehat{P}_\Sigma$. Since $\widehat{P}^1_\Sigma$ does not contain any sets of this form, we must have $|\varphi|_\Sigma \times \mathbb{R} \in \widehat{P}^0_\Sigma$. Therefore there is some ψ such that $\mathsf{E}\psi \in \Sigma$ and $|\varphi|_\Sigma \times \mathbb{R} = |\psi|_\Sigma \times \mathbb{R}$. It follows that $|\varphi|_\Sigma = |\psi|_\Sigma$, and Lemma 7.5.5 then gives $\mathsf{A}(\varphi \leftrightarrow \psi) \in \Sigma$. By Corollary 7.5.1, we have $\mathsf{E}\psi \in \Gamma$ and $\mathsf{A}(\varphi \leftrightarrow \psi) \in \Gamma$ too. By (RE_E) we get $\mathsf{E}\varphi \in \Gamma$ as required.

(S): First suppose $\mathsf{S}\varphi \in \Gamma$. Take $A \in \widehat{P}_\Sigma$ such that $\|\varphi\|_{\widehat{M}_\Sigma} \subseteq A$. By the induction hypothesis, $|\varphi|_\Sigma \times \mathbb{R} \subseteq A$. There are two cases: either $A \in \widehat{P}^0_\Sigma$ or $A \in \widehat{P}^1_\Sigma$.

If $A \in \widehat{P}^0_\Sigma$, there is ψ such that $A = |\psi|_\Sigma \times \mathbb{R}$ and $\mathsf{E}\psi \in \Sigma$. Since $|\varphi|_\Sigma \times \mathbb{R} \subseteq A$, we have $|\varphi|_\Sigma \subseteq |\psi|_\Sigma$. By Lemma 7.5.5, $\mathsf{A}(\varphi \to \psi) \in \Sigma$. By Corollary 7.5.1 we have $\mathsf{E}\psi, \mathsf{A}(\varphi \to \psi) \in \Gamma$ too. Applying (W_E) gives $\mathsf{S}\varphi \wedge \mathsf{E}\psi \to \psi \in \Gamma$; since $\mathsf{S}\varphi, \mathsf{E}\psi \in \Gamma$ we have $\mathsf{S}\varphi \wedge \mathsf{E}\psi \in \Gamma$ and thus $\psi \in \Gamma$. This means $(\Gamma, t) \in |\psi|_\Sigma \times \mathbb{R} = A$, as required.

If $A \in \widehat{P}^1_\Sigma$, A is S-closed by definition. Hence $|\mathsf{S}\varphi|_\Sigma \times \mathbb{R} \subseteq A$. Since $\mathsf{S}\varphi \in \Gamma$ we get $(\Gamma, t) \in A$ as required.

In either case we have $(\Gamma, t) \in A$. This shows $\widehat{M}_\Sigma, (\Gamma, t) \models \mathsf{S}\varphi$.

For the other direction we show the contrapositive. Take any $(\Gamma, t) \in \widehat{X}_\Sigma$ and suppose $\mathsf{S}\varphi \notin \Gamma$. We show that $\widehat{M}_\Sigma, (\Gamma, t) \not\models \mathsf{S}\varphi$, i.e. there is $A \in \widehat{P}_\Sigma$ such that $\|\varphi\|_{\widehat{M}_\Sigma} \subseteq A$ but $(\Gamma, t) \notin A$. First, set

$$\mathcal{U} = \{|\psi|_\Sigma \times \mathbb{R} \mid \psi \in \mathcal{L} \text{ and } |\psi|_\Sigma \times \mathbb{R} \not\subseteq |\mathsf{S}\varphi|_\Sigma \times \mathbb{R}\}.$$

Since $\mathcal{L}$ is countable, $\mathcal{U}$ is at most countable. Hence we may write $\mathcal{U} = \{U_n\}_{n \in N}$ for some index set $N \subseteq \mathbb{N}$. Since $U_n \not\subseteq |\mathsf{S}\varphi|_\Sigma \times \mathbb{R}$, we may choose some $(\Delta_n, t_n) \in U_n \setminus (|\mathsf{S}\varphi|_\Sigma \times \mathbb{R})$ for each n. Now write

$$\mathcal{D} = \{(\Delta_n, t_n)\}_{n \in N} \cup \{(\Gamma, t)\}.$$

Since N is at most countable, so too is $\mathcal{D}$. Since $\mathbb{R}$ is uncountable, there is some $s \in \mathbb{R}$ such that $(\Gamma, s) \notin \mathcal{D}$.[7] We necessarily have $s \neq t$. We are ready to define A: set

$$A = (|\mathsf{S}\varphi|_\Sigma \times \mathbb{R}) \cup \{(\Gamma, s)\}.$$

Note that $(\Gamma, t) \notin A$ since $\mathsf{S}\varphi \notin \Gamma$ and $s \neq t$. Next we show $\|\varphi\|_{\widehat{M}_\Sigma} \subseteq A$. By the induction hypothesis, this is equivalent to $|\varphi|_\Sigma \times \mathbb{R} \subseteq A$. By ($\mathsf{T}_\mathsf{S}$) and ($\mathsf{Nec}_\mathsf{A}$), we have $\mathsf{A}(\varphi \to \mathsf{S}\varphi) \in \Sigma$, and consequently $|\varphi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$ by Lemma 7.5.5. Hence $|\varphi|_\Sigma \times \mathbb{R} \subseteq |\mathsf{S}\varphi|_\Sigma \times \mathbb{R} \subseteq A$ as required.

It only remains to show that $A \in \widehat{P}_\Sigma$. We claim that $A \in \widehat{P}^1_\Sigma$. First, A is S-closed. Indeed, suppose $|\psi|_\Sigma \times \mathbb{R} \subseteq A$. We claim that, in fact, $|\psi|_\Sigma \times \mathbb{R} \subseteq |\mathsf{S}\varphi|_\Sigma \times \mathbb{R}$. If not, then by definition of $\mathcal{U}$ there is $n \in N$ such that $|\psi|_\Sigma \times \mathbb{R} = U_n$. Hence $U_n \subseteq A$. This means $(\Delta_n, t_n) \in A$. But $(\Delta_n, t_n) \notin |\mathsf{S}\varphi|_\Sigma \times \mathbb{R}$, so we must have $(\Delta_n, t_n) = (\Gamma, s)$. But then $(\Gamma, s) \in \mathcal{D}$: contradiction. So we do indeed have $|\psi|_\Sigma \times \mathbb{R} \subseteq |\mathsf{S}\varphi|_\Sigma \times \mathbb{R}$, and thus $|\psi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$. By Lemma 7.5.5, $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Sigma$.

Now, take any $(\Lambda, u) \in |\mathsf{S}\psi|_\Sigma \times \mathbb{R}$. Since $\Lambda \in X_\Sigma$, Corollary 7.5.1 gives $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Lambda$. By ($\mathsf{W}_\mathsf{S}$), $\mathsf{S}\psi \to \mathsf{SS}\varphi \in \Lambda$. Since $\Lambda \in |\mathsf{S}\psi|_\Sigma$, we get $\mathsf{SS}\varphi \in \Lambda$. But then ($4_\mathsf{S}$) gives $\mathsf{S}\varphi \in \Lambda$. That is, $(\Lambda, u) \in |\mathsf{S}\varphi|_\Sigma \times \mathbb{R} \subseteq A$. This shows $|\mathsf{S}\psi|_\Sigma \times \mathbb{R} \subseteq A$, so A is S-closed.

Finally, we show that for all $\psi \in \mathcal{L}$, $A \neq |\psi|_\Sigma \times \mathbb{R}$. For contradiction, suppose there is ψ with $A = |\psi|_\Sigma \times \mathbb{R}$. Then since $(\Gamma, s) \in A$, we have $\Gamma \in |\psi|_\Sigma$. But then $(\Gamma, t) \in |\psi|_\Sigma \times \mathbb{R} = A$: contradiction.

This completes the proof that $A \in \widehat{P}^1_\Sigma$. Thus $\widehat{M}_\Sigma, (\Gamma, t) \not\models \mathsf{S}\varphi$, and we are done. □

Theorem 7.5.1. L *is strongly complete*[8] *with respect to* $\mathbb{M}$.

[7] If not, then $s \mapsto (\Gamma, s)$ is an injective mapping $\mathbb{R} \to \mathcal{D}$, which would imply $\mathbb{R}$ is countable.
[8] That is, for all sets $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, if $\Gamma \models \varphi$ then $\Gamma \vdash \varphi$.

Proof. We show the contrapositive. Suppose $\Gamma \not\vdash \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is consistent. By Lindenbaum's Lemma, there is a maximally consistent set $\Sigma \supseteq \Gamma \cup \{\neg\varphi\}$. Consider the model $\widehat{M}_\Sigma$. For any $\psi \in \Gamma$ we have $\psi \in \Sigma$, so Lemma 7.5.6 (with $t = 0$, say) gives $\widehat{M}_\Sigma, (\Sigma, 0) \models \psi$. Also, $\neg\varphi \in \Gamma \subseteq \Sigma$ gives $\widehat{M}_\Sigma, (\Sigma, 0) \models \neg\varphi$, so $\widehat{M}_\Sigma, (\Sigma, 0) \not\models \varphi$. This shows that $\Gamma \not\models \varphi$, and we are done. □

Extensions of the Base Logic We now extend L to obtain axiomatisations of subclasses of $\mathbb{M}$ corresponding to closure conditions.

To start, consider closure under intersections. It was shown in Proposition 7.3.1 that the formula $\mathsf{A}(\mathsf{S}\varphi \to \varphi) \to \mathsf{E}\varphi$ characterises frames closed under intersections. It is perhaps no surprise that adding this as an axiom results in a sound and complete axiomatisation of $\mathbb{M}_{\mathsf{int}}$. Formally, let $\mathsf{L}_{\mathsf{int}}$ be the extension of L with the following axiom

$$\mathsf{A}(\mathsf{S}\varphi \to \varphi) \to \mathsf{E}\varphi \quad (\mathrm{Red}_{\mathsf{E}}),$$

so-named since together with $\mathsf{E}\varphi \to \mathsf{A}(\mathsf{S}\varphi \to \varphi)$ – which is derivable in L – it allows expertise to be reduced to soundness. That is, expertise on φ is equivalent to the statement that, in all situations, φ is only true up to lack of expertise if it is in fact true.

Theorem 7.5.2. $\mathsf{L}_{\mathsf{int}}$ *is sound and strongly complete with respect to* $\mathbb{M}_{\mathsf{int}}$.

Proof. For soundness, we only need to check that $(\mathrm{Red}_{\mathsf{E}})$ is sound for $\mathbb{M}_{\mathsf{int}}$. But this follows from Proposition 7.3.1 (1).

For completeness, we adopt a roughly similar approach to the general case. Let consistency, maximal consistency and other standard notions and notation be defined as before, but now for $\mathsf{L}_{\mathsf{int}}$ instead of L. Let $X_{\mathsf{L}_{\mathsf{int}}}$ be the set of maximally $\mathsf{L}_{\mathsf{int}}$-consistent sets. Define the relation R on $X_{\mathsf{L}_{\mathsf{int}}}$ in exactly the same way. Since $\mathsf{L}_{\mathsf{int}}$ extends L, R is again an equivalence relation, and we have the analogues of Lemma 7.5.5 and Corollary 7.5.1.

This time, however, the construction of the canonical model for a given $\Sigma \in X_{\mathsf{L}_{\mathsf{int}}}$ is much more straightforward. The set of states is simply X_Σ, i.e. the equivalence class of Σ in R. Overriding earlier terminology, say $A \subseteq X_\Sigma$ is *S-closed* iff $|\varphi|_\Sigma \subseteq A$ implies $|\mathsf{S}\varphi|_\Sigma \subseteq A$ for all $\varphi \in \mathcal{L}$. Then set

$$P_\Sigma = \{A \subseteq X_\Sigma \mid A \text{ is S-closed}\}.$$

Finally, set $V_\Sigma(p) = |p|_\Sigma$, and write $M_\Sigma = (X_\Sigma, P_\Sigma, V_\Sigma)$.

First, we have $M_\Sigma \in \mathbb{M}_{\mathsf{int}}$, i.e. intersections of S-closed sets are S-closed. Indeed, suppose $\{A_i\}_{i \in I}$ is a collection of S-closed sets, and suppose $|\varphi|_\Sigma \subseteq \bigcap_{i \in I} A_i$. Then $|\varphi|_\Sigma \subseteq A_i$ for each i, so S-closure gives $|\mathsf{S}\varphi|_\Sigma \subseteq A_i$. Hence $|\mathsf{S}\varphi|_\Sigma \subseteq \bigcap_{i \in I} A_i$.

Importantly, we have the truth lemma for M_Σ: for all $\Gamma \in X_\Sigma$ and $\varphi \in \mathcal{L}$,

$$M_\Sigma, \Gamma \models \varphi \iff \varphi \in \Gamma,$$

i.e. $\|\varphi\|_{M_\Sigma} = |\varphi|_\Sigma$.

As usual, the proof is by induction on formulas. The case for atomic propositions follows from the definition of V_Σ, the cases for conjunctions and negations hold by properties of maximally consistent sets, and the case for $\mathsf{A}\varphi$ holds by an argument identical to the one used in the general case (Lemma 7.5.6). The only interesting cases are therefore for $\mathsf{E}\varphi$ and $\mathsf{S}\varphi$ formulas:

(E): First suppose $\mathsf{E}\varphi \in \Gamma$. We claim $|\varphi|_\Sigma$ is S-closed. This will give $\|\varphi\|_{M_\Sigma} \in P_\Sigma$ by the induction hypothesis and definition of P_Σ, and therefore $M_\Sigma, \Gamma \models \mathsf{E}\varphi$.

So, suppose $|\psi|_\Sigma \subseteq |\varphi|_\Sigma$. Then $\mathsf{A}(\psi \to \varphi) \in \Sigma$. Let $\Delta \in |\mathsf{S}\psi|_\Sigma$. Since $\Delta, \Gamma, \Sigma \in X_\Sigma$, we have $\mathsf{E}\varphi \in \Delta$ and $\mathsf{A}(\psi \to \varphi) \in \Delta$ too. By (W_E), $\mathsf{S}\psi \wedge \mathsf{E}\varphi \to \varphi \in \Delta$. But $\mathsf{S}\psi \in \Delta$, so $\mathsf{S}\psi \wedge \mathsf{E}\varphi \in \Delta$ and thus $\varphi \in \Delta$, i.e. $\Delta \in |\varphi|_\Sigma$. This shows $|\mathsf{S}\psi|_\Sigma \subseteq |\varphi|_\Sigma$, so $|\varphi|_\Sigma$ is S-closed as required.

Now suppose $M_\Sigma, \Gamma \models \mathsf{E}\varphi$. Then, by the induction hypothesis, $|\varphi|_\Sigma$ is S-closed. Since $|\varphi|_\Sigma \subseteq |\varphi|_\Sigma$ clearly holds, we get $|\mathsf{S}\varphi|_\Sigma \subseteq |\varphi|_\Sigma$. This implies $\mathsf{A}(\mathsf{S}\varphi \to \varphi) \in \Sigma$, and ($\mathrm{Red}_\mathsf{E}$) gives $\mathsf{E}\varphi \in \Sigma$. Since $\Gamma \in X_\Sigma$, we get $\mathsf{E}\varphi \in \Gamma$ as required.

(S): Suppose $\mathsf{S}\varphi \in \Gamma$. Take any $A \in P_\Sigma$ such that $\|\varphi\|_{M_\Sigma} \subseteq A$. By the induction hypothesis, $|\varphi|_\Sigma \subseteq A$. By S-closure of A, $|\mathsf{S}\varphi|_\Sigma \subseteq A$. Hence $\Gamma \in |\mathsf{S}\varphi|_\Sigma \subseteq A$. This shows $M_\Sigma, \Gamma \models \mathsf{S}\varphi$.

For the other direction we show the contrapositive. Suppose $\mathsf{S}\varphi \notin \Gamma$. First, we claim $|\mathsf{S}\varphi|_\Sigma$ is S-closed. Indeed, suppose $|\psi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$. Then $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Sigma$. Take any $\Delta \in |\mathsf{S}\psi|_\Sigma$. Since $\Delta \in X_\Sigma$, $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Delta$ also. By (W_S), $\mathsf{S}\psi \to \mathsf{S}\mathsf{S}\varphi \in \Delta$. Now $\mathsf{S}\psi \in \Delta$ implies $\mathsf{S}\mathsf{S}\varphi \in \Delta$, and ($4_\mathsf{S}$) gives $\mathsf{S}\varphi \in \Delta$, i.e. $\Delta \in |\mathsf{S}\varphi|_\Sigma$. This shows $|\mathsf{S}\psi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$, and thus $|\mathsf{S}\varphi|_\Sigma$ is S-closed.

Hence $|\mathsf{S}\varphi|_\Sigma$ is a set in P_Σ not containing Γ. Moreover, $\|\varphi\|_{M_\Sigma} \subseteq |\mathsf{S}\varphi|_\Sigma$ by the induction hypothesis and (T_S). Hence $M_\Sigma, \Gamma \not\models \mathsf{S}\varphi$.

Strong completeness now follows. If $\Gamma \not\vdash_{\mathsf{L}_{\mathsf{int}}} \varphi$, then $\Gamma \cup \{\neg\varphi\}$ is consistent, so by Lindenbaum's Lemma there is $\Sigma \in X_{\mathsf{L}_{\mathsf{int}}}$ with $\Sigma \supseteq \Gamma \cup \{\neg\varphi\}$. Considering the model $M_\Sigma \in \mathbb{M}_{\mathsf{int}}$, we have $M_\Sigma, \Sigma \models \Gamma$ and $M_\Sigma, \Sigma \not\models \varphi$ by the truth lemma. Hence $\Gamma \not\models_{\mathbb{M}_{\mathsf{int}}} \varphi$. $\square$

Now we add finite unions to the mix. It was shown in Proposition 7.3.2 that within class $\mathbb{M}_{\mathsf{int}}$, the **K** axiom for the dual operator $\hat{\mathsf{S}}$ characterises closure under finite unions. Note that any frame (X, P) closed under intersections and finite unions is a topological space,[9] where P is the set of *closed* sets. Write $\mathbb{M}_{\mathsf{top}} = \mathbb{M}_{\mathsf{int}} \cap$

[9]By the convention that the empty intersection is the whole space X and the empty union is $\emptyset$, we have $X, \emptyset \in P$ too.

$\mathbb{M}_{\mathsf{finite-unions}}$ for the class of models over such frames. We obtain an axiomatisation of $\mathbb{M}_{\mathsf{top}}$ by adding **K** for $\hat{\mathsf{S}}$ and a bridge axiom linking $\hat{\mathsf{S}}$ and A:

$$\hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi) \quad (\mathrm{K}_{\mathsf{S}})$$
$$\mathsf{A}\varphi \to \hat{\mathsf{S}}\varphi \quad (\mathrm{Inc})$$

Let $\mathsf{L}_{\mathsf{top}}$ be the extension of $\mathsf{L}_{\mathsf{int}}$ by $(\mathrm{K}_{\mathsf{S}})$ and (Inc). Note that $\mathsf{L}_{\mathsf{top}}$ contains the **KT4** axioms for $\hat{\mathsf{S}}$ (recalling that $(\mathrm{T}_{\mathsf{S}})$ and (4_{S}) are the "diamond" versions of **T** and **4**). Since **KT4** together with the bridge axiom (Inc) is complete for the class of relational models $\mathbb{M}^*_{\mathsf{S4}}$, we can exploit Theorem 7.4.3 to obtain completeness of $\mathsf{L}_{\mathsf{top}}$ with respect to $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$. Since this class is included in $\mathbb{M}_{\mathsf{top}}$, we also get completeness with respect to $\mathbb{M}_{\mathsf{top}}$.[10]

Theorem 7.5.3. $\mathsf{L}_{\mathsf{top}}$ *is sound and strongly complete with respect to* $\mathbb{M}_{\mathsf{top}}$.

Proof. Soundness of $(\mathrm{K}_{\mathsf{S}})$ for $\mathbb{M}_{\mathsf{top}}$ follows from Proposition 7.3.2. For (Inc), suppose $M = (X, P, V) \in \mathbb{M}_{\mathsf{top}}$, $x \in X$ and $M, x \models \mathsf{A}\varphi$. Then $\|\varphi\|_M = X$, so $\|\neg\varphi\|_M = \emptyset$. By the convention that the empty set is the empty union $\bigcup\emptyset$ (which is a finite union), we have $\emptyset \in P$. Taking $A = \emptyset$ in the definition of the semantics for S, we have $\|\neg\varphi\|_M \subseteq A$ but clearly $x \notin A$. Hence $M, x \not\models \mathsf{S}\neg\varphi$, so $M, x \models \hat{\mathsf{S}}\varphi$.

For completeness, we go via relational semantics using the translation $t : \mathcal{L} \to \mathcal{L}_{\mathsf{KA}}$ and Theorem 7.4.3. First, let $\mathsf{L}_{\mathsf{S4A}}$ be the logic of $\mathcal{L}_{\mathsf{KA}}$ formulas formed by the axioms and inference rules shown in Table 7.2. It is well known that $\mathsf{L}_{\mathsf{S4A}}$ is strongly complete with respect to $\mathbb{M}^*_{\mathsf{S4}}$ [12, Theorem 7.2].

Table 7.2: Axioms and inference rules for $\mathsf{L}_{\mathsf{S4A}}$.

$\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)$	$(\mathrm{K}_{\mathsf{K}})$
$\mathsf{K}\varphi \to \varphi$	$(\mathrm{T}_{\mathsf{K}})$
$\mathsf{K}\varphi \to \mathsf{K}\mathsf{K}\varphi$	(4_{K})
$\mathsf{A}(\varphi \to \psi) \to (\mathsf{A}\varphi \to \mathsf{A}\psi)$	$(\mathrm{K}_{\mathsf{A}})$
$\mathsf{A}\varphi \to \varphi$	$(\mathrm{T}_{\mathsf{A}})$
$\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi$	(5_{A})
$\mathsf{A}\varphi \to \mathsf{K}\varphi$	$(\mathrm{Inc}_{\mathsf{K}})$
From φ infer $\mathsf{A}\varphi$	$(\mathrm{Nec}_{\mathsf{A}})$
From $\varphi \to \psi$ and φ infer ψ	(MP)

Now, define a translation $u : \mathcal{L}_{\mathsf{KA}} \to \mathcal{L}$ as follows:

$$\begin{aligned}
u(p) &= p \\
u(\varphi \wedge \psi) &= u(\varphi) \wedge u(\psi) \\
u(\neg\varphi) &= \neg u(\varphi) \\
u(\mathsf{K}\varphi) &= \neg\mathsf{S}\neg u(\varphi) \\
u(\mathsf{A}\varphi) &= \mathsf{A}u(\varphi).
\end{aligned}$$

[10] Note that **KT4** is also complete for topological spaces with respect to the interior semantics [10].

Recall the translation $t : \mathcal{L} \to \mathcal{L}_{\mathsf{KA}}$ from Section 7.4. While u is not the inverse of t (for instance, there is no $\psi \in \mathcal{L}_{\mathsf{KA}}$ with $u(\psi) = \mathsf{E}p$), for any $\varphi \in \mathcal{L}$ we have that φ is $\mathsf{L_{top}}$-provably equivalent to $u(t(\varphi))$.

Claim 7.5.1. *Let $\varphi \in \mathcal{L}$. Then $\vdash_{\mathsf{L_{top}}} \varphi \leftrightarrow u(t(\varphi))$.*

Proof. By induction on $\mathcal{L}$ formulas. The cases of atomic propositions and propositional connectives are straightforward. For the other cases, first note that the "replacement of equivalents" rule is derivable in L (and thus in $\mathsf{L_{top}}$) for S, E and A:

$$\text{From } \varphi \leftrightarrow \psi \text{ infer } \bigcirc\varphi \leftrightarrow \bigcirc\psi \quad (\bigcirc \in \{\mathsf{S}, \mathsf{E}, \mathsf{A}\}).$$

For S this follows from ($\mathrm{Nec}_{\mathsf{A}}$) and ($\mathrm{W}_{\mathsf{S}}$); for E from ($\mathrm{Nec}_{\mathsf{A}}$) and ($\mathrm{RE}_{\mathsf{E}}$), and for A from ($\mathrm{Nec}_{\mathsf{A}}$) and ($\mathrm{K}_{\mathsf{A}}$). Now for the inductive step, suppose $\vdash_{\mathsf{L_{top}}} \varphi \leftrightarrow u(t(\varphi))$.

- S: Note that

  $$u(t(\mathsf{S}\varphi)) = u(\neg\mathsf{K}\neg t(\varphi)) = \neg\neg\mathsf{S}\neg\neg u(t(\varphi)).$$

  By the inductive hypothesis, propositional logic and replacement of equivalents, $\vdash_{\mathsf{L_{top}}} \mathsf{S}\varphi \leftrightarrow u(t(\mathsf{S}\varphi))$.

- E: We have

  $$\begin{aligned} u(t(\mathsf{E}\varphi)) &= u(\mathsf{A}(\neg t(\varphi) \to \mathsf{K}\neg t(\varphi))) \\ &= \mathsf{A}u(\neg t(\varphi) \to \mathsf{K}\neg t(\varphi)) \\ &= \mathsf{A}(u(\neg t(\varphi)) \to u(\mathsf{K}\neg t(\varphi))) \\ &= \mathsf{A}(\neg u(t(\varphi)) \to \neg\mathsf{S}\neg u(\neg t(\varphi))) \\ &= \mathsf{A}(\neg u(t(\varphi)) \to \neg\mathsf{S}\neg\neg u(t(\varphi))). \end{aligned}$$

  Taking the contrapositive of the implication, and using replacement of equivalents together with the inductive hypothesis, we get

  $$\vdash_{\mathsf{L_{top}}} u(t(\mathsf{E}\varphi)) \leftrightarrow \mathsf{A}(\mathsf{S}\varphi \to \varphi).$$

  But we have already seen that $\vdash_{\mathsf{L_{int}}} \mathsf{E}\varphi \leftrightarrow \mathsf{A}(\mathsf{S}\varphi \to \varphi)$; since $\mathsf{L_{top}}$ extends $\mathsf{L_{int}}$, we get $\vdash_{\mathsf{L_{top}}} \mathsf{E}\varphi \leftrightarrow u(t(\mathsf{E}\varphi))$.

- A: This case is straightforward by the inductive hypothesis and replacement of equivalents, since $u(t(\mathsf{A}\varphi)) = \mathsf{A}u(t(\varphi)$.

□

Next we show that if $\varphi \in \mathcal{L}_{\mathsf{KA}}$ is a theorem of $\mathsf{L_{S4A}}$, then $u(\varphi)$ is a theorem of $\mathsf{L_{top}}$.

Claim 7.5.2. *Let $\varphi \in \mathcal{L}_{\mathsf{KA}}$. Then $\vdash_{\mathsf{L_{S4A}}} \varphi$ implies $\vdash_{\mathsf{L_{top}}} u(\varphi)$.*

Proof. By induction on the length of $\mathsf{L}_{\mathsf{S4A}}$ proofs. The base case consists of showing that if φ is an instance of an $\mathsf{L}_{\mathsf{S4A}}$ axiom or a substitution instance of a propositional tautology, then $\vdash_{\mathsf{L}_{\mathsf{top}}} u(\varphi)$. The case for instances of tautologies is straightforward, since u does not affect the structure of a propositional formula. We take the axioms of $\mathsf{L}_{\mathsf{S4A}}$ in turn.

- (K_{K}): We have

$$\begin{aligned} &u(\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)) \\ &\qquad = \neg\mathsf{S}\neg(u(\varphi) \to u(\psi)) \to (\neg\mathsf{S}\neg u(\varphi) \to \neg\mathsf{S}\neg u(\psi)) \\ &\qquad = \hat{\mathsf{S}}(u(\varphi) \to u(\psi)) \to (\hat{\mathsf{S}}u(\varphi) \to \hat{\mathsf{S}}u(\psi)) \end{aligned}$$

  which is an instance of (K_{S}).

- (T_{K}): We have

$$u(\mathsf{K}\varphi \to \varphi) = \neg\mathsf{S}\neg u(\varphi) \to u(\varphi)$$

  Taking the contrapositive, this is $\mathsf{L}_{\mathsf{top}}$-provably equivalent to $\neg u(\varphi) \to \mathsf{S}\neg u(\varphi)$, which is an instance of (T_{S}).

- (4_{K}): We have

$$u(\mathsf{K}\varphi \to \mathsf{K}\mathsf{K}\varphi) = \neg\mathsf{S}\neg u(\varphi) \to \neg\mathsf{S}\neg\neg\mathsf{S}\neg u(\varphi)$$

  This is provably equivalent to $\mathsf{S}\mathsf{S}\neg u(\varphi) \to \mathsf{S}\neg u(\varphi)$, which is an instance of (4_{S}).

- (K_{A}): We have

$$u(\mathsf{A}(\varphi \to \psi) \to (\mathsf{A}\varphi \to \mathsf{A}\psi)) = \mathsf{A}(u(\varphi) \to u(\psi)) \to (\mathsf{A}u(\varphi) \to \mathsf{A}u(\psi))$$

  which is an instance of (K_{A}) in $\mathsf{L}_{\mathsf{top}}$.

- (T_{A}): We have

$$u(\mathsf{A}\varphi \to \varphi) = \mathsf{A}u(\varphi) \to u(\varphi)$$

  which is an instance of (T_{A}) in $\mathsf{L}_{\mathsf{top}}$.

- (5_{A}): We have

$$u(\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi) = \neg\mathsf{A}u(\varphi) \to \mathsf{A}\neg\mathsf{A}u(\varphi)$$

  which is an instance of (5_{A}) in $\mathsf{L}_{\mathsf{top}}$.

- ($\mathrm{Inc}_{\mathsf{K}}$): We have

$$u(\mathsf{A}\varphi \to \mathsf{K}\varphi) = \mathsf{A}u(\varphi) \to \neg\mathsf{S}\neg u(\varphi) = \mathsf{A}u(\varphi) \to \hat{\mathsf{S}}u(\varphi)$$

  which is an instance of (Inc).

For the inductive step, we show that for each inference rule $\frac{\psi_1,\ldots,\psi_n}{\varphi}$, if $\vdash_{\mathsf{L_{top}}} u(\psi_i)$ for each i then $\vdash_{\mathsf{L_{top}}} u(\varphi)$.

- $(\mathrm{Nec}_{\mathsf{A}})$: If $\vdash_{\mathsf{L_{top}}} u(\varphi)$, then from $(\mathrm{Nec}_{\mathsf{A}})$ in $\mathsf{L_{top}}$ we get $\vdash_{\mathsf{L_{top}}} \mathsf{A}u(\varphi)$. But $\mathsf{A}u(\varphi) = u(\mathsf{A}\varphi)$, so we are done.
- (MP): Similarly, this clear from (MP) for $\mathsf{L_{top}}$ and the fact that $u(\varphi \to \psi) = u(\varphi) \to u(\psi)$.

□

Claims 7.5.1 and 7.5.2 easily imply the following.

Claim 7.5.3. *Let $\varphi \in \mathcal{L}$. Then $\vdash_{\mathsf{L_{S4A}}} t(\varphi)$ implies $\vdash_{\mathsf{L_{top}}} \varphi$.*

Proof. Suppose $\vdash_{\mathsf{L_{S4A}}} t(\varphi)$. By Claim 7.5.2, $\vdash_{\mathsf{L_{top}}} u(t(\varphi))$. By Claim 7.5.1, $\vdash_{\mathsf{L_{top}}} \varphi \leftrightarrow u(t(\varphi))$. By (MP), $\vdash_{\mathsf{L_{top}}} \varphi$. □

We can now show strong completeness. Suppose $\Gamma \subseteq \mathcal{L}$, $\varphi \in \mathcal{L}$ and $\Gamma \models_{\mathbb{M}_{\mathsf{top}}} \varphi$. We claim $t(\Gamma) \models_{\mathbb{M}^*_{\mathsf{S4}}} t(\varphi)$. Indeed, if $M^* \in \mathbb{M}^*_{\mathsf{S4}}$ and x is a state in M^* with $M^*, x \models t(\psi)$ for all $\psi \in \Gamma$, then with f as in Theorem 7.4.3 we have $f^{-1}(M^*), x \models \psi$ for all $\psi \in \Gamma$. Since $f^{-1}(M^*) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}} \subseteq \mathbb{M}_{\mathsf{top}}$, $\Gamma \models_{\mathbb{M}_{\mathsf{top}}} \varphi$ gives $f^{-1}(M^*), x \models \varphi$, and thus $M^*, x \models t(\varphi)$.

By (strong) completeness of $\mathsf{L_{S4A}}$ for $\mathbb{M}^*_{\mathsf{S4}}$, we get $t(\Gamma) \vdash_{\mathsf{L_{S4A}}} t(\varphi)$. That is, there are $\psi_0, \ldots, \psi_n \in \Gamma$ such that $\vdash_{\mathsf{L_{S4A}}} t(\psi_0) \wedge \cdots \wedge t(\psi_n) \to t(\varphi)$. Since t passes over conjunctions and implications, this means $\vdash_{\mathsf{L_{S4A}}} t(\psi_0 \wedge \cdots \wedge \psi_n \to \varphi)$. By Claim 7.5.3, $\vdash_{\mathsf{L_{top}}} \psi_0 \wedge \cdots \wedge \psi_n \to \varphi$. Hence $\Gamma \vdash_{\mathsf{L_{top}}} \varphi$, and we are done. □

Just as the connection between S4 and $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$ allowed us to obtain a complete axiomatisation of $\mathbb{M}_{\mathsf{top}}$, we can axiomatise $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$ by considering S5. For brevity, write $\mathbb{M}_{\mathsf{int-compl}} = \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$. Let $\mathsf{L_{int-compl}}$ be the extension of $\mathsf{L_{top}}$ with the **5** axiom for $\hat{\mathsf{S}}$, which we present in the "diamond" form:

$$\mathsf{S}\neg\mathsf{S}\varphi \to \neg\mathsf{S}\varphi \quad (5_{\mathsf{S}})$$

Theorem 7.5.4. $\mathsf{L_{int-compl}}$ *is sound and strongly complete with respect to* $\mathbb{M}_{\mathsf{int-compl}}$.

Proof. For soundness, we need to check that (5_{S}) is valid on $\mathbb{M}_{\mathsf{int-compl}}$. Let $M = (X, P, V)$ be closed under intersections and complements, and suppose $M, x \models \mathsf{S}\neg\mathsf{S}\varphi$. Note that $\|\mathsf{S}\varphi\|_M = \bigcap\{A \in P \mid \|\varphi\|_M \subseteq A\}$ is an intersection from P, so $\|\mathsf{S}\varphi\|_M \in P$. By closure under complements, $\|\neg\mathsf{S}\varphi\|_M \in P$ too. Hence $M, x \models \mathsf{S}\neg\mathsf{S}\varphi \wedge \mathsf{E}\neg\mathsf{S}\varphi$. By Proposition 7.2.1 (4), we get $M, x \models \neg\mathsf{S}\varphi$.

The completeness proof goes in exactly the same way as Theorem 7.5.3. Letting $\mathsf{L_{S5A}}$ be the extension of $\mathsf{L_{S4A}}$ with the (5_{K}) axiom $\neg\mathsf{K}\varphi \to \mathsf{K}\neg\mathsf{K}\varphi$, it can be shown that $\mathsf{L_{S5A}}$ is strongly complete with respect to $\mathbb{M}^*_{\mathsf{S5}}$. With u as in the proof of Theorem 7.5.3, we have that $\vdash_{\mathsf{L_{S5A}}} \varphi$ implies $\vdash_{\mathsf{L_{int-compl}}} u(\varphi)$, for $\varphi \in \mathcal{L}_{\mathsf{KA}}$ (the only new part to check there is that $u(\neg\mathsf{K}\varphi \to \mathsf{K}\neg\mathsf{K}\varphi)$ is a theorem of $\mathsf{L_{int-compl}}$, but this follows from (5_{S})). The remainder of the proof goes through as before, this time appealing to the bijection $g : \mathbb{M}_{\mathsf{int-compl}} \to \mathbb{M}^*_{\mathsf{S5}}$. □

7.6 The Multi-source Case

So far we have been able to model the expertise of only a single source. In this section we generalise the setting to handle *multiple* sources. This allows us to consider not only the expertise of different sources individually, but also notions of *collective expertise*. For example, how may sources *combine* their expertise? Is there a suitable notion of *common expertise*? To answer these questions we take inspiration from the well-studied notions of *distributed knowledge* and *common knowledge* from epistemic logic [40], and establish connections between collective expertise and collective knowledge.

7.6.1 Collective Knowledge

Let $\mathcal{J}$ be a finite, non-empty set of sources. Turning briefly to epistemic logic interpreted under relational semantics, we recount several notions of collective knowledge. First, a *multi-source relational model* is a triple $M^* = (X, \{R_j\}_{j\in\mathcal{J}}, V)$, where R_j is a binary relation on X for each j. Consider the following knowledge operators [40]:

- $\mathsf{K}_j\varphi$ (individual knowledge): for $j \in J$ and a formula φ, set

$$M^*, x \models \mathsf{K}_j\varphi \iff \forall y \in X : xR_jy \implies M^*, y \models \varphi.$$

  This is the straightforward adaptation of knowledge in the single-source case to the multi-source setting.

- $\mathsf{K}^{\mathsf{dist}}_J\varphi$ (distributed knowledge): for $J \subseteq \mathcal{J}$ non-empty, set

$$M^*, x \models \mathsf{K}^{\mathsf{dist}}_J\varphi \iff \forall y \in X : (x, y) \in \bigcap_{j\in J} R_j \implies M^*, y \models \varphi.$$

  That is, knowledge of φ is distributed among the sources in J if, by combining their accessibility relations R_j, all states possible at x satisfy φ. Here the R_j are combined by taking their intersection: a state y is possible according to the group at x iff *every* source in J considers y possible at x.

- $\mathsf{K}^{\mathsf{sh}}_J\varphi$ (shared knowledge):[11] for $J \subseteq \mathcal{J}$ non-empty, set

$$M^*, x \models \mathsf{K}^{\mathsf{sh}}_J\varphi \iff \forall j \in J : M^*, x \models \mathsf{K}_j\varphi.$$

That is, a group J have shared knowledge of φ exactly when each agent in J knows φ. Thus we have $\mathsf{K}^{\mathsf{sh}}_J\varphi \equiv \bigwedge_{j\in J}\mathsf{K}_j\varphi$.

- $\mathsf{K}^{\mathsf{com}}_J\varphi$ (common knowledge): write $\mathsf{K}^1_J\varphi$ for $\mathsf{K}^{\mathsf{sh}}_J\varphi$, and for $n\in\mathbb{N}$ write $\mathsf{K}^{n+1}_J\varphi$ for $\mathsf{K}^{\mathsf{sh}}_J\mathsf{K}^n_J\varphi$. Then

$$M^*, x \models \mathsf{K}^{\mathsf{com}}_J\varphi \iff \forall n\in\mathbb{N} : M^*, x \models \mathsf{K}^n_J\varphi.$$

  Here $\mathsf{K}^1_J\varphi$ says that everyone in J knows φ, $\mathsf{K}^2_J\varphi$ says that everybody in J knows that everybody in J knows φ, and so on. There is common knowledge of φ among J if this nesting of "everybody knows" holds for any order n.

In what follows we write $\mathcal{L}^{\mathcal{J}}_{\mathsf{KA}}$ for the language formed from Prop with knowledge operators K_j, $\mathsf{K}^{\mathsf{dist}}_J$, $\mathsf{K}^{\mathsf{sh}}_J$ and $\mathsf{K}^{\mathsf{com}}_J$, for $j\in\mathcal{J}$ and $J\subseteq\mathcal{J}$ non-empty, and the universal modality A.

7.6.2 Collective Expertise

Returning to expertise semantics, define a *multi-source expertise model* as a triple $M = (X, \{P_j\}_{j\in\mathcal{J}}, V)$, where $P_j \subseteq 2^X$ is the collection of expertise sets for source j. Say M is closed under intersections, unions, complements etc. if each P_j is. Since the connection between expertise and S4 knowledge (Theorem 7.4.3) holds for expertise models closed under unions and intersections, we restrict attention to this class of (multi-source) models in this section.

The counterpart of individual knowledge – individual expertise – is straightforward: we may simply introduce expertise and soundness operators E_j and S_j for each source $j\in\mathcal{J}$, and interpret $\mathsf{E}_j\varphi$ and $\mathsf{S}_j\varphi$ as in the single-source case using P_j. For notions of collective expertise and soundness, we define new collections P_J by combining the P_j in an appropriate way.

Distributed Expertise For distributed expertise, the intuition is clear: the sources in a group J should combine their expertise collections P_j to form a larger collection P^{dist}_J. A first candidate for P^{dist}_J would therefore be $\bigcup_{j\in J}P_j$. However, since we assume each P_j is closed under unions and intersections, we suppose that each source j has the cognitive or computational capacity to combine expertise sets $A\in P_j$ by taking unions or intersections. We argue that the same should be possible for the group J as a whole, and therefore let P^{dist}_J be the closure of $\bigcup_{j\in J}P_j$ under unions and intersections:

$$P^{\mathsf{dist}}_J = \bigcap\left\{P' \supseteq \bigcup_{j\in J}P_j \mid P' \text{ is closed under unions and intersections}\right\}.$$

[11]In Fagin et al. [40], shared knowledge is denoted $\mathsf{E}_J\varphi$ for "everybody knows φ". We opt to use the term "shared" knowledge to avoid conflict with our notation for expertise.

Note that P_J^{dist} is closed under unions and intersections, and $P_j \subseteq P_J^{\mathsf{dist}}$ for all $j \in J$ (in fact, P_J^{dist} is the smallest set with these properties). While P_J^{dist} depends on the model M, we suppress this from the notation.

Now, recall from Theorem 7.4.3 that our semantics for expertise and soundness is connected to relational semantics via the mapping $P \mapsto R_P$ (Definition 7.4.2). The following result shows that P_J^{dist} corresponds to distributed knowledge under this mapping. For ease of notation, write R_J^{dist} for $R_{P_J^{\mathsf{dist}}}$ and R_j for R_{P_j}.

Proposition 7.6.1. *For any multi-source expertise model M and $J \subseteq \mathcal{J}$ non-empty,*

$$R_J^{\mathsf{dist}} = \bigcap_{j \in J} R_j.$$

Proof. "$\subseteq$": Suppose $xR_J^{\mathsf{dist}}y$. Let $j \in J$. We need to show xR_jy. Take any $A \in P_j$ such that $y \in A$. Then $A \in P_J^{\mathsf{dist}}$, so $xR_J^{\mathsf{dist}}y$ gives $x \in A$. Hence xR_jy.

"$\supseteq$": Suppose $(x, y) \in \bigcap_{j \in J} R_j$, i.e. xR_jy for all $j \in J$. Set

$$P' = \{A \in P_J^{\mathsf{dist}} \mid y \in A \implies x \in A\} \subseteq P_J^{\mathsf{dist}}.$$

Then $P' \supseteq \bigcup_{j \in J} P_j$, since if $j \in J$ and $A \in P_j$ then $A \in P_J^{\mathsf{dist}}$ and $y \in A$ implies $x \in A$ by xR_jy. We claim P' is closed under intersections. Suppose $\{A_i\}_{i \in I} \subseteq P'$ and write $A = \bigcap_{i \in I} A_i$. Since $P' \subseteq P_J^{\mathsf{dist}}$ and P_J^{dist} is closed under intersections, $A \in P_J^{\mathsf{dist}}$. Suppose $y \in A$. Then $y \in A_i$ for each i, so $x \in A_i$ by the defining property of P'. Hence $x \in \bigcap_{i \in I} A_i = A$. This shows $A \in P'$ as desired. A similar argument shows that P' is also closed under unions.

We see from the definition of P_J^{dist} that $P_J^{\mathsf{dist}} \subseteq P'$, so in fact $P' = P_J^{\mathsf{dist}}$. It now follows that $xR_J^{\mathsf{dist}}y$: for any $A \in P_J^{\mathsf{dist}}$ with $y \in A$ we have $A \in P'$, so $x \in A$ also. □

Common Expertise Common expertise admits a straightforward definition: simply take the expertise sets in common with all P_j:

$$P_J^{\mathsf{com}} = \bigcap_{j \in J} P_j.$$

If each P_j is closed under unions and intersections, then so too is P_J^{com}.

At first this may appear *too* straightforward. The form of the definition is closer to *shared* knowledge than to common knowledge. But in fact, shared knowledge has *no* expertise counterpart which admits the type of connection established in Theorem 7.4.3. Indeed, shared knowledge may fail positive introspection (axiom **4**: $\mathsf{K}\varphi \to \mathsf{K}\mathsf{K}\varphi$), but we have seen that the knowledge derived from expertise and soundness satisfies S4 (when the collection of expertise sets is closed under unions and complements).

However, this problem is only apparent in the translation of $\mathsf{S}\varphi$ as $\neg\mathsf{K}\neg\varphi$. For our translation of $\mathsf{E}\varphi$ as $\mathsf{A}(\neg\varphi \to \mathsf{K}\neg\varphi)$, the universal quantification via A dissolves the differences between shared and common knowledge.

Proposition 7.6.2. *Let $\varphi \in \mathcal{L}_{\mathsf{KA}}^{\mathcal{J}}$ and let $J \subseteq \mathcal{J}$ be non-empty. Then*

$$\mathsf{A}(\neg\varphi \to \mathsf{K}_J^{\mathsf{com}}\neg\varphi) \equiv \mathsf{A}(\neg\varphi \to \mathsf{K}_J^{\mathsf{sh}}\neg\varphi).$$

Proof. Let $M^* = (X, \{R_j\}_{j\in\mathcal{J}}, V)$ be a multi-source relational model. Since $\mathsf{K}_J^{\mathsf{com}}\psi \to \mathsf{K}_J^{\mathsf{sh}}\psi$ is valid for any ψ, the left-to-right implication of the above equivalence is straightforward.

For the right-to-left implication, suppose $M^*, x \models \mathsf{A}(\neg\varphi \to \mathsf{K}_J^{\mathsf{sh}}\neg\varphi)$. We show by induction that $M^*, x \models \mathsf{A}(\neg\varphi \to \mathsf{K}_J^{n}\neg\varphi)$ for all $n \in \mathbb{N}$, from which the result follows.

The base case $n = 1$ is given, since $\mathsf{K}_J^{1}\neg\varphi = \mathsf{K}_J^{\mathsf{sh}}\neg\varphi$. For the inductive step, suppose $M^*, x \models \mathsf{A}(\neg\varphi \to \mathsf{K}_J^{n}\neg\varphi)$. Take $y \in X$ such that $M^*, y \models \neg\varphi$. Let $j \in J$. Take $z \in X$ such that yR_jz. From the initial assumption we have $M^*, y \models \mathsf{K}_J^{\mathsf{sh}}\neg\varphi$, so $M^*, y \models \mathsf{K}_j\neg\varphi$ and thus $M^*, z \models \neg\varphi$. By the inductive hypothesis, $M^*, z \models \mathsf{K}_J^{n}\neg\varphi$. This shows that $M^*, y \models \mathsf{K}_j\mathsf{K}_J^{n}\neg\varphi$ for all $j \in J$, and thus $M^*, y \models K_J^{n+1}\neg\varphi$. Hence $M^*, x \models \mathsf{A}(\neg\varphi \to \mathsf{K}_J^{n+1}\neg\varphi)$ as required. □

Proposition 7.6.2 shows that when interpreting collective expertise on φ as collective refutation of φ whenever φ is false, there is no difference between using common knowledge and just shared knowledge.

We now confirm that P_J^{com} does indeed correspond to common knowledge. First we recall a well-known result from Fagin et al. [40]. In what follows, write $R^+ = \bigcup_{n\in\mathbb{N}} R^n$ for the transitive closure of R.

Lemma 7.6.1 (Fagin et al. [40], Lemma 2.2.1)**.** *Let $M^* = (X, \{R_j\}_{j\in\mathcal{J}}, V)$ be a multi-source relational model and $J \subseteq \mathcal{J}$ non-empty. Write $R' = \left(\bigcup_{j\in J} R_j\right)^+$. Then for all $x \in X$ and $\varphi \in \mathcal{L}_{\mathsf{KA}}^{\mathcal{J}}$:*

$$M^*, x \models \mathsf{K}_J^{\mathsf{com}}\varphi \iff \forall y \in X : xR'y \implies M^*, y \models \varphi.$$

By Lemma 7.6.1, common knowledge has an interpretation in terms of the usual relational semantics for knowledge, where we use the transitive closure of the union of the accessibility relations of the sources in J. Writing R_J^{com} for $R_{P_J^{\mathsf{com}}}$, we have the following.

Proposition 7.6.3. *Let M be a multi-source model closed under unions and intersections. Then for $J \subseteq \mathcal{J}$ non-empty, $R_J^{\mathsf{com}} = \left(\bigcup_{j\in J} R_j\right)^+$.*

Proof. Write $R' = (\bigcup_{j\in J} R_j)^+$. Note that R_J^{com} is reflexive and transitive by Lemma 7.4.2 (1). R' is transitive by its definition as a transitive closure, and reflexive since each R_j is (and $J \neq \emptyset$). It is therefore sufficient by Lemma 7.4.1 to show that any set is downwards closed wrt R_J^{com} iff it is downwards closed wrt R'. Since each P_j is closed under unions and intersections, so too is P_J^{com}. Using Lemma 7.4.2 (2), we have

$$\begin{aligned} A \text{ downwards closed wrt } R_J^{\mathsf{com}} &\iff A \in P_J^{\mathsf{com}} \\ &\iff \forall j \in J : A \in P_j \\ &\iff \forall j \in J : A \text{ downwards closed wrt } R_j \\ &\iff A \text{ downwards closed wrt } \bigcup_{j \in J} R_j \\ &\iff A \text{ downwards closed wrt } R' \end{aligned}$$

where the last step uses the fact that A is downwards closed with respect to some relation if and only if it is downwards closed with respect to the transitive closure. This completes the proof. □

Collective semantics We now formally define the syntax and semantics of collective expertise. Let $\mathcal{L}^{\mathcal{J}}$ be the language defined by the following grammar:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \mathsf{E}_j\varphi \mid \mathsf{S}_j\varphi \mid \mathsf{E}_J^g\varphi \mid \mathsf{S}_J^g\varphi \mid \mathsf{A}\varphi$$

for $p \in \mathsf{Prop}$, $j \in \mathcal{J}$, $g \in \{\mathsf{dist}, \mathsf{com}\}$ and $J \subseteq \mathcal{J}$ non-empty. For a multi-source expertise model $M = (X, \{P_j\}_{j \in \mathcal{J}}, V)$, define the satisfaction relation as before for atomic propositions, propositional connectives and A, and set

$$\begin{array}{llll} M, x & \models E_j\varphi & \iff \|\varphi\|_M \in P_j & \\ M, x & \models E_J^g\varphi & \iff \|\varphi\|_M \in P_J^g & (g \in \{\mathsf{dist}, \mathsf{com}\}) \\ M, x & \models S_j\varphi & \iff \forall A \in P_J : \|\varphi\|_M \subseteq A \implies x \in A & \\ M, x & \models S_J^g\varphi & \iff \forall A \in P_J^g : \|\varphi\|_M \subseteq A \implies x \in A & (g \in \{\mathsf{dist}, \mathsf{com}\}) \end{array}$$

Note that expertise and soundness are interpreted as before, but with respect to different collections P. Consequently, the interactions shown in Proposition 7.2.1 still hold for individual and collective notions of expertise and soundness.

Example 7.6.1. *Extending Examples 7.2.1 and 7.2.2, consider* $\mathcal{J} = \{\mathsf{econ}, \mathsf{dr}, \mathsf{analyst}\}$, *where* econ *is the economist,* dr *is a doctor with expertise on i only, and* analyst *has access to aggregate data distinguishing three levels of virus activity: minimal ($\neg i \wedge \neg d$), high ($(i \vee d) \wedge \neg(i \wedge d)$) and very high ($i \wedge d$). This can be modelled by a multi-source model M with X, V and P_{econ} as in Example 7.2.2, and* $P_{\mathsf{dr}} = \{\emptyset, X, \{ipd, ip, id, i\}, \{pd, p, d, \emptyset\}\}$, P_{analyst} *is the closure under unions of* $\{\emptyset, X, \{ipd, id\}, \{ip, pd, i, d\}, \{p, \emptyset\}\}$.

Note that neither dr *nor* analyst *have expertise on d individually. However, if* dr *can communicate whether or not i holds, this gives* analyst *enough information to disambiguate the "high activity" case and therefore determine d. Indeed, we have* $\|d\| = \|i \wedge d\| \cup (\|i \vee d\| \setminus \|i \wedge d\| \cap \|\neg i\|)$, *which is formed by unions and intersections from* $P_{\mathsf{dr}} \cup P_{\mathsf{analyst}}$, *and thus* $\|d\| \in P^{\mathsf{dist}}_{\{\mathsf{dr},\mathsf{analyst}\}}$. *Hence* $M \models \mathsf{E}^{\mathsf{dist}}_{\{\mathsf{dr},\mathsf{analyst}\}} d$. *Similarly,* dr *and* analyst *have distributed* *expertise on $\neg d$. Bringing back* econ*, the grand coalition $\mathcal{J}$ have distributed expertise on the original report $p \wedge \neg d$ from Example 7.2.1. Consequently, the report is no longer sound at "actual" state idp: all sources together have sufficient expertise to know it is false.*

The following validities express properties specific to collective expertise.

Proposition 7.6.4. *The following formulas are valid.*

1. *For $j \in J$, $\mathsf{E}_j\varphi \to \mathsf{E}^{\mathsf{dist}}_J\varphi$*
2. $\mathsf{E}^{\mathsf{com}}_J\varphi \leftrightarrow \bigwedge_{j\in J}\mathsf{E}_j\varphi$
3. $\mathsf{S}^{\mathsf{com}}_J\varphi \leftrightarrow \bigvee_{j\in J}\mathsf{S}_j\mathsf{S}^{\mathsf{com}}_J\varphi$
4. $\mathsf{E}^{\mathsf{dist}}_{\{j\}}\varphi \leftrightarrow \mathsf{E}_j\varphi$ *is valid on $\mathbb{M}^{\mathcal{J}}_{\mathsf{int}} \cap \mathbb{M}^{\mathcal{J}}_{\mathsf{unions}}$*

Proof. We prove only (3); the others are straightforward. The right implication is valid since $\psi \to \mathsf{S}_j\psi$ is, with ψ set to $\mathsf{S}^{\mathsf{com}}_J\varphi$ and $j \in J$ arbitrary (recall J is non-empty). For the left implication, suppose there is $j \in J$ with $M, x \models \mathsf{S}_j\mathsf{S}^{\mathsf{com}}_J\varphi$. Then $x \in \bigcap\{A \in P_j \mid \|\mathsf{S}^{\mathsf{com}}_J\varphi\|_M \subseteq A\}$. Now take $B \in P^{\mathsf{com}}_J$ such that $\|\varphi\|_M \subseteq B$. Note that if $y \in \|\mathsf{S}^{\mathsf{com}}_J\varphi\|$ then $y \in B$ by the definition of the semantics for $\mathsf{S}^{\mathsf{com}}_J$, so $\|\mathsf{S}^{\mathsf{com}}_J\varphi\|_M \subseteq B$. Since $B \in P^{\mathsf{com}}_J \subseteq P_j$, we get $x \in B$. This shows $M, x \models \mathsf{S}^{\mathsf{com}}_J\varphi$. □

Validity (3) comes from the *fixed-point axiom* for common knowledge: $\mathsf{K}^{\mathsf{com}}_J\varphi \leftrightarrow \mathsf{K}^{\mathsf{sh}}_J(\varphi \wedge \mathsf{K}^{\mathsf{com}}_J\varphi)$. Our version says $\mathsf{S}^{\mathsf{com}}_J\varphi$ is a fixed-point of the function $\theta \mapsto \bigvee_{j\in J}\mathsf{S}_j\theta$. In words, φ is true up to lack of *common* expertise iff there is some source for whom $\mathsf{S}^{\mathsf{com}}_J\varphi$ is true up to their lack of (individual) expertise.

As promised, there is a tight link between our notions of collective expertise and knowledge. Define a translation $t : \mathcal{L}^{\mathcal{J}} \to \mathcal{L}^{\mathcal{J}}_{\mathsf{KA}}$ inductively by

$$\begin{aligned}
t(\mathsf{E}_j\varphi) &= \mathsf{A}(\neg t(\varphi) \to \mathsf{K}_j\neg t(\varphi)) \\
t(\mathsf{E}^g_J\varphi) &= \mathsf{A}(\neg t(\varphi) \to \mathsf{K}^g_J\neg t(\varphi)) \quad (g \in \{\mathsf{dist}, \mathsf{com}\}) \\
t(\mathsf{S}_j\varphi) &= \neg\mathsf{K}_j\neg t(\varphi) \\
t(\mathsf{S}^g_J\varphi) &= \neg\mathsf{K}^g_J\neg t(\varphi) \qquad\qquad\quad (g \in \{\mathsf{dist}, \mathsf{com}\})
\end{aligned}$$

where the other cases are as for t in Section 7.4. This is essentially the same translation as before, but with the various types of expertise and soundness matched with their knowledge counterparts. We have an analogue of Theorem 7.4.3.

Theorem 7.6.1. *The mapping $f : \mathbb{M}^{\mathcal{J}}_{\mathsf{int}} \cap \mathbb{M}^{\mathcal{J}}_{\mathsf{unions}} \to \mathbb{M}^{\mathcal{J}}_{\mathsf{S4}}$ given by $(X, \{P_j\}_{j\in\mathcal{J}}, V) \mapsto (X, \{R_{P_j}\}_{j\in\mathcal{J}}, V)$ is bijective, and for $x \in X$ and $\varphi \in \mathcal{L}^{\mathcal{J}}$:*

$$M, x \models \varphi \iff f(M), x \models t(\varphi).$$

Moreover, the restriction of this map to $\mathbb{M}^{\mathcal{J}}_{\mathsf{int}} \cap \mathbb{M}^{\mathcal{J}}_{\mathsf{compl}}$ is a bijection into $\mathbb{M}^{\mathcal{J}}_{\mathsf{S5}}$.

Proof. That the map is bijective follows easily from Theorems 7.4.1 and 7.4.2. For the stated property we proceed by induction on $\mathcal{L}^{\mathcal{J}}$ formulas. As in Theorem 7.4.3, the cases for atomic propositions, propositional connectives and A are straightforward. For expertise and soundness, the argument in the proof of Theorem 7.4.3 showed that $\mathsf{E}\varphi$ and $\mathsf{S}\varphi$ interpreted via some collection P is equivalent to $t(\mathsf{E}\varphi)$ and $t(\mathsf{S}\varphi)$ interpreted wrt relational semantics via R_P. It is therefore sufficient to show that for each notion of individual and collective expertise interpreted in M via P, its corresponding notion of individual or collective knowledge (used in the translation t) is interpreted in $f(M)$ via R_P. This is self-evident for individual expertise. For distributive expertise this was shown in Proposition 7.6.1. For common expertise this was shown in Lemma 7.6.1 and Proposition 7.6.3. □

Theorem 7.6.1 can be used to adapt any sound and complete axiomatisation for $\mathbb{M}^{\mathcal{J}}_{\mathsf{S4}}$ (resp., $\mathbb{M}^{\mathcal{J}}_{\mathsf{S5}}$) over the language $\mathcal{L}^{\mathcal{J}}_{\mathsf{KA}}$ to obtain an axiomatisation for $\mathbb{M}^{\mathcal{J}}_{\mathsf{int}} \cap \mathbb{M}^{\mathcal{J}}_{\mathsf{unions}}$ (resp., $\mathbb{M}^{\mathcal{J}}_{\mathsf{int}} \cap \mathbb{M}^{\mathcal{J}}_{\mathsf{compl}}$) over $\mathcal{L}^{\mathcal{J}}$, in the same way as we did earlier when adapting S4 and S5 in Theorems 7.5.3 and 7.5.4.

7.7 Dynamic Extension

So far our picture has been entirely static. We cannot speak of expertise changing over time, nor of the information in a model changing via announcements from sources. To remedy this, we extend the framework with two *dynamic* operators: one to account for *increases in expertise* – e.g. after a process of learning or acquisition of new evidence – and one to model *sound announcements*. For simplicity, we return to the single-source case.

7.7.1 Expertise Increase

As a source interacts with the world over time, they may learn to make more distinctions between possible states of the world, and thereby increase their expertise. Leaving the particulars of the learning mechanism unspecified, we study only the end result: the source's expertise collection P is expanded to include a new set A.

However, this may not be so simple as setting $P' = P \cup \{A\}$ in light of the closure properties that may be imposed P. As remarked in Section 7.3, closure conditions correspond to assumptions about the source's cognitive or computational capabilities. It seems natural that if the source has the ability to combine sets in P by taking intersections, for example, then they should also be able to do after the learning, i.e. P' should also be closed under intersections. Thus, the new collection P' should inherit any closure properties from P, while extending $P \cup \{A\}$. In principle, we could therefore consider an expertise increase operation for *each* combination of closure properties.

For concreteness we will not do this, and will instead focus on the class $\mathbb{M}_{\mathsf{int}}$ of models closed under intersections. Conceptually, this is a minimal requirement, since we argued in section Section 7.3 that closure under intersections is a natural property. There are also technical advantages: we will later show that closure under intersections allows us to find reduction axioms which allow the formulas involving expertise increase to be equivalently expressed in the static language.

Definition 7.7.1. *Given an expertise model $M = (X, P, V)$ and a formula φ, define the model $M^{+\varphi} = (X, P^{+\varphi}, V)$ by setting*

$$P^{+\varphi} = \left\{ \bigcap \mathcal{A} \mid \mathcal{A} \subseteq P \cup \{\|\varphi\|_M\} \right\}.$$

That is, $P^{+\varphi}$ is obtained by adding $\|\varphi\|_M$ to P and closing under intersections.

Syntactically, we introduce formulas of the form $[+\varphi]\psi$, which are to be read as "ψ holds after the source gains expertise on φ". The truth condition for $[+\varphi]\psi$ in a model M is defined in terms of $M^{+\varphi}$:

$$M, x \models [+\varphi]\psi \iff M^{+\varphi}, x \models \psi.$$

If $\mathcal{L}_0$ denotes the propositional language built from Prop, then $[+\alpha]\mathsf{E}\alpha$ is valid for all $\alpha \in \mathcal{L}_0$. That is, expertise increase is successful for any propositional formula. However, this is not the case for general formulas $\varphi \in \mathcal{L}$. This comes from the fact that expertise is represented *semantically* via sets of states. The operator $[+\varphi]$ represents the source obtaining expertise on the set of φ states, where φ is interpreted *before the increase took place*. If φ refers to expertise (with E or S) then the meaning of φ may change after the increase. For example, consider the model $M = (X, P, V)$ with

$$\begin{aligned} X &= \{1, 2, 3, 4\} \\ P &= \{\emptyset, X, \{1, 3\}\} \\ V(p) &= \{1\} \\ V(q) &= \{2, 3\} \end{aligned}$$

Then, with $\varphi = p \vee (q \wedge \neg \mathsf{S}p)$ we have $M, 1 \not\models [+\varphi]\mathsf{E}\varphi$.[12] This counterexample is reminiscent of *Moore sentences* as formalised in Dynamic Epistemic Logic; e.g. an agent cannot know $p \wedge \neg \mathsf{K}p$ ("p is true but I do not know it") after this is truthfully announced [9].

Next we give reduction axioms to express any formula involving $[+\varphi]$ by an equivalent formula in the static language $\mathcal{L}$.

[12] In detail, we have $\|\varphi\|_M = \{1, 2\}$, so $P^{+\varphi} = \{\emptyset, X, \{1, 3\}, \{1, 2\}, \{1\}\}$. Then $\|\varphi\|_{M^{+\varphi}} = \{1, 2, 3\} \notin P^{+\varphi}$, so $M^{+\varphi}, 1 \not\models \mathsf{E}\varphi$.

Proposition 7.7.1. *The following formulas are valid on* $\mathbb{M}$*:*

$$\begin{array}{rcl} p & \leftrightarrow & [+\varphi]p \\ [+\varphi](\psi \wedge \theta) & \leftrightarrow & [+\varphi]\psi \wedge [+\varphi]\theta \\ [+\varphi]\neg\psi & \leftrightarrow & \neg[+\varphi]\psi \\ [+\varphi]\mathsf{A}\psi & \leftrightarrow & \mathsf{A}[+\varphi]\psi \\ [+\varphi]\mathsf{S}\psi & \leftrightarrow & \mathsf{S}[+\varphi]\psi \wedge (\mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi) \\ [+\varphi]\mathsf{E}\psi & \leftrightarrow & \mathsf{A}((\mathsf{S}[+\varphi]\psi \wedge (\mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi)) \to [+\varphi]\psi) \end{array}$$

Proof. The cases for atomic propositions, propositional connectives and A are straightforward. We show the reduction axiom for S. Let $M = (X, P, V)$ be a model and $x \in X$.

"$\to$": Suppose $M, x \models [+\varphi]\mathsf{S}\psi$. Then $M^{+\varphi}, x \models \mathsf{S}\psi$. Hence

$$x \in \bigcap\{A \in P^{+\varphi} \mid \|\psi\|_{M^{+\varphi}} \subseteq A\} \quad (*)$$

Note that $\|\psi\|_{M^{+\varphi}} = \|[+\varphi]\psi\|_M$. Now take $A \in P$ such that $\|[+\varphi]\psi\|_M \subseteq A$. Since $P \subseteq P^{+\varphi}$, we get $x \in A$ from $(*)$. Hence $M, x \models \mathsf{S}[+\varphi]\psi$.

Now suppose $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi)$. Then $\|[+\varphi]\psi\|_M \subseteq \|\varphi\|_M$, so $\|\psi\|_{M^{+\varphi}} \subseteq \|\varphi\|_M$. Since $\|\varphi\|_M \in P^{+\varphi}$, we get $x \in \|\varphi\|_M$ from $(*)$, i.e. $M, x \models \varphi$ as required.

"$\leftarrow$": Suppose $M, x \models \mathsf{S}[+\varphi]\psi$ and $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi$. Take $A \in P^{+\varphi}$ such that $\|\psi\|_{M^{+\varphi}} \subseteq A$. Then $\|[+\varphi]\psi\|_M \subseteq A$. By definition of $P^{+\varphi}$, there is a collection $\mathcal{A} \subseteq P \cup \{\|\varphi\|_M\}$ such that $A = \bigcap\mathcal{A}$. Let $B \in \mathcal{A}$. If $B \in P$, then $\|[+\varphi]\psi\|_M \subseteq A \subseteq B$ and $M, x \models \mathsf{S}[+\varphi]\psi$ give $x \in B$. Otherwise, $B = \|\varphi\|_M$. Hence $\|[+\varphi]\psi\|_M \subseteq \|\varphi\|_M$, so $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi)$. By the second assumption, we get $M, x \models \varphi$, i.e. $x \in \|\varphi\|_M = B$. We have now shown that $x \in \bigcap\mathcal{A} = A$, and thus $M^{+\varphi}, x \models \mathsf{S}\psi$ and $M, x \models [+\varphi]\mathsf{S}\psi$.

For the reduction axiom for E, note that since $M^{+\varphi} \in \mathbb{M}_{\mathsf{int}}$ we have $M^{+\varphi}, x \models E\psi$ iff $M^{+\varphi}, x \models \mathsf{A}(\mathsf{S}\psi \to \psi)$. Using the reduction axioms for A and S (and the reduction axiom for the implication, derived from those for $\neg$ and $\wedge$), we obtain the desired equivalence. $\square$

Note that only the reduction axiom for $[+\varphi]\mathsf{E}\psi$ requires $M^{+\varphi}$ to be closed under intersections.

7.7.2 Sound Announcements

In logics of public announcement [84, 29], the dynamic operator $[!\varphi]$ represents a *public* and *truthful* announcement of φ; the formula $[!\varphi]\psi$ is read as "after φ is announced, ψ holds". Such an announcement changes the information available in a model: after the announcement, all $\neg\varphi$ states are eliminated.

Since the premise of our work is to deal with non-expert sources, the truthfulness requirement is too strong for an announcement operator in our setting. Instead, we consider *sound announcements*: the source may announce φ whenever φ is sound at the current state. That is, the source may announce any (possibly false) statement which is true up to their lack of expertise.

Such an announcement is denoted syntactically by $[?\varphi]$. As with the expertise increase operator, we define a model update operation $M \mapsto M^{?\varphi}$.

It is clear how one should define new set of states: since the announcement tells us φ is sound, we eliminate unsound states by setting $X^{?\varphi} = \|\mathsf{S}\varphi\|_M$. The valuation is also straightforward, since announcements should not change the meaning of atomic propositions.

What about the new expertise collection $P^{?\varphi}$? If we restrict attention to models closed under intersections, as we did for expertise increase, then a natural choice is to simply restrict each $A \in P$ to $X^{?\varphi}$ by intersection. Since $X^{?\varphi} = \|\mathsf{S}\varphi\|_M = \bigcap\{B \in P \mid \|\varphi\|_M \subseteq B\}$, by the closure property we will have $P^{?\varphi} \subseteq P$, so that announcements do not increase expertise. This assumption will also permit us to find reduction axioms later on.

Definition 7.7.2. *Let $M = (X, P, V)$ be an expertise model. For a formula φ, define the model $M^{?\varphi} = (X^{?\varphi}, P^{?\varphi}, V^{?\varphi})$ by setting*

$$
\begin{aligned}
X^{?\varphi} &= \|\mathsf{S}\varphi\|_M \\
P^{?\varphi} &= \{A \cap X^{?\varphi} \mid A \in P\} \\
V^{?\varphi} &= V(p) \cap X^{?\varphi}
\end{aligned}
$$

Semantically, the truth condition for $[?\varphi]\psi$ is as follows.

$$M, x \models [?\varphi]\psi \iff M, x \models \mathsf{S}\varphi \implies M^{?\varphi}, x \models \psi.$$

Here we have the precondition that $\mathsf{S}\varphi$ is true: if φ is unsound, $[?\varphi]\psi$ is true for *any* ψ. Note that a sound announcement of φ can also be seen as a public (*truthful*) announcement of $\mathsf{S}\varphi$.

Example 7.7.1. *The report of the economist in Example 7.2.1 can be modelled as $[?(p \wedge \neg d)]$. Note that, with M as in Example 7.2.2, $\|\mathsf{S}(p \wedge \neg d)\|_M = \|p\|_M$. The updated model $M^{?(p\wedge\neg d)}$ therefore consists only of the bottom half of M as shown in Fig. 7.1. We see that $M, idp \models [?(p \wedge \neg d)]d$ – showing that even propositional announcements can "fail" due to lack of expertise – and $M \models [?(p \wedge \neg d)]\mathsf{A}p$ – showing that the parts of the report on which the sources does have expertise are* always *true after their announcement.*

As with the expertise increase operator, sound announcements remain sound for purely propositional formulas $\alpha \in \mathcal{L}_0$: $[?\alpha]\mathsf{S}\alpha$ is valid on $\mathbb{M}_{\mathsf{int}}$. This is not true for general formulas $\varphi \in \mathcal{L}$ which may refer to expertise itself. For example, in the model $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}}$ given by $X = \{1, 2, 3, 4\}$, $P = \{\emptyset, X, \{1\}, \{2\}, \{1, 2, 3\}\}$, $V(p) = \{1, 2\}$ and $V(q) = \{2, 4\}$, with $\varphi = p \wedge \neg \mathsf{E} q$ we have $M, 1 \not\models [?\varphi]\mathsf{S}\varphi$.

The following reduction axioms allow formulas involving announcements to be expressed in the static language.

Proposition 7.7.2. *The following formulas are valid on* $\mathbb{M}$*:*

$$\begin{aligned} p &\leftrightarrow \mathsf{S}\varphi \to p \\ [?\varphi](\psi \wedge \theta) &\leftrightarrow [?\varphi]\psi \wedge [?\varphi]\theta \\ [?\varphi]\neg\psi &\leftrightarrow \mathsf{S}\varphi \to \neg[?\varphi]\psi \\ [?\varphi]\mathsf{A}\psi &\leftrightarrow \mathsf{S}\varphi \to \mathsf{A}[?\varphi]\psi \\ [?\varphi]\mathsf{S}\psi &\leftrightarrow \mathsf{S}\varphi \to \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi) \end{aligned}$$

and the following is valid on $\mathbb{M}_{\mathsf{int}}$*:*

$$[?\varphi]\mathsf{E}\psi \quad \leftrightarrow \quad \mathsf{S}\varphi \to \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$$

Proof. The cases of atomic propositions, propositional connectives and the universal modality A are straightforward.

For the reduction axiom for S, first note that $\|\psi\|_{M^{?\varphi}} = \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M$. We need to show that $M, x \models [?\varphi]\mathsf{S}\psi$ iff $M, x \models \mathsf{S}\varphi \to \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$. If $M, x \not\models \mathsf{S}\varphi$ this is clear. Otherwise $x \in \|\mathsf{S}\varphi\|_M$, and we have

$$\begin{aligned} M, x \models [?\varphi]\mathsf{S}\psi &\iff M^{?\varphi}, x \models \mathsf{S}\psi \\ &\iff \forall B \in P^{?\varphi} : \|\psi\|_{M^{?\varphi}} \subseteq B \implies x \in B \\ &\iff \forall A \in P : \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq A \cap \|\mathsf{S}\varphi\|_M \implies x \in A \cap \|\mathsf{S}\varphi\|_M \\ &\iff \forall A \in P : \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq A \implies x \in A \\ &\iff M, x \models \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi) \end{aligned}$$

and the result follows.

For the E reduction axiom, take $M \in \mathbb{M}_{\mathsf{int}}$. Again, suppose without loss of generality that $x \in \|\mathsf{S}\varphi\|_M$. Then we have

$$\begin{aligned} M, x \models [?\varphi]\mathsf{E}\psi &\iff M^{?\varphi}, x \models \mathsf{E}\psi \\ &\iff \|\psi\|_{M^{?\varphi}} \in P^{?\varphi} \\ &\iff \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \in P^{?\varphi} \\ &\iff \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \in P \\ &\iff M, x \models \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi) \end{aligned}$$

where the forwards direction of the penultimate equivalence holds since $P^{?\varphi} \subseteq P$ when M is closed under intersections, and the backwards direction holds since $\|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq \|\mathsf{S}\varphi\|_M = X^{?\varphi}$. It follows that $M, x \models [?\varphi]\mathsf{E}\psi$ iff $M, x \models \mathsf{S}\varphi \to \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$, as required. □

To conclude, we note some interesting validities involving the dynamic operators and their interaction.

Proposition 7.7.3. *For any $\alpha, \beta \in \mathcal{L}_0$, the following formulas are valid on $\mathbb{M}_{\mathsf{int}}$:*

1. $\mathsf{E}\alpha \leftrightarrow \mathsf{A}[?\alpha]\alpha$
2. $\mathsf{A}(\alpha \to \beta) \to [+\beta][?\alpha]\beta$
3. $[+\alpha][?\alpha]\alpha$

Proof.

1. Using the reduction axioms for atomic propositions, conjunctions and negations, one can show by induction that $[?\varphi]\alpha$ is equivalent to $\mathsf{S}\varphi \to \alpha$. Applying this with $\varphi = \alpha$, we have that $\mathsf{A}[?\alpha]\alpha$ is equivalent to $\mathsf{A}(\mathsf{S}\alpha \to \alpha)$, which is equivalent to $\mathsf{E}\alpha$ for models closed under intersections.

2. We use the following fact, whose proof is straightforward by induction on $\mathcal{L}_0$ formulas.

   - For $\alpha \in \mathcal{L}_0$, $\varphi \in \mathcal{L}$ and any model M, $\|\alpha\|_{M^{+\varphi}} = \|\alpha\|_M$ and $\|\alpha\|_{M^{?\varphi}} = \|\alpha \wedge \mathsf{S}\varphi\|_M$.

   Now, take $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}}$, $x \in X$, and suppose $M, x \models \mathsf{A}(\alpha \to \beta)$. Then $\|\alpha\|_M \subseteq \|\beta\|_M$.

   We need to show $M, x \models [+\beta][?\alpha]\beta$, i.e. $M^{+\beta}, x \models [?\alpha]\beta$. Suppose $M^{+\beta}, x \models \mathsf{S}\alpha$. To show $(M^{+\beta})^{?\alpha}, x \models \beta$, we need

   $$x \in \|\beta\|_{(M^{+\beta})^{?\alpha}} = \|\beta \wedge \mathsf{S}\alpha\|_{M^{+\beta}}$$

   where the equality follows from the claim above. By assumption $M^{+\beta}, x \models \mathsf{S}\alpha$, so we only need to show $M^{+\beta}, x \models \beta$.

   Since $[+\beta]\mathsf{E}\beta$ is valid in M, we have $M^{+\beta}, x \models \mathsf{E}\beta$. From Proposition 7.2.1 (3), $M^{+\beta}, x \models \mathsf{A}(\alpha \to \beta) \to (\mathsf{S}\alpha \wedge \mathsf{E}\beta \to \beta)$. But from the above claim and $\|\alpha\|_M \subseteq \|\beta\|_M$ we have $\|\alpha\|_{M^{+\beta}} \subseteq \|\beta\|_{M^{+\beta}}$, i.e. $M^{+\beta}, x \models \mathsf{A}(\alpha \to \beta)$. Hence $M^{+\beta}, x \models \beta$, and we are done.

3. Taking $\beta = \alpha$, this validity follows from (2).

$\square$

In words, (1) says that expertise on a propositional formula α is equivalent to the guarantee that α is true whenever it is soundly announced. (2) is essentially a reformulation of Proposition 7.2.1 (3); it says that if β is logically weaker than α, gaining expertise on β ensures that β is at least true after a sound announcement of the stronger formula α. (3) is the special case of (2) with $\beta = \alpha$, which says that α is true following a sound announcement after the sources gains expertise on α.

7.8 Conclusion

This chapter presented a simple modal logic framework to reason about the expertise of information sources and soundness of information, generalising the framework of Singleton [87]. We investigated both conceptual and technical issues, establishing several completeness for various classes of expertise models. The connection with epistemic logic showed how expertise and soundness may be given precise interpretations in terms of knowledge; if expertise is closed under intersections and unions this results in S4 knowledge, and closure under complements strengthens this to S5. Finally, we extended the framework to handle multiple sources and studied notions of collective expertise.

There are many directions for future work. First, our approach allows one to reason about soundness of information only if the extent of a source's expertise is known up-front. In practical situations it is more likely that one has to *estimate* a source's expertise, e.g. on the basis of previous reports [56, 24]; such approaches could be combined with our logical framework in future work.

Expertise is also not static: it may change over time as sources learn and acquire new evidence. To model this one could introduce *dynamic expertise operators*, as in Dynamic Epistemic Logic. One source of inspiration here is *dynamic evidence logics* [95, 94], which study how evidence (and beliefs formed on the basis of evidence) change over time. Such logics also use neighbourhood semantics to interpret evidence modalities, which is technically (and possibly conceptually) similar to our semantics for expertise.

Finally, there is scope to study the interaction between interaction between expertise and *trust*, which has been extensively studied from a logical perspective [13, 71, 72, 55]. Intuitively, source i should trust j on φ if i believes that j has expertise on φ. "Belief in expertise" in this manner is not particularly meaningful in the current framework, since $\mathsf{E}_j\varphi$ either holds everywhere or nowhere. Future work could extend the semantics to allow the expertise collection P_j to vary between states, so as to model one source's uncertainty about the expertise of another.

8 Belief Change with Non-Expert Sources

8.1 Introduction

Consider the following belief change scenario in a hospital. We observe the results of a blood test of patient 1, confirming condition X. Assuming the test is reliable, the AGM paradigm [1] tells us how to revise our beliefs in light of the new information. Dr. A then claims that patient 2 suffers from the same condition, but Dr. B disagrees. Given that doctors specialise in different areas and may make mistakes, who should we trust? Since the *Success* postulate ($\alpha \in K * \alpha$) assumes information is reliable, we are outside the realm of AGM revision, and must instead apply some form of *non-prioritised* revision [52].

Suppose it now emerges that Dr. A had earlier claimed patient 1 did *not* suffer from condition X, contrary to the test results. We now have reason to suspect Dr. A may *lack expertise* on diagnosing X, and may subsequently revise beliefs about Dr. A's domain of expertise and the status of patient 2 (e.g. by opting to trust Dr. B instead).

While simple, this example illustrates the key features of the belief change problem we study: we consider multiple sources, whose expertise is *a priori* unknown, providing reports on various instances of a problem domain. On the basis of these reports we form beliefs both about the expertise of the sources and the state of the world in each instance.

By including a distinguished *completely reliable* source (the test results in the example) we extend AGM revision. In some respects we also extend approaches to non-prioritised revision (e.g. selective revision [41], credibility-limited revision [53], and trust-based revision [13]), which assume information about the reliability of sources is known up front. The problem is also related to *belief merging* [64] which deals with combining belief bases from multiple sources; a detailed comparison will be given in Section 8.7.

Our work is also connected to trust and belief revision, if one interprets trust as *belief in expertise*. As Yasser and Ismail [107] note in recent work, trust and belief are inexorably linked: we should accept reports from sources we believe are trustworthy, and we should trust sources whose reports turn out to be reliable. Trust and belief should also be revised in tandem, so that we may increase or decrease trust in a source as more reports are received, and revoke or reinstate previous reports from a source as its perceived trustworthiness changes.[1]

To unify the trust and belief aspects, we enrich a propositional language with *expertise statements* $\mathsf{E}_i\varphi$, read as "source i has expertise on φ". The output of our belief change problem is then a collection of belief and knowledge sets in the extended language, describing what we *know* and *believe* about the expertise of the sources and the state of the world in each instance. For example, we should *know* reports from the reliable source are true, whereas reports from ordinary sources may only be believed.

Following recent work on logical approaches to expertise [87, 13], we formally model expertise using a partition of propositional valuations for each source. Equivalently, each source has an *indistinguishability equivalence relation* over valuations. A source is an expert on a proposition φ exactly when they can distinguish every φ valuation from every $\neg\varphi$ valuation.[2] As in Singleton [87], we also use *soundness formulas* $\mathsf{S}_i\varphi$, which intuitively say that φ is true *up the expertise of* i. For example, if i has expertise on p but not q, then the conjunction $p \wedge q$ is sound for i whenever p holds, since we can effectively ignore q. Formally, φ is sound for i if the "actual" state of the world is indistinguishable from a φ valuation. Note that expertise does not depend on the "actual" state, whereas soundness does. This provides a crucial link between expertise and truthfulness of information.

We then make the assumption that *sources only report sound propositions*. That is, reports are only false due to sources overstepping the bounds of their expertise. In particular, we assume sources are honest in their reports, and that experts are always right.

Note that in our introductory example, the fact that we had a report from Dr. A on patient 1 (together with reliable information on patient 1) was essential for determining the expertise of Dr. A, and subsequently the status of patient 2. While the patients are independent, reports on one can cause beliefs about the other to change, as we update our beliefs about the expertise of the sources.

In general we consider an arbitrary number of *cases*, which are seen as labels for instances of the domain. For example, a crowdsourcing worker may label multiple images, or a weather forecaster may give predictions for different locations. Each report in the input to the problem then refers to a specific case. Via these

[1] This mutual dependence between trust and belief is also the core idea in *truth discovery* [69].

[2] The relationship between this notion of expertise and *S5 epistemic logic* is explored in a modal logic setting in Singleton [87], and we revisit this connection in Section 8.7.

cases and the presence of the completely reliable source, we are able to model scenarios where some “ground truth" is available, listing how often sources have been correct/incorrect on a proposition (e.g. the *report histories* of Hunter [57]). We can also generalise this scenario, e.g. by having only partial information about “previous" cases.

Throughout the chapter we make the assumption that *expertise is fixed across cases*: the expertise of a source does not depend on the particular instance of the domain we look at. For instance, the expertise of Dr. A is the same for patient 1 as for patient 2. This is a simplifying assumption, and may rule out certain interpretations of the cases (e.g. if cases represent different points in time, it would be natural to let expertise evolve over time).

Contribution. Our contributions are threefold. First, we develop a logical framework for reasoning about the expertise of multiple sources and the state of the world in multiple cases. Second, we formulate a belief change problem within this framework, which allows us to explore how trust and belief should interact and evolve as reports are received from the various sources. Finally, we put forward several postulates and two concrete classes of operators – with a representation result for one class – and analyse these operators with respect to the postulates.

Chapter Outline. In Section 8.2 we develop the formal framework. Section 8.3 introduces the problem and lists some core postulates. We give two constructions and specific example operators in Section 8.4. Section 8.5 introduces some further postulates concerning belief change on the basis of one new report. An analogue of selective revision [41] is presented Section 8.6. Section 8.7 discusses related work, and we conclude in Section 8.8. **[TODO:** mention expertise and selectivity?]

8.2 The Framework

Let $\mathcal{S}$ be a finite set of information sources. For convenience, we assume there is a *completely reliable* source in $\mathcal{S}$, which we denote by $*$. For example, we can treat our first-hand observations as if they are reported by $*$. Other sources besides $*$ will be termed *ordinary sources*. Let $\mathcal{C}$ be a finite set of *cases*, which we interpret as labels for different instances of the problem domain.

Syntax. There are two levels to our formal language. To describe properties of the world in each case $c \in \mathcal{C}$, we assume a fixed finite set $\mathcal{P}$ of propositional variables, and let $\mathcal{L}_0$ denote the set of propositional formulas generated from $\mathcal{P}$ using the usual propositional connectives. We use lower case Greek letters (φ, ψ etc) for formulas in $\mathcal{L}_0$. The classical logical consequence operator will be denoted by Cn_0, and $\equiv$ denotes equivalence of propositional formulas.

The extended *language of expertise* $\mathcal{L}$ additionally describes the expertise of the sources, and is defined by the following grammar:

$$\Phi ::= \varphi \mid \Phi \wedge \Phi \mid \neg\Phi \mid \mathsf{E}_i\varphi \mid \mathsf{S}_i\varphi$$

where $i \in \mathcal{S}$ and $\varphi \in \mathcal{L}_0$. We introduce Boolean connectives $\vee$, $\rightarrow$, $\leftrightarrow$ and $\bot$ as abbreviations. We use upper case Greek letters (Φ, Ψ etc) for formulas in $\mathcal{L}$. For $\Gamma \subseteq \mathcal{L}$, we write $[\Gamma] = \Gamma \cap \mathcal{L}_0$ for the propositional formulas in Γ.

The intuitive reading of $\mathsf{E}_i\varphi$ is *source i has expertise on φ*, i.e., i is able to correctly identify the truth value of φ in any possible state. The intuitive reading of $\mathsf{S}_i\varphi$ is that φ *sound* for i to report: that φ is true up to the expertise of i. That is, the parts of φ on which i has expertise are true. Note that both operators are restricted to propositional formulas, so we will not consider iterated formulas such as $\mathsf{E}_i\mathsf{S}_j\varphi$.

Semantics. Let $\mathcal{V}$ denote the set of propositional valuations over $\mathcal{P}$. For each $\varphi \in \mathcal{L}_0$, the set of valuations making φ true is denoted by $\mathrm{mod}_0(\varphi)$. A *world* $W = \langle \{v_c\}_{c\in\mathcal{C}}, \{\Pi_i\}_{i\in\mathcal{S}} \rangle$ is a possible complete specification of the environment we find ourselves in:

- $v_c \in \mathcal{V}$ is the "true" valuation at case $c \in \mathcal{C}$;
- Π_i is a partition of $\mathcal{V}$ for each $i \in \mathcal{S}$, representing the "true" expertise of source i; and
- Π_* is the unit partition $\{\{v\} \mid v \in \mathcal{V}\}$.

Let $\mathcal{W}$ denote the set of all worlds. Note that the partition corresponding to the distinguished source $*$ is fixed in all worlds as the finest possible partition, reflecting the fact that $*$ is completely reliable.

For any partition Π and valuation v, write $\Pi[v]$ for the unique cell in Π containing v. For a set of valuations U, write $\Pi[U] = \bigcup_{v\in U} \Pi[v]$. For brevity, we write $\Pi[\varphi]$ for $\Pi[\mathrm{mod}_0(\varphi)]$. Then $\Pi[\varphi]$ is the set of valuations indistinguishable from a φ valuation.

For our belief change problem we will be interested in maintaining a collection of several belief sets, describing beliefs about each case $c \in \mathcal{C}$. Towards determining when a world W models such a collection, we define semantics for $\mathcal{L}$ formulas with respect to a world and a case:

$$\begin{aligned} W, c \models \varphi &\iff v_c \in \mathrm{mod}_0(\varphi) \\ W, c \models \mathsf{E}_i\varphi &\iff \Pi_i[\varphi] = \mathrm{mod}_0(\varphi) \\ W, c \models \mathsf{S}_i\varphi &\iff v_c \in \Pi_i[\varphi] \end{aligned}$$

where $i \in \mathcal{S}$, $\varphi \in \mathcal{L}_0$, and the clauses for conjunction and negation are the expected ones. Since $\mathrm{mod}_0(\varphi) \subseteq \Pi_i[\varphi]$ always holds, we have that $\mathsf{E}_i\varphi$ holds iff there is no $\neg\varphi$ valuation which is indistinguishable from a φ valuation (c.f. Booth and Hunter [13]). Note that since each source i has only a single partition Π_i used to interpret the expertise formulas, the truth value of $\mathsf{E}_i\varphi$ does not depend on the case c. On the other hand, $\mathsf{S}_i\varphi$ holds in case c iff the c-valuation of W is indistinguishable from some model of φ. That is, it is consistent with i's expertise that φ is true.

Note that the mapping $2^{\mathcal{V}} \to 2^{\mathcal{V}}$ given by $U \mapsto \Pi[U]$ satisfies the *Kuratowski closure axioms*,[3] so can be considered a closure operator of the set of propositional valuations. Then $W, c \models \mathsf{E}_i\varphi$ iff $\mathrm{mod}_0(\varphi)$ is closed in V, and $W, c \models \mathsf{S}_i\varphi$ iff v_c lies in the closure of $\mathrm{mod}_0(\varphi)$, i.e. φ is true after closing $\mathrm{mod}_0(\varphi)$ along the lines of the expertise of source i. Also note that $\Pi[U] = U$ iff U can be expressed as a union of the partition cells in Π, so that $W, c \models \mathsf{E}_i\varphi$ can alternatively be interpreted as saying φ is a disjunction of stronger formulas on which i also has expertise.

Also note that if φ is a propositional tautology, $\mathsf{E}_i\varphi$ holds for every source i. Thus, all sources are experts on *something*, even if just the tautologies.

Example 8.2.1. *Let us extend the hospital example from the introduction. Let $\mathcal{S} = \{*, a, b\}$ denote the reliable source, Dr. A and Dr. B, and let $\mathcal{C} = \{c_1, c_2\}$ denote patients 1 and 2. Consider propositional variables $\mathcal{P} = \{x, y\}$, standing for condition X and Y respectively. Suppose that Dr. A has expertise on diagnosing condition Y only, whereas Dr. B only has expertise on X. For the sake of the example, suppose that patient 1 suffers from both conditions, and patient 2 suffers only from condition Y. This situation is modelled by the following world $W = \langle \{v_c\}_{c \in \{c_1, c_2\}}, \{\Pi_i\}_{i \in \{*, a, b\}} \rangle$:*

$$\begin{array}{cc} v_{c_1} = xy; & v_{c_2} = \bar{x}y; \\ \Pi_a = xy, \bar{x}y \mid x\bar{y}, \bar{x}\bar{y}; & \Pi_b = xy, x\bar{y} \mid \bar{x}y, \bar{x}\bar{y}. \end{array}$$

We have $W, c \models \mathsf{E}_a y \wedge \mathsf{E}_b x$ for each $c \in \{c_1, c_2\}$. Also note that $W, c_1 \models x$ (patient 1 suffers from X), $W, c_1 \models \mathsf{S}_a \neg x$ (it is sound for Dr. A to report otherwise; this holds since $\Pi_a[\neg x] = \{xy, \bar{x}y\} \cup \{x\bar{y}, \bar{x}\bar{y}\} \ni xy = v_{c_1}$), but $W, c_1 \models \neg \mathsf{S}_b \neg x$ (the same formula is not *sound for Dr. B; we have $\Pi_b[\neg x] = \{\bar{x}y, \bar{x}\bar{y}\} = \mathrm{mod}_0(\neg x) \not\ni xy = v_{c_1}$).*

Say Φ is *valid* if $W, c \models \Phi$ for all $W \in \mathcal{W}$ and $c \in \mathcal{C}$. For future reference we collect a list of validities.

Proposition 8.2.1. *For any $i \in \mathcal{S}$, $c \in \mathcal{C}$ and $\varphi, \psi \in \mathcal{L}_0$, the following formulas are valid*

1. $\mathsf{S}_i\varphi \leftrightarrow \mathsf{S}_i\psi$ *and* $\mathsf{E}_i\varphi \leftrightarrow \mathsf{E}_i\psi$*, whenever* $\varphi \equiv \psi$

2. $\mathsf{E}_i\varphi \leftrightarrow \mathsf{E}_i\neg\varphi$ *and* $\mathsf{E}_i\varphi \wedge \mathsf{E}_i\psi \to \mathsf{E}_i(\varphi \wedge \psi)$

[3] That is, (i) $\Pi[\emptyset] = \emptyset$, (ii) $U \subseteq \Pi[U]$, (iii) $\Pi[\Pi[U]] = \Pi[U]$, and (iv) $\Pi[U_1 \cup U_2] = \Pi[U_1] \cup \Pi[U_2]$.

3. $\mathsf{E}_i p_1 \wedge \cdots \wedge \mathsf{E}_i p_k \to \mathsf{E}_i \varphi$, *where* $p_1, \ldots, p_k$ *are the propositional variables appearing in* φ

4. $\mathsf{E}_i \varphi \wedge \mathsf{S}_i \varphi \to \varphi$, *and* $\mathsf{S}_i \varphi \wedge \neg\varphi \to \neg\mathsf{E}_i \varphi$

5. $\mathsf{S}_i \varphi \wedge \mathsf{S}_i \neg\varphi \to \neg\mathsf{E}_i \varphi$

6. $\mathsf{S}_* \varphi \leftrightarrow \varphi$ *and* $\mathsf{E}_* \varphi$

We comment on each property before giving the proof. (1) states syntax-irrelevance properties. (2) says that expertise is symmetric with respect to negation, and closed under conjunctions. Intuitively, symmetry means that i is an expert on φ if they know *whether or not* φ holds. (3) says that expertise on each propositional variable in φ is sufficient for expertise on φ itself. (4) says that, in the presence of expertise, soundness of φ is sufficient for φ to in fact be true. (5) says that if both φ and $\neg\varphi$ are true up to the expertise of i, then i cannot have expertise on φ. Finally, (6) says that the reliable source $*$ has expertise on *all* formulas, and thus φ is sound for $*$ iff it is true.

Proof.

1. If $\varphi \equiv \psi$ then $\mathrm{mod}_0(\varphi) = \mathrm{mod}_0(\psi)$; since the semantics for $\mathsf{S}_i\varphi$ and $\mathsf{E}_i\varphi$ only refer to $\mathrm{mod}_0(\varphi)$ (and likewise for ψ), we have that $\mathsf{S}_i\varphi \leftrightarrow \mathsf{S}_i\psi$ and $\mathsf{E}_i\varphi \leftrightarrow \mathsf{E}_i\psi$ are valid.

2. For the first validity, suppose $W, c \models \mathsf{E}_i\varphi$. Then $\mathrm{mod}_0(\varphi) = \Pi_i[\varphi]$. We show $W, c \models \mathsf{E}_i\neg\varphi$. Indeed, take $v \in \Pi_i[\neg\varphi]$. Then there is $v' \in \mathrm{mod}_0(\neg\varphi)$ such that $v \in \Pi_i[v']$. Thus $v' \in \Pi_i[v]$ also. Supposing for contradiction that $v \in \mathrm{mod}_0(\varphi)$, we get
$$v' \in \Pi_i[v] \subseteq \Pi_i[\varphi] = \mathrm{mod}_0(\varphi).$$
But then $v' \in \mathrm{mod}_0(\neg\varphi) \cap \mathrm{mod}_0(\varphi) = \emptyset$; contradiction. Hence $v \notin \mathrm{mod}_0(\varphi)$, i.e. $v \in \mathrm{mod}_0(\neg\varphi)$. This shows that $\Pi_i[\neg\varphi] \subseteq \mathrm{mod}_0(\neg\varphi)$, so $W, c \models \mathsf{E}_i\neg\varphi$.

   We have shown that $\mathsf{E}_i\varphi \to \mathsf{E}_i\neg\varphi$ is valid. For the converse note that, by symmetry, $\mathsf{E}_i\neg\varphi \to \mathsf{E}_i\neg\neg\varphi$ is valid; since $\mathsf{E}_i\neg\neg\varphi$ is equivalent to $\mathsf{E}_i\varphi$ by (1) we get $\mathsf{E}_i\varphi \leftrightarrow \mathsf{E}_i\neg\varphi$.

   For the second validity, suppose $W, c \models \mathsf{E}_i\varphi \wedge \mathsf{E}_i\psi$. Note that
$$\Pi_i[\varphi \wedge \psi] \subseteq \Pi_i[\varphi] = \mathrm{mod}_0(\varphi)$$
and, similarly, $\Pi_i[\varphi \wedge \psi] \subseteq \mathrm{mod}_0(\psi)$. Hence
$$\Pi_i[\varphi \wedge \psi] \subseteq \mathrm{mod}_0(\varphi) \cap \mathrm{mod}_0(\psi) = \mathrm{mod}_0(\varphi \wedge \psi),$$
which shows $W, c \models \mathsf{E}_i(\varphi \wedge \psi)$.

3. Let φ be a propositional formula, and let $p_1, \ldots, p_k$ be the variables appearing in φ. Let $\widehat{\mathcal{L}_0} \subseteq \mathcal{L}_0$ be the propositional formulas over $p_1, \ldots p_k$ generated only using conjunction and negation. Then there is some $\psi \in \widehat{\mathcal{L}_0}$ with $\varphi \equiv \psi$.

   Suppose $W, c \models \mathsf{E}_i p_1 \wedge \cdots \wedge \mathsf{E}_i p_k$. By this assumption and the properties in (2), one can show by induction that $W, c \models \mathsf{E}_i \theta$ for all $\theta \in \widehat{\mathcal{L}_0}$. In particular, $W, c \models \mathsf{E}_i \psi$. Since $\varphi \equiv \psi$, we get $W, c \models \mathsf{E}_i \varphi$.

4. Suppose $W, c \models \mathsf{E}_i \varphi \wedge \mathsf{S}_i \varphi$. Then $v_c \in \Pi_i[\varphi] = \mathrm{mod}_0(\varphi)$, so $W, c \models \varphi$. Hence $\mathsf{E}_i \varphi \wedge \mathsf{S}_i \varphi \to \varphi$ is valid. Similarly, $\mathsf{S}_i \varphi \wedge \neg\varphi \to \neg \mathsf{E}_i \varphi$ is valid.

5. Suppose $W, c \models \mathsf{S}_i \varphi \wedge \mathsf{S}_i \neg\varphi$, and, for contradiction, $W, c \models \mathsf{E}_i \varphi$. On the one hand we have $W, c \models \mathsf{E}_i \varphi \wedge \mathsf{S}_i \varphi$, so (4) gives $W, c \models \varphi$. On the other hand, $W, c \models \mathsf{E}_i \varphi$ gives $W, c \models \mathsf{E}_i \neg\varphi$ by (2), so $W, c \models \mathsf{E}_i \neg\varphi \wedge \mathsf{S}_i \neg\varphi$; by (4) again we have $W, c \models \neg\varphi$. But then $W, c \models \varphi \wedge \neg\varphi$ – contradiction.

6. Since the distinguished source $*$ has the unit partition Π_* in any world W, we have $\Pi_*[\varphi] = \mathrm{mod}_0(\varphi)$, so $W, c \models \mathsf{E}_* \varphi$. Similarly, $W, c \models \mathsf{S}_i \varphi$ iff $v_c \in \Pi_*[\varphi] = \mathrm{mod}_0(\varphi)$ iff $W, c \models \varphi$.

$\square$

Case-Indexed Collections. In the remainder of this chapter we will be interested in forming beliefs about each case $c \in \mathcal{C}$. To do so we use collections of belief sets $G = \{\Gamma_c\}_{c \in \mathcal{C}}$, with $\Gamma_c \subseteq \mathcal{L}$, indexed by cases. Say a world W is a *model* of G iff

$$W, c \models \Phi \text{ for all } c \in \mathcal{C} \text{ and } \Phi \in \Gamma_c,$$

i.e. iff W satisfies all formulas in G in the relevant case. Let $\mathrm{mod}(G)$ denote the models of G, and say that G is *consistent* if $\mathrm{mod}(G) \neq \emptyset$. For $c \in \mathcal{C}$, define the *c-consequences*

$$\mathrm{Cn}_c(G) = \{\Phi \in \mathcal{L} \mid \forall W \in \mathrm{mod}(G), W, c \models \Phi\}.$$

We write $\mathrm{Cn}(G)$ for the collection $\{\mathrm{Cn}_c(G)\}_{c \in \mathcal{C}}$.

Example 8.2.2. *Suppose $\mathcal{C} = \{c_1, c_2, c_3\}$, and define G by $\Gamma_{c_1} = \{\mathsf{S}_i(p \wedge q)\}$, $\Gamma_{c_2} = \{\mathsf{E}_i p\}$ and $\Gamma_{c_3} = \{\mathsf{E}_i q\}$. Then, since expertise holds independently of case, any model W of G has $W, c_1 \models \mathsf{E}_i p \wedge \mathsf{E}_i q$. By Proposition 8.2.1 part (3), $W, c_1 \models \mathsf{E}_i(p \wedge q)$. Since W satisfies Γ_{c_1} in case c_1, Proposition 8.2.1 part (4) gives $W, c_1 \models p \wedge q$. Since W was an arbitrary model of G, we have $p \wedge q \in \mathrm{Cn}_{c_1}(G)$, i.e. $p \wedge q$ is a c_1-consequence of G. This illustrates how information about distinct cases can be brought together to have consequences for other cases.*

For two collections $G = \{\Gamma_c\}_{c \in \mathcal{C}}$, $D = \{\Delta_c\}_{c \in \mathcal{C}}$, write $G \sqsubseteq D$ iff $\Gamma_c \subseteq \Delta_c$ for all c, and let $G \sqcup D$ denote the collection $\{\Gamma_c \cup \Delta_c\}_{c \in \mathcal{C}}$. With this notation, the case-indexed consequence operator satisfies analogues of the Tarskian consequence properties.[4]

Say a collection G is *closed* if $\mathrm{Cn}(G) = G$. Closed collections provide an idealised representation of beliefs, which will become useful later on. For instance, when G is closed we have $\mathsf{E}_i\varphi \in \Gamma_c$ iff $\mathsf{E}_i\varphi \in \Gamma_d$ for all $c, d \in \mathcal{C}$; i.e. expertise statements are either present for all cases or for none. We also have $\mathrm{Cn}_0\left[\Gamma_c\right] = \left[\Gamma_c\right]$, i.e. the propositional parts of G are (classically) closed.

In propositional logic, mod_0 is a 1-to-1 correspondence between closed sets of formulas and sets of valuations. This is not so in our setting, since some subsets of $\mathcal{W}$ do not arise as the models of any collection. Instead, we have a 1-to-1 correspondence into a restricted collection of sets of worlds. Borrowing the terminology of Delgrande, Peppas, and Woltran [26], say a set of worlds $S \subseteq \mathcal{W}$ is *elementary* if $S = \mathrm{mod}(G)$ for some collection $G = \{\Gamma_c\}_{c\in\mathcal{C}}$.[5]

Elementariness is characterised by a certain closure condition. Say that two worlds W, W' are *partition-equivalent* if $\Pi_i^W = \Pi_i^{W'}$ for all sources i, and say W is a *valuation combination* from a set $S \subseteq \mathcal{W}$ if for all cases c there is $W_c \in S$ such that $v_c^W = v_c^{W_c}$. Then a set is elementary iff it is closed under valuation combinations of partition-equivalent worlds.

Proposition 8.2.2. *$S \subseteq \mathcal{W}$ is elementary if and only if the following condition holds: for all $W \in \mathcal{W}$ and $W_1, W_2 \in S$, if W is partition-equivalent to both W_1, W_2 and W is a valuation combination from $\{W_1, W_2\}$, then $W \in S$.*

Proof. "if": Suppose the stated condition holds for $S \subseteq \mathcal{W}$. Form a collection $G = \{\Gamma_c\}_{c\in\mathcal{C}}$ by setting $\Gamma_c = \{\Phi \in \mathcal{L} \mid S \subseteq \mathrm{mod}_c(\Phi)\}$. Clearly $S \subseteq \mathrm{mod}(G)$. For the reverse inclusion, suppose $W \in \mathrm{mod}(G)$. For any set of valuations $U \subseteq \mathcal{V}$, let φ_U be any propositional sentence with $\mathrm{mod}_0(\varphi_U) = U$. For each $c \in \mathcal{C}$, consider the sentence

$$\Phi_c = \bigvee_{W'\in S} \left(\varphi_{\{v_c^{W'}\}} \wedge \bigwedge_{i\in\mathcal{S}} \bigwedge_{U\subseteq\mathcal{V}} R_{W',i,U} \right)$$

where

$$R_{W',i,U} = \begin{cases} \mathsf{E}_i\varphi_U, & W', c_0 \models \mathsf{E}_i\varphi_U \\ \neg\mathsf{E}_i\varphi_U, & \text{otherwise} \end{cases}$$

for some fixed case $c_0 \in \mathcal{C}$. It is straightforward to see that each $W' \in S$ satisfies its corresponding disjunct at case c, so $\Phi_c \in \Gamma_c$. Hence $W \in \mathrm{mod}(G)$ implies $W, c \models \Phi_c$ for each c. Consequently, for each c there is some $W_c \in S$ such that (i)

[4] That is, (i) $G \sqsubseteq \mathrm{Cn}(G)$, (ii) $G \sqsubseteq D$ implies $\mathrm{Cn}(G) \sqsubseteq \mathrm{Cn}(D)$, and (iii) $\mathrm{Cn}(\mathrm{Cn}(G)) = \mathrm{Cn}(G)$.

[5] Non-elementary sets can also exist for weaker logics (such as Horn logic [26]) which lack the syntactic expressivity to identify all sets of models. In our framework, $\mathcal{C}$-indexed collections are not expressive enough to specify *combinations of valuations*, since each Γ_c only says something about the valuation for c.

$v_c^W = v_c^{W_c}$; and (ii) for each $i \in \mathcal{S}$ and $U \subseteq V$, $W, c \models \mathsf{E}_i\varphi_U$ iff $W_c, c \models \mathsf{E}_i\varphi_U$. From (i), W is a valuation combination from $\{W_c\}_{c\in\mathcal{C}}$. From (ii) it can be shown that in fact $\Pi_i^W = \Pi_i^{W_c}$ for each c and i; that is, W is partition-equivalent to each W_c. In particular, all the W_c are partition-equivalent to each other. Repeatedly applying the closure condition assumed to hold for S, we see that $W \in S$ as required.

"only if": Suppose S is elementary, i.e. $S = \mathrm{mod}(G)$ for some collection $G = \{\Gamma_c\}_{c\in\mathcal{C}}$, and let W, W_1, W_2 be as in the statement of the proposition. Take $c \in \mathcal{C}$ and $\Phi \in \Gamma_c$. We will show $W, c \models \Phi$. By assumption, there is $n \in \{1, 2\}$ such that $v_c^W = v_c^{W_n}$. It can be shown by induction on $\mathcal{L}$ formulas that, since W and W_n are partition-equivalent and have the same c valuation, $W, c \models \Phi$ iff $W_n, c \models \Phi$. But $W_n \in S = \mathrm{mod}(G)$ implies $W_n, c \models \Phi$, so $W, c \models \Phi$ too. Since $\Phi \in \Gamma_c$ was arbitrary, we have $W \in \mathrm{mod}(G) = S$ as required.

□

8.3 The Problem

With the framework set out, we can formally define the problem. We seek an operator with the following behaviour:

- **Input:** A sequence of reports σ, where each report is a triple $\langle i, c, \varphi\rangle \in \mathcal{S} \times \mathcal{C} \times \mathcal{L}_0$ and $\varphi \not\equiv \bot$. Such a report represents that *source i reports φ to hold in case c*. Note that we only allow sources to make *propositional* reports.

- **Output:** A pair $\langle B^\sigma, K^\sigma\rangle$, where $B^\sigma = \{B_c^\sigma\}_{c\in\mathcal{C}}$ is a collection of *belief sets* $B_c^\sigma \subseteq \mathcal{L}$ and $K^\sigma = \{K_c^\sigma\}_{c\in\mathcal{C}}$ is a collection of *knowledge sets* $K_c^\sigma \subseteq \mathcal{L}$.

8.3.1 Basic Postulates

We immediately narrow the scope of operators under consideration by introducing some basic postulates which are expected to hold. In what follows, say a sequence σ is $*$*-consistent* if for each $c \in \mathcal{C}$ the set $\{\varphi \mid \langle *, c, \varphi\rangle \in \sigma\} \subseteq \mathcal{L}_0$ is classically consistent. Write G_{snd}^σ for the collection with $(G_{\mathsf{snd}}^\sigma)_c = \{\mathsf{S}_i\varphi \mid \langle i, c, \varphi\rangle \in \sigma\}$, i.e. the collection of soundness statements corresponding to the reports in σ.

Closure *$B^\sigma = \mathrm{Cn}(B^\sigma)$ and $K^\sigma = \mathrm{Cn}(K^\sigma)$*

Containment *$K^\sigma \sqsubseteq B^\sigma$*

Consistency *If σ is $*$-consistent, B^σ and K^σ are consistent*

Soundness *If $\langle i, c, \varphi\rangle \in \sigma$, then $\mathsf{S}_i\varphi \in K_c^\sigma$*

K-bound *$K^\sigma \sqsubseteq \mathrm{Cn}(G_{\mathsf{snd}}^\sigma \sqcup K^\emptyset)$*

Prior-Extension $K^{\emptyset} \sqsubseteq K^{\sigma}$

Rearrangement *If σ is a permutation of ρ, then $B^{\sigma} = B^{\rho}$ and $K^{\sigma} = K^{\rho}$*

Equivalence *If $\varphi \equiv \psi$ then $B^{\sigma \cdot \langle i,c,\varphi \rangle} = B^{\sigma \cdot \langle i,c,\psi \rangle}$ and $K^{\sigma \cdot \langle i,c,\varphi \rangle} = K^{\sigma \cdot \langle i,c,\psi \rangle}$*

Closure says that the belief and knowledge collections are closed under logical consequence. In light of earlier remarks, this implies that the propositional belief sets $[B^{\sigma}_c]$ are closed under (propositional) consequence, and that $\mathsf{E}_i\varphi \in B^{\sigma}_c$ iff $\mathsf{E}_i\varphi \in B^{\sigma}_d$. *Containment* says that everything which is known is also believed. *Consistency* ensures the output is always consistent, provided we are not in the degenerate case where $*$ gives inconsistent reports. *Soundness* says we *know* that all reports are sound in their respective cases. This formalises our assumption that sources are *honest*, i.e. that false reports only arise due to lack of expertise. By Proposition 8.2.1 part (4) it also implies *experts are always right*: if a source has expertise on their report then it must be true. While *Soundness* places a lower bound on knowledge, *K-bound* places an upper bound: knowledge cannot go beyond the soundness statements corresponding to the reports in σ together with the prior knowledge $K^{\emptyset}$. That is, from the point view of knowledge, a new report of $\langle i, c, \varphi \rangle$ only allows us to learn $\mathsf{S}_i\varphi$ in case c (and to combine this with other reports and prior knowledge). Note that the analogous property for belief is *not* desirable: we want to be more liberal when it comes to beliefs, and allow for *defeasible inferences* going beyond the mere fact that reports are sound. *Prior-Extension* says that knowledge after a sequence σ extends the prior knowledge on the empty sequence $\emptyset$. *Rearrangement* says that the order in which reports are received is irrelevant. This can be justified on the basis that we are reasoning about *static worlds* for each case c, so that there is no reason to see more "recent" reports as any more or less important or truthful than earlier ones.[6] Consequently, we can essentially view the input as a *multi-set* of belief sets – one for each source – bringing us close to the setting of belief merging. This postulate also appears as the commutativity postulate **(Com)** in the work of Schwind and Konieczny [86]. Finally, *Equivalence* says that the syntactic form of reports is irrelevant.

Taking all the basic postulates together, the knowledge component K^{σ} is fully determined once $K^{\emptyset}$ is chosen.

Proposition 8.3.1. *Suppose an operator satisfies the basic postulates. Then*

1. $K^{\sigma} = \mathrm{Cn}(G^{\sigma}_{\mathsf{snd}} \sqcup K^{\emptyset})$

2. $K^{\emptyset} = \mathrm{Cn}(\emptyset)$ *iff* $K^{\sigma} = \mathrm{Cn}(G^{\sigma}_{\mathsf{snd}})$ *for all σ.*

Proof.

[6] This argument is from [25].

1. The "$\sqsubseteq$" inclusion is just *K-bound*. For the "$\sqsupseteq$" inclusion, note that $G^\sigma_{\mathsf{snd}} \sqsubseteq K^\sigma$ by *Soundness*, and $K^\emptyset \sqsubseteq K^\sigma$ by *Prior-Extension*. Hence

$$G^\sigma_{\mathsf{snd}} \sqcup K^\emptyset \sqsubseteq K^\sigma.$$

By monotonicity of Cn,

$$\mathrm{Cn}(G^\sigma_{\mathsf{snd}} \sqcup K^\emptyset) \sqsubseteq \mathrm{Cn}(K^\sigma) = K^\sigma$$

where we use *Closure* in the final step.

2. "if": Suppose $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$ for all σ. Taking $\sigma = \emptyset$ we obtain

$$K^\sigma = \mathrm{Cn}(G^\emptyset_{\mathsf{snd}}) = \mathrm{Cn}(\emptyset).$$

"only if": Suppose $K^\emptyset = \mathrm{Cn}(\emptyset)$. Take any sequence σ. By *K-bound*,

$$K^\sigma \sqsubseteq \mathrm{Cn}(G^\sigma_{\mathsf{snd}} \sqcup \mathrm{Cn}(\emptyset)) = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$$

On the other hand, *Soundness* and *Closure* give $\mathrm{Cn}(G^\sigma_{\mathsf{snd}}) \sqsubseteq K^\sigma$. Hence $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$.

□

The choice of $K^\emptyset$ depends on the scenario one wishes to model. While $\mathrm{Cn}(\emptyset)$ is a sensible choice if the sequence σ is all we have to go on, we allow $K^\emptyset \neq \mathrm{Cn}(\emptyset)$ in case *prior knowledge* is available (for example, the expertise of particular sources may be known ahead of time).

Another important property of knowledge, which follows from the basic postulates, says that *knowledge is monotonic*: knowledge after receiving σ and ρ together is just the case-wise union of K^σ and K^ρ.

K-conjunction $K^{\sigma\cdot\rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$

K-conjunction reflects the idea that one should be cautious when it comes to knowledge: a formula should only be accepted as known if it won't be given up in light of new information.

Proposition 8.3.2. *Any operator satisfying the basic postulates satisfies* K-conjunction.

Proof. Suppose an operator satisfies the basic postulates, and take sequences σ and ρ. By Proposition 8.3.1,

$$K^{\sigma\cdot\rho} = \mathrm{Cn}(G^{\sigma\cdot\rho}_{\mathsf{snd}} \sqcup K^\emptyset)$$

Note that $G^{\sigma\cdot\rho}_{\mathsf{snd}} = G^\sigma_{\mathsf{snd}} \sqcup G^\rho_{\mathsf{snd}}$. Hence we may write

$$\begin{aligned} K^{\sigma\cdot\rho} &= \mathrm{Cn}(G^\sigma_{\mathsf{snd}} \sqcup G^\rho_{\mathsf{snd}} \sqcup K^\emptyset) \\ &= \mathrm{Cn}((G^\sigma_{\mathsf{snd}} \sqcup K^\emptyset) \sqcup (G^\rho_{\mathsf{snd}} \sqcup K^\emptyset)) \end{aligned}$$

By Proposition 8.3.1 again, we have $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}} \sqcup K^\emptyset)$ and $K^\rho = \mathrm{Cn}(G^\rho_{\mathsf{snd}} \sqcup K^\emptyset)$. It is easily verified that for any collections G, D, we have

$$\mathrm{Cn}(G \sqcup D) = \mathrm{Cn}(\mathrm{Cn}(G) \sqcup \mathrm{Cn}(D)).$$

Consequently,

$$\begin{aligned} K^{\sigma\cdot\rho} &= \mathrm{Cn}(\mathrm{Cn}(G^\sigma_{\mathsf{snd}} \sqcup K^\emptyset) \sqcup \mathrm{Cn}(G^\rho_{\mathsf{snd}} \sqcup K^\emptyset)) \\ &= \mathrm{Cn}(K^\sigma \sqcup K^\rho) \end{aligned}$$

as required for *K-conjunction*. □

The postulates also imply some useful properties linking *trust* (seen as belief in expertise) and *belief/knowledge*.

Proposition 8.3.3. *Suppose an operator satisfies the basic postulates. Then*

1. *If* $\varphi \in K^\sigma_c$ *and* $\neg\psi \in \mathrm{Cn}_0(\varphi)$ *then* $\neg\mathsf{E}_i\psi \in K^{\sigma\cdot\langle i,c,\psi\rangle}_c$.

2. *If* $\langle i, c, \varphi\rangle \in \sigma$ *and* $\mathsf{E}_i\varphi \in B^\sigma_c$ *then* $\varphi \in B^\sigma_c$.

Proof.

1. Suppose $\varphi \in K^\sigma_c$ and $\neg\psi \in \mathrm{Cn}_0(\varphi)$. Write $\rho = \sigma \cdot \langle i, c, \psi\rangle$. By *Soundness*, $\mathsf{S}_i\psi \in K^\rho_c$. By *K-conjunction*, $\varphi \in K^\sigma_c \subseteq (K^\sigma \sqcup K^{\langle i,c,\psi\rangle})_c \subseteq \mathrm{Cn}_c(K^\sigma \sqcup K^{\langle i,c,\psi\rangle}) = K^\rho_c$. Since $\neg\psi \in \mathrm{Cn}_0(\varphi)$ and $\varphi \in K^\rho_c$, *Closure* gives $\neg\psi \in K^\rho_c$. Recalling from Proposition 8.2.1 part (4) that $\mathsf{S}_i\psi \wedge \neg\psi \to \neg\mathsf{E}_i\psi$, *Closure* gives $\neg\mathsf{E}_i\psi \in K^\rho_c$, as desired.

2. Suppose $\langle i, c, \varphi\rangle \in \sigma$ and $\mathsf{E}_i\varphi \in B^\sigma_c$. By *Soundness* and *Containment*, $\mathsf{S}_i\varphi \in B^\sigma_c$. From Proposition 8.2.1 part (4) again we have $\mathsf{E}_i\varphi \wedge \mathsf{S}_i\varphi \to \varphi$. By *Closure*, $\varphi \in B^\sigma_c$.

□

(1) expresses how knowledge can negatively affect trust: we should distrust sources who make reports we know to be false. (2) expresses how trust affects belief: we should believe reports from trusted sources. It can also be seen as a form of *success* for ordinary sources, and implies AGM success when $i = *$ (by Proposition 8.2.1 part (6) and *Closure*). We illustrate the basic postulates by formalising the introductory hospital example.

Example 8.3.1. *Set* $\mathcal{S}, \mathcal{C}$ *and* $\mathcal{P}$ *as in Example 8.2.1, and consider the sequence*

$$\sigma = (\langle *, c_1, x\rangle, \langle a, c_2, x\rangle, \langle b, c_2, \neg x\rangle, \langle a, c_1, \neg x\rangle).$$

What do we know on the basis of this sequence, assuming the basic postulates? First note that by Soundness*, Proposition 8.2.1 part (6) and* Closure*, the report from $*$ gives $x \in K^{\sigma}_{c_1}$, i.e. reliable reports are known.* Soundness *also gives $\mathsf{S}_a x \wedge \mathsf{S}_b \neg x \in K^{\sigma}_{c_2}$. Combined with Proposition 8.2.1 parts (2), (4) and* Closure*, this yields $\neg(\mathsf{E}_a x \wedge \mathsf{E}_b x) \in K^{\sigma}_{c}$ for all c, formalising the intuitive idea that Drs. A and B cannot both be experts on X, since they give conflicting reports. Considering the final report from a, we get $x \wedge \mathsf{S}_a \neg x \in K^{\sigma}_{c_1}$, and thus $\neg \mathsf{E}_a x \in K^{\sigma}_{c}$ by* Closure*. So in fact Dr. A is known to be a non-expert on X.*

What about beliefs? The basic postulates do not require beliefs to go beyond knowledge, so we cannot say much in general. An "optimistic" operator, however, may opt to believe that sources are experts unless we know otherwise, and thus maximise the information that can be (defeasibly) inferred from the sequence (in the next section we will introduce concrete operators obeying this principle). In this case we may believe that at least one source has expertise on x (i.e. $\mathsf{E}_a x \vee \mathsf{E}_b x \in B^{\sigma}_{c}$). Combined with $\neg \mathsf{E}_a x \in K^{\sigma}_{c}$, Closure *and* Containment*, we get $\mathsf{E}_b x \in B^{\sigma}_{c_2}$. Symmetry of expertise together with Proposition 8.3.3 part (2) then gives $\neg x \in B^{\sigma}_{c_2}$, i.e. we trust Dr. B in the example and believe patient 2 does not suffer from condition X.*

8.3.2 Model-Based Operators

While an operator is a purely syntactic object, it will be convenient to specify K^{σ} and B^{σ} in semantic terms by selecting a set of *possible* and *most plausible* worlds for each sequence σ. We call such operators *model-based*.

Definition 8.3.1. *An operator is* model-based *if for every σ there are sets $\mathcal{X}_\sigma, \mathcal{Y}_\sigma \subseteq \mathcal{W}$ such that (i) $\mathcal{X}_\sigma \supseteq \mathcal{Y}_\sigma$; (ii) $\Phi \in K^{\sigma}_{c}$ iff $W, c \models \Phi$ for all $W \in \mathcal{X}_\sigma$; and (iii) $\Phi \in B^{\sigma}_{c}$ iff $W, c \models \Phi$ for all $W \in \mathcal{Y}_\sigma$.*

In other words, K^{σ}_{c} (resp., B^{σ}_{c}) contains the formulas which hold at case c in *all worlds* in $\mathcal{X}_\sigma$ (resp., $\mathcal{Y}_\sigma$). It follows from the relevant definitions that $\mathcal{X}_\sigma \subseteq \mathrm{mod}(K^{\sigma})$, and equality holds if and only if $\mathcal{X}_\sigma$ is elementary (similarly for $\mathcal{Y}^{\sigma}$ and B^{σ}). Model-based operators are characterised by our first two basic postulates.

Theorem 8.3.1. *An operator satisfies* Closure *and* Containment *if and only if it is model-based.*

Proof. For ease of notation in what follows, write $\mathrm{mod}_c(\Phi) = \{W \in \mathcal{W} \mid W, c \models \Phi\}$.

"if": Suppose an operator $\sigma \mapsto \langle B^{\sigma}, K^{\sigma} \rangle$ is model-based. For *Closure*, we need to show that $B^{\sigma}_{c} \supseteq \mathrm{Cn}_c(B^{\sigma})$ and $K^{\sigma}_{c} \supseteq \mathrm{Cn}_c(K^{\sigma})$, for each c. Take any $\Phi \in \mathrm{Cn}_c(B^{\sigma})$. Then $\mathrm{mod}(B^{\sigma}) \subseteq \mathrm{mod}_c(\Phi)$. From the relevant definitions, one can easily check that $\mathcal{Y}_\sigma \subseteq \mathrm{mod}(B^{\sigma})$, so we have $\mathcal{Y}_\sigma \subseteq \mathrm{mod}_c(\Phi)$. That is, $W, c \models \Phi$ for all $W \in \mathcal{Y}_\sigma$. By definition of model-based operators, $\Phi \in B^{\sigma}_{c}$. The fact that $K^{\sigma}_{c} \supseteq \mathrm{Cn}_c(K^{\sigma})$ follows by an identical argument upon noticing that $\mathcal{X}_\sigma \subseteq \mathrm{mod}(K^{\sigma})$.

Containment follows from $\mathcal{X}_\sigma \supseteq \mathcal{Y}_\sigma$: if $\Phi \in K_c^\sigma$ then $W, c \models \Phi$ for all $W \in \mathcal{X}_\sigma$, and in particular this holds for all $W \in \mathcal{Y}_\sigma$. Hence $\Phi \in B_c^\sigma$, so $K^\sigma \sqsubseteq B^\sigma$.

"only if": Suppose an operator satisfies *Closure* and *Containment*. For any σ, set

$$\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma)$$

$$\mathcal{Y}_\sigma = \mathrm{mod}(B^\sigma)$$

We show the three properties required in Definition 8.3.1. $\mathcal{X}_\sigma \supseteq \mathcal{Y}_\sigma$ follows from *Containment* and the definition of a model of a collection. For the second property, note that $\Phi \in K_c^\sigma$ iff $\Phi \in \mathrm{Cn}_c(K^\sigma)$ by *Closure*, i.e. iff $\mathrm{mod}(K^\sigma) \subseteq \mathrm{mod}_c(\Phi)$. By choice of $\mathcal{X}_\sigma$, this holds exactly when $W, c \models \Phi$ for all $W \in \mathcal{X}_\sigma$, as required. The third property is proved using an identical argument. □

Since we take *Closure* and *Containment* to be fundamental properties, all operators we consider from now on will be model-based. We introduce our first concrete operator.

Definition 8.3.2. *Define the model-based operator* weak-mb *by*

$$\mathcal{X}_\sigma = \mathcal{Y}_\sigma = \{W \mid W, c \models \mathsf{S}_i\varphi \textit{ for all } \langle i, c, \varphi\rangle \in \sigma\}.$$

That is, the possible worlds $\mathcal{X}_\sigma$ are exactly those satisfying the soundness constraint for each report in σ, i.e. false reports are only due to lack of expertise of the corresponding source. Syntactically, $K^\sigma = B^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$.

Clearly weak-mb satisfies *Soundness*, and one can show that it satisfies all of the basic postulates of Section 8.3.1.[7] In fact, it is the *weakest* operator satisfying *Closure*, *Containment* and *Soundness*, in that for any other operator $\sigma \mapsto \langle \hat{B}^\sigma, \hat{K}^\sigma\rangle$ with these properties we have $B^\sigma \sqsubseteq \hat{B}^\sigma$ and $K^\sigma \sqsubseteq \hat{K}^\sigma$ for any σ.

Example 8.3.2. *Consider* weak-mb *applied to the sequence* $\sigma = (\langle *, c, p\rangle, \langle i, c, \neg p \wedge q\rangle)$. *By* Soundness, Closure *and the validities from Proposition 8.2.1, we have* $p \in K_c^\sigma$ *and* $\neg\mathsf{E}_i p \in K_c^\sigma$. *In fact, by* Closure, *we have* $\neg\mathsf{E}_i p \in K_d^\sigma$ *for all cases* d. *However, we cannot say much about* q*: neither* q, $\neg q$, $\mathsf{E}_i q$ *nor* $\neg\mathsf{E}_i q$ *are in* $B_c^\sigma = K_c^\sigma$.

8.4 Constructions

For model-based operators in Definition 8.3.1, the sets $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$ – which determine knowledge and belief – can depend on σ in a completely arbitrary manner. This lack of structure leads to very wide class of operators, and one cannot say much about them in general beyond the satisfaction of *Closure* and *Containment*. In this section we specialise model-based operators by providing two constructions.

[7] For *Consistency*, note that for any $*$-consistent sequence σ one can form a world W such that v_c is a model of all reports from $*$ at case c, and $\Pi_i = \{\mathcal{V}\}$ for all $i \neq *$. This satisfies all the soundness constraints, so $W \in \mathcal{X}_\sigma = \mathcal{Y}_\sigma$.

8.4.1 Conditioning Operators

Intuitively, $\mathcal{Y}_\sigma$ is supposed to represent the *most plausible* worlds among the possible worlds in $\mathcal{X}_\sigma$. This suggests the presence of a *plausibility ordering* on $\mathcal{X}_\sigma$, which is used to select $\mathcal{Y}_\sigma$. For our first construction we take this approach: we condition a fixed plausibility total preorder[8] on the knowledge $\mathcal{X}_\sigma$, and obtain $\mathcal{Y}_\sigma$ by selecting the minimal (i.e. most plausible) worlds.

Definition 8.4.1. *An operator is a* conditioning operator *if there is a total preorder $\leq$ on $\mathcal{W}$ and a mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma \rangle$ as in Definition 8.3.1 such that $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$ for all σ.*

Note that $\leq$ is independent of σ: it is fixed before receiving any reports. All conditioning operators are model-based by definition. Clearly $\mathcal{Y}_\sigma$ is determined by $\mathcal{X}_\sigma$ and the plausibility order, so that to define a conditioning operator it is enough to specify $\leq$ and the mapping $\sigma \mapsto \mathcal{X}_\sigma$. Write $W \simeq W'$ iff both $W \leq W'$ and $W' \leq W$. We now present examples of how such an ordering can be defined.

Definition 8.4.2. *Define the conditioning operator* var-based-cond *by setting $\mathcal{X}_\sigma$ in the same way as* weak-mb *in Definition 8.3.2, and $W \leq W'$ iff $r(W) \leq r(W')$, where*

$$r(W) = -\sum_{i \in \mathcal{S}} \left| \left\{ p \in \mathcal{P} \mid \Pi_i^W[p] = \mathrm{mod}_0(p) \right\} \right|.$$

var-based-cond aims to trust each source on *as many propositional variables* as possible. One can check that var-based-cond satisfies the basic postulates.

Example 8.4.1. *Revisiting the sequence $\sigma = (\langle *, c, p \rangle, \langle i, c, \neg p \wedge q \rangle)$ from Example 8.3.2 with* var-based-cond*, the knowledge set K_c^σ is the same as before, but we now have $q \wedge \mathsf{E}_i q \in B_c^\sigma$. This reflects the "credulous" behaviour of the ranking $\leq$: while it is not possible to believe i is an expert on p, we should believe they* are *an expert on q so long as this does not conflict with soundness. For the propositional beliefs generally, we have $[B_c^\sigma] = \mathrm{Cn}_0(p \wedge q)$. That is,* var-based-cond *takes the q part of the report from i (on which i is credulously trusted) while ignoring the $\neg p$ part (which is false due to report from $*$).*

Definition 8.4.3. *Define a conditioning operator* part-based-cond *with $\mathcal{X}_\sigma$ as for* var-based-cond*, and $\leq$ defined by the ranking function*

$$r(W) = -\sum_{i \in \mathcal{S}} |\Pi_i^W|.$$

part-based-cond aims to maximise the *number of cells* in the sources' partitions, and thereby maximise the number of propositions on which they have expertise. Unlike var-based-cond, the propositional variables play no special role. As expected, part-based-cond satisfies the basic postulates.

[8] A total preorder is a reflexive, transitive and total relation.

Example 8.4.2. *Applying* part-based-cond *to σ from Examples 8.3.2 and 8.4.1, we no longer extract q from the report of i: $q \notin B_c^\sigma$ and $\mathsf{E}_i q \notin B_c^\sigma$. Instead, we have $[B_c^\sigma] = \mathrm{Cn}_0(p)$, and $\mathsf{E}_i(p \vee q) \in B_c^\sigma$.*

Note that both var-based-cond and part-based-cond are based on the general principle of maximising the expertise of sources, subject to the constraint that all reports are sound. This intuition is formalised by the following postulate for conditioning operators. In what follows, write $W \preceq W'$ iff Π_i^W refines $\Pi_i^{W'}$ for all $i \in \mathcal{S}$, i.e. if all sources have broadly more expertise in W than in W'.[9]

Refinement *If $W \preceq W'$ then $W \leq W'$*

Since $\preceq$ is only a partial order on $\mathcal{W}$ there are many possible total extensions; var-based-cond and part-based-cond provide two specific examples.

We now turn to an axiomatic characterisation of conditioning operators. Taken with the basic postulates from Section 8.3.1, conditioning operators can be characterised using an approach similar to that of Delgrande, Peppas, and Woltran [26] in their account of *generalised AGM belief revision*.[10] This involves a technical property Delgrande, Peppas, and Woltran call *Acyc*, which finds its roots in the *Loop* property of Kraus, Lehmann, and Magidor [66].

Duplicate-removal *If $\rho_1 = \sigma \cdot \langle i, c, \varphi \rangle$ and $\rho_2 = \rho_1 \cdot \langle i, c, \varphi \rangle$ then $B^{\rho_1} = B^{\rho_2}$ and $K^{\rho_1} = K^{\rho_2}$*

Conditional-consistency *If K^σ is consistent then so is B^σ*

Inclusion-vacuity *$B^{\sigma \cdot \rho} \sqsubseteq \mathrm{Cn}(B^\sigma \sqcup K^\rho)$, with equality if $B^\sigma \sqcup K^\rho$ is consistent*

Acyc *If $\sigma_0, \ldots, \sigma_n$ are such that $K^{\sigma_j} \sqcup B^{\sigma_{j+1}}$ is consistent for all $0 \leq j < n$ and $K^{\sigma_n} \sqcup B^{\sigma_0}$ is consistent, then $K^{\sigma_0} \sqcup B^{\sigma_n}$ is consistent*

Inclusion-vacuity is so-named since it is analogous to the combination of *Inclusion* and *Vacuity* from AGM revision, if one informally views $B^{\sigma \cdot \rho}$ as the revision of B^σ by K^ρ. *Conditional-consistency* is another consistency postulate, which follows from *Consistency*, *Closure* and *Soundness*. *Acyc* is the analogue of the postulate of Delgrande, Peppas, and Woltran, which rules out cycles in the plausibility order constructed in the representation result.

As with the result of Delgrande, Peppas, and Woltran, a technical condition beyond Definition 8.4.1 is required to obtain the characterisation: say that a conditioning operator is *elementary* if for each σ the sets of worlds $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$ are elementary.[11]

[9] Π refines Π' if $\forall A \in \Pi, \exists B \in \Pi'$ such that $A \subseteq B$.

[10] Note that while the result is similar, our framework is not an instance of theirs.

[11] Equivalently, there is a total preorder $\leq$ such that $\mathrm{mod}(B^\sigma) = \min_\leq \mathrm{mod}(K^\sigma)$ for all σ.

Theorem 8.4.1. *Suppose an operator satisfies the basic postulates of Section 8.3.1.*[12] *Then it is an elementary conditioning operator if and only if it satisfies* Duplicate-removal, Conditional-consistency, Inclusion-vacuity *and* Acyc*.*

The proof is roughly follows the lines of Theorem 4.9 in [26], although some differences arise due to the form of our input as finite sequences of reports. First, we need a preliminary result.

Lemma 8.4.1. *For any model-based operator and sequence σ, $\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma)$ iff $\mathcal{X}_\sigma$ is elementary, and $\mathcal{Y}_\sigma = \mathrm{mod}(B^\sigma)$ iff $\mathcal{Y}_\sigma$ is elementary.*

Proof. We prove the result for $\mathcal{X}_\sigma$ and K^σ only. The "only if" direction is clear from the definition of an elementary set. For the "if" direction, suppose $\mathcal{X}_\sigma$ is elementary, i.e. $\mathcal{X}_\sigma = \mathrm{mod}(G)$ for some collection G. Since $\Phi \in K^\sigma_c$ iff $\mathcal{X}_\sigma \subseteq \mathrm{mod}_c(\Phi)$, we have $K^\sigma_c = \mathrm{Cn}_c(G)$, i.e. $K^\sigma = \mathrm{Cn}(G)$. Consequently $\mathrm{mod}(K^\sigma) = \mathrm{mod}(\mathrm{Cn}(G)) = \mathrm{mod}(G) = \mathcal{X}_\sigma$. □

We will prove the following result – slightly more general than Theorem 8.4.1 – from which Theorem 8.4.1 immediately follows.

Proposition 8.4.1. *Suppose an operator satisfies* Closure, Containment, K-conjunction *and* Equivalence*. Then it is an elementary conditioning operator if and only if it satisfies* Rearrangement, Duplicate-removal, Conditional-consistency, Inclusion-vacuity *and* Acyc*.*

Proof. Take some operator $\sigma \mapsto \langle B^\sigma, K^\sigma \rangle$ satisfying *Closure*, *Containment*, *K-conjunction* and *Equivalence*.

"if": Suppose the operator in question additionally satisfies *Rearrangement*, *Duplicate-removal*, *Conditional-consistency*, *Inclusion-vacuity* and *Acyc*. For any σ, set

$$\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma)$$
$$\mathcal{Y}_\sigma = \mathrm{mod}(B^\sigma)$$

Then – by *Closure* and *Containment* as shown in the proof of Theorem 8.3.1 – our operator is model based corresponding to this choice of $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$. Clearly both are elementary. We will construct a total preorder $\leq$ over $\mathcal{W}$ such that $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$; this will show the operator is an elementary conditioning operator.

First, fix a function $c : \mathcal{L}_0/\equiv \to \mathcal{L}_0$ which chooses a fixed representative of each equivalence class of logically equivalent propositional formulas, i.e. any mapping such that $c([\varphi]_\equiv) \equiv \varphi$. To simplify notation, write $\widehat{\varphi}$ for $c([\varphi]_\equiv)$. Then $\varphi \equiv \widehat{\varphi}$. Write $\widehat{\mathcal{L}_0} = \{\widehat{\varphi} \mid \varphi \in \mathcal{L}_0\}$. Note that $\widehat{\mathcal{L}_0}$ is finite (since we work with only finitely

[12] Strictly speaking, we only need *Closure*, *Containment*, *K-conjunction*, *Equivalence* and *Rearrangement*.

many propositional variables) and every formula in $\mathcal{L}_0$ is equivalent to exactly one formula in $\widehat{\mathcal{L}_0}$. For a sequence σ, let $\widehat{\sigma}$ be the result of replacing each report $\langle i, c, \varphi \rangle$ with $\langle i, c, \widehat{\varphi} \rangle$. Note that by *Rearrangement* and *Equivalence*, $\mathcal{X}_{\widehat{\sigma}} = \mathcal{X}_\sigma$ and $\mathcal{Y}_{\widehat{\sigma}} = \mathcal{Y}_\sigma$.

Now, for any world W, set

$$\mathcal{R}(W) = \{\langle i, c, \varphi \rangle \in \mathcal{S} \times \mathcal{C} \times \widehat{\mathcal{L}_0} \mid W \in \mathcal{X}_{\langle i,c,\varphi \rangle}\}$$

Note that $\mathcal{R}(W)$ is finite. For any pair of worlds W_1, W_2, let $\rho(W_1, W_2)$ be some enumeration of $\mathcal{R}(W_1) \cap \mathcal{R}(W_2)$. We establish some useful properties of $\rho(W_1, W_2)$.

Claim 8.4.1. *If $\rho(W_1, W_2) \neq \emptyset$, $W_1, W_2 \in \mathcal{X}_{\rho(W_1,W_2)}$.*

Proof. By *K-conjunction*, for any sequences σ, ρ we have $K^{\sigma \cdot \rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$. Taking the models of both sides, we have $\mathcal{X}_{\sigma \cdot \rho} = \mathcal{X}_\sigma \cap \mathcal{X}_\rho$. It follows that for $\rho(W_1, W_2) \neq \emptyset$,

$$\mathcal{X}_{\rho(W_1,W_2)} = \bigcap_{\langle i,c,\varphi \rangle \in \rho(W_1,W_2)} \mathcal{X}_{\langle i,c,\varphi \rangle}$$

If $\langle i, c, \varphi \rangle \in \rho(W_1, W_2)$ then $W_1, W_2 \in \mathcal{X}_{\langle i,c,\varphi \rangle}$ by definition. Hence $W_1, W_2 \in \mathcal{X}_{\rho(W_1,W_2)}$. $\square$

Claim 8.4.2. *If a sequence σ contains no equivalent reports (i.e. no distinct tuples $\langle i, c, \varphi \rangle$, $\langle i, c, \psi \rangle$ with $\varphi \equiv \psi$) and $W_1, W_2 \in \mathcal{X}_\sigma$, there is a sequence δ such that $W_1, W_2 \in \mathcal{X}_\delta$ and $\rho(W_1, W_2)$ is a permutation of $\widehat{\sigma} \cdot \delta$.*

Proof. If $\sigma = \emptyset$ then we can simply take $\delta = \rho(W_1, W_2)$. So suppose $\sigma \neq \emptyset$. By the same argument as in the proof of Claim 8.4.1, we have

$$\mathcal{X}_\sigma = \bigcap_{\langle i,c,\varphi \rangle \in \sigma} \mathcal{X}_{\langle i,c,\varphi \rangle}$$

Take any $\langle i, c, \varphi \rangle \in \widehat{\sigma}$. Then $\varphi \in \widehat{\mathcal{L}_0}$, and there is $\psi \equiv \varphi$ such that $\langle i, c, \psi \rangle \in \sigma$. By *Equivalence*, we have

$$W_1, W_2 \in \mathcal{X}_\sigma \subseteq \mathcal{X}_{\langle i,c,\psi \rangle} = \mathcal{X}_{\langle i,c,\varphi \rangle}$$

i.e. $\langle i, c, \varphi \rangle \in \mathcal{R}(W_1) \cap \mathcal{R}(W_2)$. Hence $\langle i, c, \varphi \rangle$ appears in $\rho(W_1, W_2)$. By the assumption that σ contains no equivalent reports, $\widehat{\sigma}$ contains no duplicates. It follows that $\rho(W_1, W_2)$ can be permuted so that $\widehat{\sigma}$ appears as a prefix. Taking δ to be the sequence that remains after $\widehat{\sigma}$ in this permutation, we clearly have that $\rho(W_1, W_2)$ is a permutation of $\widehat{\sigma} \cdot \delta$. Since $\sigma \neq \emptyset$ implies $\widehat{\sigma} \neq \emptyset$ and thus $\rho(W_1, W_2) \neq \emptyset$, by *Rearrangement*, *K-conjunction* and Claim 8.4.1 we get

$$W_1, W_2 \in \mathcal{X}_{\rho(W_1,W_2)} = \mathcal{X}_{\widehat{\sigma} \cdot \delta} = \mathcal{X}_{\widehat{\sigma}} \cap \mathcal{X}_\delta \subseteq \mathcal{X}_\delta$$

and we are done. $\square$

Now define a relation R on $\mathcal{W}$ by

$$WRW' \iff W = W' \text{ or } W \in \mathcal{Y}_{\rho(W,W')}$$

We have that any world in $\mathcal{Y}_\sigma$ R-precedes all worlds $\mathcal{X}_\sigma$.

Claim 8.4.3. *If $W \in \mathcal{Y}_\sigma$, then for all $W' \in \mathcal{X}_\sigma$ we have WRW'*

Proof. By *Rearrangement*, *Equivalence* and *Duplicate-removal*, we may assume without loss of generality that σ contains no distinct equivalent reports.

Let $W \in \mathcal{Y}_\sigma$ and $W' \in \mathcal{X}_\sigma$. Then $W \in \mathcal{X}_\sigma$ too. By Claim 8.4.2 and *Rearrangement*, there is some sequence δ such that $\mathcal{Y}_{\rho(W,W')} = \mathcal{Y}_{\widehat{\sigma}\cdot\delta}$ and $W, W' \in \mathcal{X}_\delta$. Consequently $W \in \mathcal{Y}_\sigma \cap \mathcal{X}_\delta = \mathcal{Y}_{\widehat{\sigma}} \cap \mathcal{X}_\delta$. Thus $B^{\widehat{\sigma}} \sqcup K^\delta$ is consistent. From *Inclusion-vacuity* we get

$$\mathcal{Y}_{\widehat{\sigma}\cdot\delta} = \mathcal{Y}_{\widehat{\sigma}} \cap \mathcal{X}_\delta$$

Thus

$$W \in \mathcal{Y}_{\widehat{\sigma}} \cap \mathcal{X}_\delta = \mathcal{Y}_{\widehat{\sigma}\cdot\delta} = \mathcal{Y}_{\rho(W,W')}$$

so WRW' as required. □

Now let $\leq_0$ be the transitive closure of R. Then $\leq_0$ is a (partial) preorder. By Claim 8.4.3, every world in $\mathcal{Y}_\sigma$ is $\leq_0$-minimal in $\mathcal{X}_\sigma$. In fact, the converse is also true.

Claim 8.4.4. *If $W \in \mathcal{X}_\sigma$ and there is no $W' \in \mathcal{X}_\sigma$ with $W' <_0 W$, then $W \in \mathcal{Y}_\sigma$.*

Proof. As before, assume without loss of generality that σ contains no distinct equivalent reports.

Take W as in the statement of the claim. Then $\mathcal{X}_\sigma \neq \emptyset$, so $\mathcal{Y}_\sigma \neq \emptyset$ by *Conditional-consistency*. Let $W' \in \mathcal{Y}_\sigma$. By Claim 8.4.3, $W'RW$ and thus $W' \leq_0 W$. But by assumption, $W' \not<_0 W$. So we must have $W \leq_0 W'$. By definition of $\leq_0$ as the transitive closure of R, there are $W_0, \dots W_n$ such that $W_0 = W$, $W_n = W'$ and

$$W_j R W_{j+1} \qquad (0 \leq j < n)$$

Without loss of generality, $n > 0$ and each of the W_j are distinct. From the definition of R, we therefore have that

$$W_j \in \rho(W_j, W_{j+1}) \qquad (0 \leq j < n)$$

Now set

$$\rho_j = \rho(W_j, W_{j+1}) \qquad (0 \leq j < n)$$
$$\rho_n = \rho(W_0, W_n)$$

Since $W'RW$, i.e. $W_n R W_0$, we in fact have $W_j \in \mathcal{Y}_{\rho_j}$ for all j (including $j = n$). For $j < n$, we also have $W_{j+1} \in \mathcal{X}_{\rho_j}$.[13] Consequently, for $j < n$ we have

$$W_{j+1} \in \mathcal{X}_{\rho_j} \cap \mathcal{Y}_{\rho_{j+1}}$$

i.e. $K^{\rho_j} \sqcup B^{\rho_{j+1}}$ is consistent. Moreover, $W_0 \in \mathcal{X}_{\rho_n} \cap \mathcal{Y}_{\rho_0}$, so $K^{\rho_n} \sqcup B^{\rho_0}$ is consistent. We can now apply *Acyc*: we get that $K^{\rho_0} \sqcup B^{\rho_n}$ is also consistent. On the one hand, *Inclusion-vacuity* and consistency of $K^{\rho_n} \sqcup B^{\rho_0}$ gives

$$B^{\rho_0 \cdot \rho_n} = \mathrm{Cn}(B^{\rho_0} \sqcup K^{\rho_n})$$

On the other, consistency of $B^{\rho_n} \sqcup K^{\rho_0}$ and *Rearrangement* gives

$$B^{\rho_0 \cdot \rho_n} = B^{\rho_n \cdot \rho_0} = \mathrm{Cn}(B^{\rho_n} \sqcup K^{\rho_0})$$

Combining these and taking models, we find

$$\mathcal{Y}_{\rho_0} \cap \mathcal{X}_{\rho_n} = \mathcal{Y}_{\rho_n} \cap \mathcal{X}_{\rho_0}$$

In particular, since W_0 lies in the set on the left-hand side, we have $W_0 \in \mathcal{Y}_{\rho_n}$.

Now, since $W_0, W_n \in \mathcal{X}_\sigma$ and $\rho_n = \rho(W_0, W_n)$, Claim 8.4.2 gives that there is δ with $W_0, W_n \in \mathcal{X}_\delta$ such that ρ_n is a permutation of $\widehat{\sigma} \cdot \delta$. Recalling that $W_n = W' \in \mathcal{Y}_\sigma = \mathcal{Y}_{\widehat{\sigma}}$ by assumption, we have $W_n \in \mathcal{Y}_{\widehat{\sigma}} \cap \mathcal{X}_\delta$, i.e. $B^{\widehat{\sigma}} \sqcup K^\delta$ is consistent. Applying *Inclusion-vacuity* once more, we get

$$B^{\rho_n} = B^{\widehat{\sigma} \cdot \delta} = \mathrm{Cn}(B^{\widehat{\sigma}} \sqcup K^\delta) = \mathrm{Cn}(B^\sigma \sqcup K^\delta)$$

Taking models of both sides,

$$\mathcal{Y}_{\rho_n} = \mathcal{Y}_\sigma \cap \mathcal{X}_\delta \subseteq \mathcal{Y}_\sigma$$

But we already saw that $W_0 \in \mathcal{Y}_{\rho_n}$. Hence $W_0 \in \mathcal{Y}_\sigma$. Since $W_0 = W$, we are done. $\square$

To complete the proof we extend $\leq_0$ to a *total* preorder and show that this does not affect the minimal elements of each $\mathcal{X}_\sigma$. Indeed, let $\leq$ be any total preorder extending $\leq_0$ and preserving strict inequalities, i.e. $\leq$ such that (i) $W \leq_0 W'$ implies $W \leq W'$; and (ii) $W <_0 W'$ implies $W < W'$.[14]

Claim 8.4.5. *For any sequence σ, $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$*

[13] If $\rho_j \neq \emptyset$ this follows from Claim 8.4.1. Otherwise, $W_{j+1} \in \mathcal{Y}_{\rho_{j+1}} \subseteq \mathcal{X}_{\rho_{j+1}} = \mathcal{X}_{\rho_{j+1} \cdot \emptyset} = \mathcal{X}_{\rho_{j+1}} \cap \mathcal{X}_\emptyset \subseteq \mathcal{X}_\emptyset = \mathcal{X}_{\rho_j}$ by *K-conjunction*.

[14] Such $\leq$ always exists. Indeed, note that $\leq_0$ induces a partial order on the equivalence classes of $\mathcal{W}$ with respect to the symmetric part of $\leq_0$ given by $W \simeq_0 W'$ iff $W \leq_0 W'$ and $W' \leq_0 W$. This partial order can be extended to a linear order $\leq^*$ on the equivalence classes. Taking $W \leq W'$ iff $[W] \leq^* [W']$, we obtain a total preorder on $\mathcal{W}$ with the desired properties.

Proof. Take any σ. For the left-to-right inclusion, take $W \in \mathcal{Y}_\sigma$. Then $W \in \mathcal{X}_\sigma$. Let $W' \in \mathcal{X}_\sigma$. By Claim 8.4.3, WRW', so $W \leq_0 W'$ and $W \leq W'$. Hence W is $\leq$-minimal in $\mathcal{X}_\sigma$.

For the right-to-left inclusion, take $W \in \min_\leq \mathcal{X}_\sigma$. Then for any $W' \in \mathcal{X}_\sigma$ we have $W \leq W'$. In particular, $W' \not< W$. By property (ii) of $\leq$, we have $W' \not<_0 W$. Since W' was an arbitrary member of $\mathcal{X}_\sigma$ and $W \in \mathcal{X}_\sigma$, the conditions of Claim 8.4.4 are satisfied, and we get $W \in \mathcal{Y}_\sigma$. □

This shows that our operator is an elementary conditioning operator as required.

"only if": Now suppose the operator is an elementary conditioning operator. i.e. there is a total preorder $\leq$ on $\mathcal{W}$ and a mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma \rangle$ such that for each σ, $\mathcal{Y}_\sigma = \min_\leq X_\sigma$, $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$ are elementary, and K^σ, B^σ are determined by $\mathcal{X}_\sigma$, $\mathcal{Y}_\sigma$ respectively according to Definition 8.3.1. By elementariness and Lemma 8.4.1, $\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma)$ and $\mathcal{Y}_\sigma = \mathrm{mod}(B^\sigma)$.

The following claim will be useful at various points.

Claim 8.4.6. *Suppose σ and ρ are such that $\mathcal{X}_\sigma = \mathcal{X}_\rho$. Then $K^\sigma = K^\rho$ and $B^\sigma = B^\rho$.*

Proof. Since the total preorder $\leq$ is fixed, we have

$$Y_\sigma = \min_\leq \mathcal{X}_\sigma = \min_\leq \mathcal{X}_\rho = \mathcal{Y}_\rho$$

Now, $\mathcal{X}_\sigma = \mathcal{X}_\rho$ means $\mathrm{mod}(K^\sigma) = \mathrm{mod}(K^\rho)$, so $\mathrm{Cn}(K^\sigma) = \mathrm{Cn}(K^\rho)$. By *Closure*, $K^\sigma = K^\rho$. Similarly, $\mathcal{Y}_\sigma = \mathcal{Y}_\rho$ gives $B^\sigma = B^\rho$. □

We take the postulates to be shown in turn.

- *Rearrangement*: Suppose σ is a permutation of ρ. Without loss of generality, $\sigma, \rho \neq \emptyset$. Repeated application of *K-conjunction* gives

$$\mathcal{X}_\sigma = \bigcap_{\langle i,c,\varphi \rangle \in \sigma} \mathcal{X}_{\langle i,c,\varphi \rangle}$$

  Since σ and ρ contain exactly the same reports – just in a different order – commutativity and associativity of intersection of sets gives $\mathcal{X}_\sigma = \mathcal{X}_\rho$. *Rearrangement* follows from Claim 8.4.6.

- *Duplicate-removal*: Let σ, ρ_1 and ρ_2 be as in the statement of *Duplicate-removal*. Then by *K-conjunction*,

$$\begin{aligned}
\mathcal{X}_{\rho_2} &= \mathcal{X}_{\rho_1 \cdot \langle i,c,\varphi \rangle} \\
&= \mathcal{X}_{\rho_1} \cap \mathcal{X}_{\langle i,c,\varphi \rangle} \\
&= \mathcal{X}_{\sigma \cdot \langle i,c,\varphi \rangle} \cap \mathcal{X}_{\langle i,c,\varphi \rangle} \\
&= \mathcal{X}_\sigma \cap \mathcal{X}_{\langle i,c,\varphi \rangle} \cap \mathcal{X}_{\langle i,c,\varphi \rangle} \\
&= \mathcal{X}_\sigma \cap \mathcal{X}_{\langle i,c,\varphi \rangle} \\
&= \mathcal{X}_{\rho_1}
\end{aligned}$$

and we may conclude by Claim 8.4.6.

- *Conditional-consistency*: Suppose K^σ is consistent, i.e. $\mathcal{X}_\sigma \neq \emptyset$. Since $\mathcal{W}$ is finite, $\mathcal{X}_\sigma$ is finite and thus some $\leq$-minimal world must exist in $\mathcal{X}_\sigma$. Hence $\mathcal{Y}_\sigma \neq \emptyset$, so B^σ is consistent.

- *Inclusion-vacuity*: Take any sequences σ, ρ. First we show $B^{\sigma\cdot\rho} \sqsubseteq \mathrm{Cn}(B^\sigma \sqcup K^\rho)$, or equivalently, $\mathcal{Y}_{\sigma\cdot\rho} \supseteq \mathcal{Y}_\sigma \cap \mathcal{X}_\rho$. Suppose $W \in \mathcal{Y}_\sigma \cap \mathcal{X}_\rho$. Since $\mathcal{Y}_\sigma \subseteq \mathcal{X}_\sigma$, we have $W \in \mathcal{X}_\sigma \cap \mathcal{X}_\rho = \mathcal{X}_{\sigma\cdot\rho}$ by *K-conjunction*. We need to show W is minimal. Take any $W' \in \mathcal{X}_{\sigma\cdot\rho}$. Then $W' \in \mathcal{X}_\sigma$, so $W \in \mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$ gives $W \leq W'$. Hence $W \in \min_\leq X_{\sigma\cdot\rho} = \mathcal{Y}_{\sigma\cdot\rho}$.

  Now suppose $B^\sigma \sqcup K^\rho$ is consistent, i.e. $\mathcal{Y}_\sigma \cap \mathcal{X}_\rho \neq \emptyset$. Take some $\widehat{W} \in \mathcal{Y}_\sigma \cap \mathcal{X}_\rho$. We need to show $B^{\sigma\cdot\rho} \sqsupseteq \mathrm{Cn}(B^\sigma \sqcup K^\rho)$, i.e. $\mathcal{Y}_{\sigma\cdot\rho} \subseteq \mathcal{Y}_\sigma \cap \mathcal{X}_\rho$. To that end, let $W \in \mathcal{Y}_{\sigma\cdot\rho}$. Then $W \in \mathcal{X}_{\sigma\cdot\rho} = \mathcal{X}_\sigma \cap \mathcal{X}_\rho \subseteq \mathcal{X}_\rho$, so we only need to show $W \in \mathcal{Y}_\sigma$. Take any $W' \in \mathcal{X}_\sigma$. Then $\widehat{W} \in \mathcal{Y}_\sigma$ gives $\widehat{W} \leq W'$. But $\widehat{W} \in \mathcal{X}_\sigma \cap \mathcal{X}_\rho = \mathcal{X}_{\sigma\cdot\rho}$ and $W \in \mathcal{Y}_{\sigma\cdot\rho}$ gives $W \leq \widehat{W}$. By transitivity of $\leq$, we have $W \leq W'$. Hence $W \in \min_\leq \mathcal{X}_\sigma = \mathcal{Y}_\sigma$.

- *Acyc*: Let $\sigma_0, \ldots, \sigma_n$ be as in the statement of *Acyc*. Without loss of generality, $n > 0$. Then there are $W_0, \ldots, W_n$ such that

$$\begin{aligned} &W_j \in \mathcal{X}_{\sigma_j} \cap \mathcal{Y}_{\sigma_{j+1}} \qquad (0 \leq j < n) \\ &W_n \in \mathcal{X}_{\sigma_n} \cap \mathcal{Y}_{\sigma_0} \end{aligned}$$

Note that $W_j \in \mathcal{X}_{\sigma_j}$ for all j. For $j < n$, we also have $W_j \in \mathcal{Y}_{\sigma_{j+1}} = \min_\leq X_{\sigma_{j+1}}$. It follows that $W_j \leq W_{j+1}$ for such j, so

$$W_0 \leq \cdots \leq W_n$$

But we also have $W_n \in \mathcal{Y}_{\sigma_0} = \min_\leq \mathcal{X}_{\sigma_0}$ and $W_0 \in \mathcal{X}_{\sigma_0}$, so $W_n \leq W_0$. By transitivity of $\leq$, the chain flattens: we have

$$W_0 \simeq \cdots \simeq W_n$$

Now note that since $W_{n-1} \in \mathcal{Y}_{\sigma_n}$, W_{n-1} is minimal in $\mathcal{X}_{\sigma_n}$. But $W_n \in \mathcal{X}_{\sigma_n}$ and $W_{n-1} \simeq W_n$ by the above, so in fact $W_n \in \mathcal{Y}_{\sigma_n}$ too. Hence

$$\begin{aligned} W_n &\in \mathcal{Y}_{\sigma_0} \cap \mathcal{Y}_{\sigma_n} \\ &\subseteq \mathcal{X}_{\sigma_0} \cap \mathcal{Y}_{\sigma_n} \\ &= \mathrm{mod}(K^{\sigma_0} \sqcup B^{\sigma_n}) \end{aligned}$$

i.e. $K^{\sigma_0} \sqcup B^{\sigma_n}$ is consistent, as required for *Acyc*.

□

Note that while the requirement in Theorem 8.4.1 that $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$ are elementary is a technical condition,[15] the characterisation in Proposition 8.2.2 implies a simple sufficient condition for elementariness.

Proposition 8.4.2. *Suppose $\leq$ is such that $W \simeq W'$ whenever W and W' are partition-equivalent. Then $\min_\leq S$ is elementary for any elementary set $S \subseteq \mathcal{W}$.*

Proof. We use the characterisation of elementary sets from Proposition 8.2.2. Take $S \subseteq \mathcal{W}$ elementary. Suppose $W \in \mathcal{W}$, $W_1, W_2 \in \min_\leq S$ are such that W is partition-equivalent to both W_1, W_2 and W is a valuation combination from $\{W_1, W_2\}$. By hypothesis we have $W \simeq W_1 \simeq W_2$.

Now since $\min_\leq S \subseteq S$, we have $W_1, W_2 \in S$. Since S is elementary, $W \in S$. But now $W \simeq W_1$ and $W_1 \in \min_\leq S$ gives $W \in \min_\leq S$. This shows the required closure property for $\min_\leq S$, and we are done. $\square$

Proposition 8.4.2 implies that var-based-cond and part-based-cond are elementary. Indeed, for both operators $\mathcal{X}_\sigma = \mathrm{mod}(G^\sigma_{\mathsf{snd}})$ so is elementary by definition. Since the ranking $\leq$ for each operator only depends on the partitions of worlds, $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$ is elementary also.

8.4.2 Score-Based Operators

The fact that the plausibility order $\leq$ of a conditioning operator is fixed may be too limiting. For example, consider

$$\sigma = (\langle i, c, p\rangle, \langle j, c, \neg p\rangle, \langle i, d, p\rangle).$$

If one sets $\mathcal{X}_\sigma$ to satisfy the soundness constraints (i.e. as in weak-mb), there is a possible world $W_1 \in \mathcal{X}_\sigma$ with $W_1, d \models \neg \mathsf{E}_i p \wedge \mathsf{E}_j p \wedge \neg p$ (i.e. W_1 sides with source j and p is false at d) and another world $W_2 \in \mathcal{X}_\sigma$ with $W_2, d \models \mathsf{E}_i p \wedge \neg \mathsf{E}_j p \wedge p$ (i.e. W_2 sides with source i). Appealing to symmetry, one may argue that neither world is *a priori* more plausible than the other, so any fixed plausibility order should have $W_1 \simeq W_2$. If these worlds are maximally plausible (e.g. if taking the "optimistic" view outlined in Example 8.3.1), conditioning gives $p \notin B^\sigma_d$ and $\neg p \notin B^\sigma_d$. However, there is an argument that W_2 should be considered more plausible than W_1 *given the sequence* σ, since W_2 validates the final report $\langle i, d, p\rangle$ whereas W_1 does not. Consequently, there is an argument that we should in fact have $p \in B^\sigma_d$.[16] This shows that we need the plausibility order to be responsive to the input sequence for adequate belief change.[17]

[15] *Inclusion-vacuity* may fail for non-elementary conditioning.
[16] At the very least, the case $p \in B^\sigma_d$ should not be *excluded*.
[17] In Section 8.5 we make this argument more precise by providing an impossibility result which shows conditioning operators with some basic properties cannot accept p in sequences such as this.

As a result of this discussion, we look for operators whose plausibility ordering can depend on σ. One approach to achieve this in a controlled way is to have a ranking for each *report* $\langle i, c, \varphi\rangle$, and combine these to construct a ranking for each sequence σ. We represent these rankings by *scoring functions*, and call the resulting operators *score-based*.

Definition 8.4.4. *An operator is* score-based *if there is a mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma\rangle$ as in Definition 8.3.1 and functions $r_0 : \mathcal{W} \to \mathbb{N} \cup \{\infty\}$, $d : \mathcal{W} \times (\mathcal{S} \times \mathcal{C} \times \mathcal{L}_0) \to \mathbb{N} \cup \{\infty\}$ such that $\mathcal{X}_\sigma = \{W \mid r_\sigma(W) < \infty\}$ and $\mathcal{Y}_\sigma = \operatorname{argmin}_{W\in\mathcal{X}_\sigma} r_\sigma(W)$, where*

$$r_\sigma(W) = r_0(W) + \sum\nolimits_{\langle i,c,\varphi\rangle\in\sigma} d(W, \langle i, c, \varphi\rangle).$$

Here $r_0(W)$ is the *prior implausibility score* of W, and $d(W, \langle i, c, \varphi\rangle)$ is the *disagreement score* for world W and $\langle i, c, \varphi\rangle$. The set of most plausible worlds $\mathcal{Y}_\sigma$ consists of those W which minimise the sum of the prior implausibility and the total disagreement with σ. Note that by summing the scores of each report $\langle i, c, \varphi\rangle$ with equal weight, we treat each report independently. Score-based operators generalise elementary conditioning operators with *K-conjunction*.

Proposition 8.4.3. *Any elementary conditioning operator satisfying* K-conjunction *is score-based.*

Proof. Take any elementary conditioning operator corresponding to some mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma\rangle$ and total preorder $\leq$, and suppose *K-conjunction* holds. Write

$$k(W) = |\{W' \in \mathcal{W} \mid W' \leq W\}|$$

Then we have $W \leq W'$ iff $k(W) \leq k(W')$. Set

$$r_0(W) = \begin{cases} \infty, & W \notin \mathcal{X}_\emptyset \\ k(W), & W \in \mathcal{X}_\emptyset \end{cases},$$

$$d(W, \langle i, c, \varphi\rangle) = \begin{cases} \infty, & W \notin \mathcal{X}_{\langle i,c,\varphi\rangle} \\ 0, & W \in \mathcal{X}_{\langle i,c,\varphi\rangle} \end{cases}.$$

For any sequence σ, repeated applications of *K-conjunction* (and the fact that $\mathcal{X}_\sigma$ is elementary) give $r_\sigma(W) < \infty$ iff $W \in \mathcal{X}_\sigma$. Similarly, the choice of r_0 gives $\operatorname{argmin}_{W\in\mathcal{X}_\sigma} r_\sigma(W) = \min_\leq \mathcal{X}_\sigma = \mathcal{Y}_\sigma$. Hence the operator is score-based.

□

We now give a concrete example.

Definition 8.4.5. *Define a score-based operator* excess-min *by setting $r_0(W) = 0$ and*

$$d(W, \langle i, c, \varphi\rangle) = \begin{cases} |\Pi_i^W[\varphi] \setminus \mathrm{mod}_0(\varphi)|, & W, c \models \mathsf{S}_i\varphi \\ \infty, & \textit{otherwise.} \end{cases}$$

The set of possible worlds $\mathcal{X}_\sigma$ is the same as for the earlier operators. All worlds are *a priori* equiplausible according to r_0. The disagreement score d is defined as the number of propositional valuations in the "excess" of $\Pi_i^W[\varphi]$ which are not models of φ, i.e. the number of $\neg\varphi$ valuations which are indistinguishable from some φ valuation. The intuition here is that *sources tend to only report formulas on which they have expertise*. The minimum score 0 is attained exactly when i has expertise on φ; other worlds are ordered by how much they deviate from this ideal.

One can verify that excess-min satisfies the basic postulates of Section 8.3.1. It can also be seen that $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$ are elementary, and excess-min fails *Duplicate-removal* and *Inclusion-vacuity*. It follows from Theorem 8.4.1 that excess-min is *not* a conditioning operator.[18]

Example 8.4.3. *To illustrate the differences between* excess-min *and conditioning, consider a more elaborate version of the example given at the start of this section:*

$$\sigma = (\langle i, c, p \to q\rangle, \langle j, c, p \to \neg q\rangle, \langle *, c, p\rangle, \langle i, d, p\rangle, \langle i, d, q\rangle).$$

Here the reports of i and j in case c are consistent, but inconsistent when taken with the reliable information p from $*$. Should we believe q or $\neg q$? Both our conditioning operators var-based-cond *and* part-based-cond *decline to decide, and have $[B_c^\sigma] = \mathrm{Cn}_0(p)$. However, since* excess-min *takes into account each report in the sequence, the fact that i reports both p and q in case d leads to $\mathsf{E}_i p \wedge \mathsf{E}_i q \in B_c^\sigma$. This gives $\mathsf{E}_i(p \to q) \in B_c^\sigma$ by Proposition 8.2.1 part (3), so we can make use of the report from i in case c: we have $[B_c^\sigma] = \mathrm{Cn}_0(p \wedge q)$. This example shows that score-based operators can be* more credulous *than conditioning operators (e.g. we can believe $\mathsf{E}_i p$ when i reports p), and can consequently hold stronger propositional beliefs.*

8.5 One-Step Revision

The postulates of Section 8.3.1 only set out very basic requirements for an operator. In this section we introduce some more demanding postulates which address how beliefs should change when a sequence σ is extended by a new report $\langle i, c, \varphi\rangle$. In view of *Rearrangement*, we do not view this process as *revision* of B^σ by $\langle i, c, \varphi\rangle$, but rather as *reinterpretation* of σ in light of a new report $\langle i, c, \varphi\rangle$. The postulates we introduce can therefore be seen as *coherency* requirements, which place some constraints on this reinterpretation.

First, we address how propositional beliefs should be affected by reliable information.

[18] We will later give an alternative proof of this fact, via an impossibility result for conditioning operators (Proposition 8.5.3).

AGM-$*$ *For any σ and $c \in \mathcal{C}$ there is an AGM operator $\star$ for $[B_c^\sigma]$ such that $\left[B_c^{\sigma \cdot \langle *, c, \varphi \rangle}\right] = [B_c^\sigma] \star \varphi$ whenever $\neg\varphi \notin K_c^\sigma$*

AGM-$*$ says that receiving information from the reliable source $*$ acts in accordance with the well-known AGM postulates [1] for propositional belief revision (provided we are not in the degenerate case where the new report φ was already *known* to be false). Since AGM revision operators are characterised by total preorders over valuations [50, 59], it is no surprise that our order-based constructions are consistent with *AGM-$*$*.

[TODO: flow in what follows.]

Proposition 8.5.1. var-based-cond, part-based-cond *and* excess-min *satisfy* AGM-$*$.

We require some preliminary results. For a case $c \in \mathcal{C}$ and valuation $v \in \mathcal{V}$, write $\mathcal{W}_{c\,:\,v} = \{W \in \mathcal{W} \mid v_c^W = v\}$ for the set of worlds whose c valuation is v.

Lemma 8.5.1. *For any model-based operator, sequence σ, case c, and valuation v in$\mathcal{V}$,*

$$v \in \mathrm{mod}_0\,[B_c^\sigma] \iff \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v} \neq \emptyset$$

Proof. " $\implies$ ": We show the contrapositive. Suppose $\mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v} = \emptyset$. Let ψ be any propositional formula such that $\mathrm{mod}_0(\psi) = \mathcal{V} \setminus \{v\}$. Now for any $W \in \mathcal{Y}_\sigma$, we have $W \notin \mathcal{W}_{c\,:\,v}$, i.e. $v_c^W \neq v$. Hence $v_c^W \in \mathrm{mod}_0(\psi)$, so $W, c \models \psi$. By definition of the belief set of a model-based operator, we have $\psi \in B_c^\sigma$. But ψ is a propositional formula, so $\psi \in [B_c^\sigma]$. Since $v \notin \mathrm{mod}_0(\psi)$, we have $v \notin \mathrm{mod}_0\,[B_c^\sigma]$.

" $\impliedby$ ": Suppose there is some $W \in \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v}$. Let $\varphi \in [B_c^\sigma]$. Then, in particular, $\varphi \in B_c^\sigma$, so $W, c \models \varphi$ by $W \in \mathcal{Y}_\sigma$ and the definition of the model-based belief set. That is, $v = v_c^W \in \mathrm{mod}_0(\varphi)$. Since $\varphi \in [B_c^\sigma]$ was arbitrary, we have $v \in \mathrm{mod}_0\,[B_c^\sigma]$.

□

We have a sufficient condition for *AGM-$*$* for score-based operators.

Lemma 8.5.2. *Suppose a score-based operator is such that for each $c \in \mathcal{C}$ and $\varphi \in \mathcal{L}_0$ there is a constant $M \in \mathbb{N}$ with*

$$d(W, \langle *, c, \varphi \rangle) = \begin{cases} M, & W, c \models \varphi \\ \infty, & W, c \models \neg\varphi \end{cases}$$

for all W. Then AGM-$*$ *holds.*

Proof. Take a score-based operator with the stated property. Let σ be a sequence and take $c \in \mathcal{C}$. Without loss of generality, there is some $\varphi \in \mathcal{L}_0$ such that $\neg\varphi \notin K_c^\sigma$ (otherwise *AGM-$*$* trivially holds). Since any score-based operator is model-based and therefore satisfies *Closure*, we have that K^σ is inconsistent iff $K^\sigma_c = \mathcal{L}$. But since K^σ_c does not contain $\neg\varphi$, it must be the case that K^σ is consistent.

Now, set

$$k(v) = \min\{r_\sigma(W) \mid W \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v}\}$$

where $\min\emptyset = \infty$. Note that $k(v) = \infty$ if and only if $\mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v} = \emptyset$. Then k defines a total preorder $\preceq$ on valuations, where $v \preceq v'$ iff $k(v) \leq k(v')$. Define a propositional revision operator $\star$ for $[B^\sigma_c]$ by

$$[B^\sigma_c] \star \varphi = \{\psi \in \mathcal{L}_0 \mid \min\nolimits_\preceq \mathrm{mod}_0(\varphi) \subseteq \mathrm{mod}_0(\psi)\}$$

To show that $\star$ satisfies the AGM postulates (for $[B^\sigma_c]$) it is sufficient to show that the models of $[B^\sigma_c]$ are exactly the $\preceq$-minimal valuations.

Claim 8.5.1. $\mathrm{mod}_0\,[B^\sigma_c] = \min_\preceq \mathcal{V}$.

Proof. "$\subseteq$": let $v \in \mathrm{mod}_0\,[B^\sigma_c]$. By Lemma 8.5.1, there is some $W \in \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v}$. Since $W \in \mathcal{X}_\sigma$ too, by definition of k we have $k(v) \leq r_\sigma(W) < \infty$. Now let $v' \in \mathcal{V}$. Without loss of generality assume $k(v') < \infty$. Then there is some $W' \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v'}$ such that $k(v') = r_\sigma(W')$. But $W' \in \mathcal{X}_\sigma$ and $W \in \mathcal{Y}_\sigma$ gives $r_\sigma(W) \leq r_\sigma(W')$, so

$$k(v) \leq r_\sigma(W) \leq r_\sigma(W') = k(v')$$

i.e. $v \preceq v'$. Hence v is $\preceq$-minimal.

"$\supseteq$": let $v \in \min_\preceq \mathcal{V}$. Since K^σ is consistent, there is some $\hat{W} \in \mathcal{X}_\sigma$. Writing $\hat{v} = v^{\hat{W}}_c$, we have $\hat{W} \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,\hat{v}}$, so $v \preceq \hat{v}$ implies

$$k(v) \leq k(\hat{v}) \leq r_\sigma(\hat{W}) < \infty$$

Hence there must be some $W \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v}$ such that $k(v) = r_\sigma(W)$. We claim that, in fact, $W \in \mathcal{Y}_\sigma$. Indeed, for any $W' \in \mathcal{X}_\sigma$ we have $v \preceq v^{W'}_c$, so

$$r_\sigma(W) = k(v) \leq k(v^{W'}_c) \leq r_\sigma(W')$$

That is, $W \in \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v}$. By Lemma 8.5.1, $v \in \mathrm{mod}_0\,[B^\sigma_c]$.

□

So, $\star$ is indeed an AGM operator for $[B^\sigma_c]$. Now take $\varphi \in \mathcal{L}_0$ such that $\neg\varphi \notin K^\sigma_c$. Write $\rho = \sigma \cdot \langle *, c, \varphi\rangle$. We claim the following.

Claim 8.5.2. $\mathrm{mod}_0\,[B^\rho_c] = \min_\preceq \mathrm{mod}_0(\varphi)$.

Proof. "$\subseteq$": let $v \in \mathrm{mod}_0\,[B^\rho_c]$. By Lemma 8.5.1 again, there is some $W \in \mathcal{Y}_\rho \cap \mathcal{W}_{c\,:\,v}$. Since $\langle *, c, \varphi\rangle \in \rho$ and $d(W, \langle *, c, \varphi\rangle) \leq r_\rho(W) < \infty$, we must have $W, c \models \varphi$ by the assumed property of the score function d. Hence $v = v^W_c \in \mathrm{mod}_0(\varphi)$.

Now since $\mathcal{Y}_\rho \subseteq \mathcal{X}_\rho$, we have $W \in \mathcal{Y}_\rho \subseteq \mathcal{X}_\rho \subseteq \mathcal{X}_\sigma$, so $W \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v}$. By definition of k, we have $k(v) \leq r_\sigma(W)$. Take any $v' \in \mathrm{mod}_0(\varphi)$. Without loss of generality, assume $k(v') < \infty$, so that there is some $W' \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v'}$ with $k(v') = r_\sigma(W')$. Since $v_c^{W'} = v' \in \mathrm{mod}_0(\varphi)$, we have $W', c \models \varphi$. Consequently, by the property of d again, $d(W', \langle *, c, \varphi\rangle) = M$. Since $W' \in \mathcal{X}_\sigma$ gives $r_\sigma(W') < \infty$, it follows that

$$r_\rho(W') = r_\sigma(W') + M < \infty$$

so $W' \in \mathcal{X}_\rho$.

Recall that $W, c \models \varphi$ too, so $d(W, \langle *, c, \varphi\rangle) = M$ also. From $W \in \mathcal{Y}_\rho$ and $W' \in \mathcal{X}_\rho$ we get

$$\begin{aligned} r_\sigma(W) &= r_\rho(W) - M \\ &\leq r_\rho(W') - M \\ &= r_\rho(W') - d(W', \langle *, c, \varphi\rangle) \\ &= r_\sigma(W') \end{aligned}$$

This yields

$$k(v) \leq r_\sigma(W) \leq r_\sigma(W') = k(v')$$

and $v \preceq v'$ as required.

"$\supseteq$": let $v \in \min_{\preceq} \mathrm{mod}_0(\varphi)$. Since $\neg\varphi \notin K_c^\sigma$, there is some $\hat{W} \in \mathcal{X}_\sigma$ such that $\hat{W}, c \models \varphi$. Writing $\hat{v} = v_c^{\hat{W}}$, we have $\hat{v} \in \mathrm{mod}_0(\varphi)$. Hence $v \preceq \hat{v}$. This implies

$$k(v) \leq k(\hat{v}) \leq r_\sigma(\hat{W}) < \infty$$

so there must be some $W \in \mathcal{X}_\sigma \cap \mathcal{W}_{c\,:\,v}$ with $k(v) = r_\sigma(W)$. Since $v_c^W = v \in \mathrm{mod}_0(\varphi)$, we have $W, c \models \varphi$. By the assumed property of d, we get $d(W, \langle *, c, \varphi\rangle) = M$. Hence

$$r_\rho(W) = r_\sigma(W) + d(W, \langle *, c, \varphi\rangle) = r_\sigma(W) + M < \infty$$

so $W \in \mathcal{X}_\rho$ too. We will show that $W \in \mathcal{Y}_\rho$. Let $W' \in \mathcal{X}_\rho$. Then we must have $d(W', \langle *, c, \varphi\rangle) = M$ and $W', c \models \varphi$. That is, $v_c^{W'} \in \mathrm{mod}_0(\varphi)$. By minimality of v, we have $v \preceq v_c^{W'}$. Noting that $W' \in \mathcal{X}_\rho \subseteq \mathcal{X}_\sigma$, we get

$$r_\sigma(W) = k(v) \leq k(v_c^{W'}) \leq r_\sigma(W')$$

Consequently,

$$r_\rho(W) = r_\sigma(W) + M \leq r_\sigma(W') + M = r_\rho(W')$$

This shows $W \in \mathcal{Y}_\rho$, i.e. $\mathcal{Y}_\rho \cap \mathcal{W}_{c\,:\,v} \neq \emptyset$. By Lemma 8.5.1, we are done.

□

Noting that $\mathrm{mod}_0\,[B_c^\sigma] \star \varphi = \min_{\preceq} \mathrm{mod}_0(\varphi)$, it follows from Claim 8.5.2 that $\mathrm{Cn}_0([B_c^\rho]) = \mathrm{Cn}_0([B_c^\sigma] \star \varphi)$. But $[B_c^\rho]$ is deductively closed by *Closure*, and $[B_c^\sigma] \star \varphi$ is deductively closed by construction. Hence $[B_c^\rho] = [B_c^\sigma] \star \varphi$, as required for *AGM-$*$*.

□

As a consequence of Proposition 8.4.3 (and the construction of d in its proof), one can apply Lemma 8.5.2 with $M = 0$ for conditioning operators with *K-conjunction* and a certain natural property.

Corollary 8.5.1. *Suppose an elementary conditioning operator satisfying* K-conjunction *has the property that*

$$W \in \mathcal{X}_{\langle *, c, \varphi \rangle} \iff W, c \models \varphi$$

Then AGM-$*$ *holds.*

We can now prove Proposition 8.5.1.

Proof of Proposition 8.5.1. For the conditioning operators var-based-cond and part-based-cond, it is easily verified that the condition in Corollary 8.5.1 holds, and thus *AGM-$*$* does also.

For the score-based operator excess-min, we may use Lemma 8.5.2 with $M = 0$. □

Thus, we do indeed extend AGM revision in the case of reliable information. What about non-reliable information? First note that the analogue of *AGM-$*$* for ordinary sources $i \neq *$ is *not* desirable. In particular, we should not have the *Success* postulate:

$$\varphi \in B_c^{\sigma \cdot \langle i, c, \varphi \rangle}.$$

Indeed, the sequence in Example 8.3.2 with $\varphi = \neg p \wedge q$ already shows that *Success* would conflict with the basic postulates. However, there are weaker modifications of *Success* which may be more appropriate. We consider two such postulates.

Cond-success *If* $\mathsf{E}_i\varphi \in B_c^\sigma$ *and* $\neg\varphi \notin B_c^\sigma$*, then* $\varphi \in B_c^{\sigma \cdot \langle i, c, \varphi \rangle}$

Strong-cond-success *If* $\neg(\mathsf{E}_i\varphi \wedge \varphi) \notin B_c^\sigma$*, then* $\varphi \in B_c^{\sigma \cdot \langle i, c, \varphi \rangle}$

Cond-success says that if i is deemed an expert on φ, which is consistent with current beliefs, then φ is accepted after i reports it. That is, the acceptance of φ is *conditional* on prior beliefs about the expertise of i (on φ). *Strong-cond-success* weakens the antecedent by only requiring that $\mathsf{E}_i\varphi$ and φ are jointly consistent with current beliefs (i.e. i need not be considered an expert on φ). In other words, we should believe reports if there is no reason not to. It is easily shown that *Closure* and *Strong-cond-success* implies *Cond-success*. We once again revisit our examples. **[TODO:** flow, again.]

Proposition 8.5.2. var-based-cond, part-based-cond *and* excess-min *satisfy* Cond-success, *and* excess-min *additionally satisfies* Strong-cond-success.

As a first step in the proof, we present sufficient conditions for conditioning operators to satisfy *Cond-success*. In fact, we do not need to impose any condition on the total preorder $\leq$: a natural constraint on the mapping $\sigma \mapsto \mathcal{X}_\sigma$ (together with some basic postulates) is enough.

Lemma 8.5.3. *Suppose an elementary conditioning operator satisfies* K-conjunction, Soundness *and*

$$W, c \models \varphi \implies W \in \mathcal{X}_{\langle i,c,\varphi\rangle}$$

Then Cond-success *holds.*

Proof. Suppose an elementary conditioning operator corresponding to the mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma \rangle$ and total preorder $\leq$ satisfies *K-conjunction*, *Soundness* and has the stated property.

Let σ be a sequence and $c \in \mathcal{C}$. Suppose $\mathsf{E}_i\varphi \in B_c^\sigma$ and $\neg\varphi \notin B_c^\sigma$. Write $\rho = \sigma \cdot \langle i, c, \varphi \rangle$. We need to show $\varphi \in B_c^\rho$.

By $\neg\varphi \notin B_c^\sigma$, there is some $W \in \mathcal{Y}_\sigma$ such that $W, c \models \varphi$. Hence $W \in \mathcal{X}_{\langle i,c,\varphi\rangle}$. By elementariness and *K-conjunction*, we have $\mathcal{X}_\rho = \mathcal{X}_\sigma \cap \mathcal{X}_{\langle i,c,\varphi\rangle}$. Since $W \in \mathcal{Y}_\sigma \subseteq \mathcal{X}_\sigma$, we get $W \in \mathcal{X}_\rho$.

Now take any $W' \in \mathcal{Y}_\rho$. Then W' is $\leq$-minimal in $\mathcal{X}_\rho$, so $W' \leq W$. But W is $\leq$-minimal in $\mathcal{X}_\sigma$, so $W' \in \mathcal{Y}_\rho \subseteq \mathcal{X}_\rho \subseteq \mathcal{X}_\sigma$ gives $W' \in \mathcal{Y}_\sigma$ also. Consequently, $\mathsf{E}_i\varphi \in B_c^\sigma$ means $W', c \models \mathsf{E}_i\varphi$. On the other hand, *Soundness* together with $\langle i, c, \varphi \rangle \in \rho$ and $W' \in \mathcal{X}_\rho$ means $W', c \models \mathsf{S}_i\varphi$. Hence $W', c \models \mathsf{E}_i\varphi \wedge \mathsf{S}_i\varphi$. From Proposition 8.2.1 part (4), we get $W', c \models \varphi$.

We have shown that φ holds in case c at an arbitrary world in $\mathcal{Y}_\rho$. Hence $\varphi \in B_c^\rho$, as required.

□

Similarly, we have sufficient conditioning for score-based operators to satisfy *Strong-cond-success*: the postulate follows if worlds in which i makes a expert, truthful report are strictly more plausible than worlds in which i makes a false report.

Lemma 8.5.4. *Suppose a score-based operator is such that for any $i \in \mathcal{S}$, $c \in \mathcal{C}$, $\varphi \in \mathcal{L}_0$ and $W, W' \in \mathcal{W}$,*

$$\begin{aligned} &W, c \models \mathsf{E}_i\varphi \wedge \varphi \text{ and } W', c \models \neg\varphi \\ &\qquad \implies d(W, \langle i, c, \varphi \rangle) < d(W', \langle i, c, \varphi \rangle) \end{aligned}$$

Then Strong-cond-success *holds.*

Proof. Suppose a score-based operator has the stated property. Take σ such that $\neg(\mathsf{E}_i\varphi \wedge \varphi) \notin B_c^\sigma$. Write $\rho = \sigma \cdot \langle i, c, \varphi\rangle$. We need to show that $\varphi \in B_c^\rho$.

First note that by $\neg(\mathsf{E}_i\varphi \wedge \varphi) \notin B_c^\sigma$ and the definition of B^σ for score-based operators, there is $W \in \mathcal{Y}_\sigma$ such that $W, c \models \mathsf{E}_i\varphi \wedge \varphi$.

Take any $W' \in \mathcal{Y}_\rho$. Suppose, for the sake of contradiction, that $W', c \not\models \varphi$. Then by the hypothesised property of the score function d, we have

$$d(W, \langle i, c, \varphi\rangle) < d(W', \langle i, c, \varphi\rangle)$$

Now, $W \in \mathcal{Y}_\sigma$ and $W' \in \mathcal{Y}_\rho \subseteq \mathcal{X}_\rho \subseteq \mathcal{X}_\sigma$ gives $r_\sigma(W) \leq r_\sigma(W')$. Thus

$$\begin{aligned} r_\rho(W) &= r_\sigma(W) + d(W, \langle i, c, \varphi\rangle) \\ &\leq r_\sigma(W') + d(W, \langle i, c, \varphi\rangle) \\ &< r_\sigma(W') + d(W', \langle i, c, \varphi\rangle) \\ &= r_\rho(W') < \infty \end{aligned}$$

i.e. $r_\rho(W) < r_\rho(W') < \infty$. But this means $W \in \mathcal{X}_\rho$ and W' is not minimal in $\mathcal{X}_\rho$ under r_ρ, contradicting $W' \in \mathcal{Y}_\rho$. Hence $W', c \models \varphi$.

Since W' was an arbitrary member of $\mathcal{Y}_\rho$, we have shown $\varphi \in B_c^\rho$, and thus *Strong-cond-success* is shown. □

The main result now follows.

Proof of Proposition 8.5.2. For the conditioning operators var-based-cond and part-based-cond, *Cond-success* follows from Lemma 8.5.3 since $W, c \models \varphi$ implies $W, c \models \mathsf{S}_i\varphi$. For the score-based operator excess-min, one can easily check that the condition in Lemma 8.5.4 holds, and thus *Strong-cond-success* and *Cond-success* follow. □

By omission, the reader may suppose that the conditioning operators fail *Strong-cond-success*. This is correct, and we can in fact say even more: *no* conditioning operator with a few basic properties – all of which are satisfied by var-based-cond and part-based-cond – can satisfy *Strong-cond-success*. In what follows, for a permutation $\pi : \mathcal{S} \to \mathcal{S}$ with $\pi(*) = *$, write $\pi(W)$ for the world with $v_c^{\pi(W)} = v_c^W$ and $\Pi_i^{\pi(W)} = \Pi_{\pi(i)}^W$. We have an impossibility result.

Proposition 8.5.3. *No elementary conditioning operator satisfying the basic postulates can simultaneously satisfy the following properties:*

1. $K^\emptyset = \mathrm{Cn}(\emptyset)$
2. *If π is a permutation of $\mathcal{S}$ with $\pi(*) = *$, $W \simeq \pi(W)$*
3. Refinement
4. Strong-cond-success

However, any proper subset of (1) - (4) is satisfiable.

(1) says that before any reports are received, we only know tautologies. As remarked earlier, this is not an *essential* property, but is reasonable when no prior knowledge is available. (2) is an anonymity postulate: it says that permuting the "names" of sources does not affect the plausibility of a world, and is a desirable property in light of (1). *Refinement*, introduced in Section 8.4.1, says that worlds in which all sources have more expertise are preferred.

Proof. Take distinct sources $i_1, i_2 \in \mathcal{S} \setminus \{*\}$, distinct cases $c, d \in \mathcal{C}$, and distinct valuations $v_1, v_2 \in \mathcal{V}$. Let $\varphi_1, \varphi_2 \in \mathcal{L}_0$ be propositional formulas with $\mathrm{mod}_0(\varphi_k) = v_k$ ($k \in \{1, 2\}$). Suppose for contradiction that some elementary conditioning operator – satisfying the basic postulates – has the stated properties.

Define a sequence

$$\sigma = (\langle *, c, \varphi_1 \vee \varphi_2 \rangle, \langle i_1, c, \varphi_1 \rangle, \langle i_2, c, \varphi_2 \rangle).$$

Let $\Pi_\perp$ denote the unit partition $\{\{u\} \mid u \in \mathcal{V}\}$, and let $\widehat{\Pi}$ denote the partition

$$\{\{v_1, v_2\}\} \cup \{\{u\} \mid u \in \mathcal{V} \setminus \{v_1, v_2\}\},$$

i.e. the partition obtained from $\Pi_\perp$ by merging the cells of v_1 and v_2.

Consider worlds W_1, W_2 given by

$$v_{c'}^{W_k} = v_k \qquad (c' \in \mathcal{C})$$

$$\Pi_i^{W_k} = \begin{cases} \widehat{\Pi}, & (k = 1 \text{ and } i = i_2) \text{ or } (k = 2 \text{ and } i = i_1) \\ \Pi_\perp, & \text{otherwise} \end{cases}$$

That is, W_1 has v_1 as its valuation for all cases, i_2 has partition $\widehat{\Pi}$, and all other sources have the finest partition $\Pi_\perp$; similarly W_2 has v_2 for its valuations and all sources except i_1 have $\Pi_\perp$.

Let $\leq$ denote the total preorder associated with the conditioning operator.

Claim 8.5.3. *$W_1 \simeq W_2$.*

Proof. Let π be the permutation of $\mathcal{S}$ which swaps i_1 and i_2. It is easily observed that $\pi(W_1)$ is partition-equivalent to W_2. By reflexivity of partition refinement, $\pi(W_1) \preceq W_2$ and $W_2 \preceq \pi(W_1)$. By *Refinement*, we get $\pi(W_1) \simeq W_2$. By property (2), $W_1 \simeq \pi(W_1)$. By transitivity of $\simeq$ we get $W_1 \simeq W_2$ as desired. □

Now, from the basic postulates, property (1) and Proposition 8.3.1 we have $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$. By elementariness and Lemma 8.4.1, we get $\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma) = \mathrm{mod}(G^\sigma_{\mathsf{snd}})$. It is easily checked that both W_1 and W_2 satisfy the soundness statements corresponding to σ, and thus $W_1, W_2 \in \mathrm{mod}(G^\sigma_{\mathsf{snd}}) = \mathcal{X}_\sigma$.

Claim 8.5.4. $W_1, W_2 \in \mathcal{Y}_\sigma$.

Proof. We show W_1 and W_2 are $\leq$-minimal in $\mathcal{X}_\sigma$. Take any $W \in \mathcal{X}_\sigma$. Then $W \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$, so $W, c \models \mathsf{S}_*(\varphi_1 \vee \varphi_2)$, i.e. $V^W_c \in \{v_1, v_2\}$. We consider two cases.

- **Case 1** ($v^W_c = v_1$). By $W \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$ again we have $W, c \models \mathsf{S}_{i_2}\varphi_2$, i.e

$$v_1 = v^W_c \in \Pi^W_{i_2}[\varphi_2] = \Pi^W_{i_2}[v_2].$$

  It follows that $\{v_1, v_2\} \subseteq \Pi^W_{i_2}[v_2]$, and that $\widehat{\Pi}$ refines $\Pi^W_{i_2}$. Since $\widehat{\Pi}$ is the partition of i_2 in W_1, and all other sources have the finest partition $\Pi_\perp$, we get $W_1 \preceq W$. By *Refinement*, $W_1 \leq W$. Since $W_1 \simeq W_2$ we have $W_2 \leq W$ also.

- **Case 2** ($v^W_c = v_2$). Applying a near-identical argument to that used in case 1 with soundness of the report $\langle i_1, c, \varphi_1 \rangle$, we get $W_1, W_2 \leq W$.

In either case, both $W_1 \leq W$ and $W_2 \leq W$, so $W_1, W_2 \in \mathcal{Y}_\sigma$. □

Now we consider case d. Since

$$W_1, d \models \mathsf{E}_{i_1}\varphi_1 \wedge \varphi_1$$

and $W_1 \in \mathcal{Y}_\sigma$, $\neg(\mathsf{E}_{i_1}\varphi_1 \wedge \varphi_1) \notin B^\sigma_d$. Writing $\rho = \sigma \cdot \langle i_1, d, \varphi_1 \rangle$, we get from *Strong-cond-success* that $\varphi_1 \in B^\rho_d$.

Note that $W_2, d \models \mathsf{S}_{i_1}\varphi_1$, so $W_2 \in \mathrm{mod}(G^\rho_{\mathsf{snd}}) = \mathrm{mod}(K^\rho) = \mathcal{X}_\rho$. Since W_2 is $\leq$-minimal in $\mathcal{X}_\sigma$ and

$$X_\rho = \mathrm{mod}(G^\rho_{\mathsf{snd}}) \subseteq \mathrm{mod}(G^\sigma_{\mathsf{snd}}) = \mathcal{X}_\sigma,$$

W_2 is also $\leq$-minimal in $\mathcal{X}_\rho$, i.e. $W_2 \in \mathcal{Y}_\rho$. Now $\varphi_1 \in B^\rho_d$ gives $W_2, d \models \varphi_1$. Since $v^{W_2}_d = v_2$ and $\mathrm{mod}_0(\varphi_1) = \{v_1\}$, this means $v_1 = v_2$. But v_1 and v_2 were assumed to be distinct: contradiction.

[TODO: Show that any strict subset is satisfiable] □

Proposition 8.5.3 highlights an important difference between conditioning and score-based operators, and hints that a fixed plausibility order may be too restrictive: we need to allow the order to be responsive to new reports in order to satisfy properties such as *Strong-cond-success*.

8.6 Selective Change

In the previous section we saw how a single formula φ may be accepted when it is received as an additional report. But what can we say about propositional beliefs when taking into account the *whole sequence* σ? To investigate this we introduce an analogue of *selective revision* [41], in which propositional beliefs are formed by "selecting" part of each input report (intuitively, some part consistent with the source's expertise). In what follows, write $\sigma \upharpoonright c = \{\langle i, \varphi\rangle \mid \langle i, c, \varphi\rangle \in \sigma\}$ for the c-reports in σ.

Definition 8.6.1. *A* selection scheme *is a mapping f assigning to each $*$-consistent sequence σ a function $f_\sigma : \mathcal{S} \times \mathcal{C} \times \mathcal{L}_0 \to \mathcal{L}_0$ such that $f_\sigma(i, c, \varphi) \in \mathrm{Cn}_0(\varphi)$. An operator is* selective *if there is a selection scheme f such that for all $*$-consistent σ and $c \in \mathcal{C}$,*

$$[B_c^\sigma] = \mathrm{Cn}_0(\{f_\sigma(i, c, \varphi) \mid \langle i, \varphi\rangle \in \sigma \upharpoonright c\}).$$

Thus, an operator is selective if its propositional beliefs in case c are formed by weakening each c-report and taking their consequences. Note that for $\sigma = \emptyset$ we get $[B_c^\sigma] = \mathrm{Cn}_0(\emptyset)$, so selectivity already rules out non-tautological prior propositional beliefs. Also note that in the presence of *Closure*, *Containment* and *Soundness*, selectivity implies that $[B_c^\sigma] = [B_c^\rho]$, where ρ is obtained by replacing each report $\langle i, c, \varphi\rangle$ with $\langle *, c, f_\sigma(i, c, \varphi)\rangle$.

Selectivity can be characterised by a natural postulate placing an upper bound on the propositional part of B_c^σ. In what follows, let $\Gamma_c^\sigma = \{\varphi \in \mathcal{L}_0 \mid \exists i \in \mathcal{S} : \langle i, \varphi\rangle \in \sigma \upharpoonright c\}$.

Boundedness *If σ is $*$-consistent, $[B_c^\sigma] \subseteq \mathrm{Cn}_0(\Gamma_c^\sigma)$*

Boundedness says that the propositional beliefs in case c should not go beyond the consequences of the formulas reported in case c. In some sense this can be seen as an iterated version of *Inclusion* from AGM revision, in the case where $[B_c^\sigma] = \mathrm{Cn}_0(\emptyset)$. We have the following characterisation.

Theorem 8.6.1. *A model-based operator is selective if and only if it satisfies* Boundedness*.*

Proof. "if": Suppose a model-based operator satisfies *Boundedness*. Take any $*$-consistent σ. For $c \in \mathcal{C}$, set

$$M_c = \mathrm{mod}_0\,[B_c^\sigma]\,.$$

By *Boundedness*, we have $M_c \supseteq \mathrm{mod}_0(\Gamma_c^\sigma)$. Now set

$$F_\sigma(i, c, \varphi) = \mathrm{mod}_0(\varphi) \cup M_c.$$

Define a selection function f_σ by letting $f_\sigma(i, c, \varphi)$ be any formula with $\mathrm{mod}_0(f_\sigma(i, c, \varphi)) = F_\sigma(i, c, \varphi)$. Since $F_\sigma(i, c, \varphi)$ contains the models of φ, clearly $f_\sigma(i, c, \varphi) \in \mathrm{Cn}_0(\varphi)$. Therefore f is indeed a selection function.

We claim that, for any $c \in \mathcal{C}$,

$$M_c = \bigcap_{\langle i, \varphi\rangle \in \sigma \upharpoonright c} F_\sigma(i, c, \varphi).$$

The "$\subseteq$" inclusion is clear since, by definition, $F_\sigma(i,c,\varphi) \supseteq M_c$. For the "$\supseteq$" inclusion, suppose for contradiction that there is some $v \in \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i,c,\varphi)$ with $v \notin M_c$.

Take any $\varphi \in \Gamma_c^\sigma$. Then there is $i \in \mathcal{S}$ such that $\langle i,\varphi\rangle \in \sigma \restriction c$, and hence $v \in F_\sigma(i,c,\varphi)$. But $v \notin M_c$ by assumption, so $v \in \mathrm{mod}_0(\varphi)$. This shows $v \in \mathrm{mod}_0(\Gamma_c^\sigma)$. But $\mathrm{mod}_0(\Gamma_c^\sigma) \subseteq M_c$ by *Boundedness*, so $v \in M_c$; contradiction.

From this we get

$$\begin{aligned}\mathrm{mod}_0\,[B_c^\sigma] &= M_c \\ &= \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i,c,\varphi) \\ &= \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} \mathrm{mod}_0(f_\sigma(i,c,\varphi)) \\ &= \mathrm{mod}_0\left(\{f_\sigma(i,c,\varphi) \mid \langle i,\varphi\rangle \in \sigma \restriction c\}\right)\end{aligned}$$

Since $[B_c^\sigma]$ is deductively closed (by *Closure*, which holds for all model-based operators), we get

$$[B_c^\sigma] = \mathrm{Cn}_0\left(\{f_\sigma(i,c,\varphi) \mid \langle i,\varphi\rangle \in \sigma \restriction c\}\right)$$

as required for selectivity.

"only if": Suppose a model-based operator is selective according to some selection scheme f. Take any $*$-consistent σ and $c \in \mathcal{C}$. Write

$$\Delta = \{f_\sigma(i,c,\varphi) \mid \langle i,\varphi\rangle \in \sigma \restriction c\}.$$

so that $[B_c^\sigma] = \mathrm{Cn}_0(\Delta)$. For $\langle i,\varphi\rangle \in \sigma \restriction c$ we have $f_\sigma(i,c,\varphi) \in \mathrm{Cn}_0(\varphi) \subseteq \mathrm{Cn}_0(\Gamma_c^\sigma)$ from the definition of a selection scheme and the fact that $\varphi \in \Gamma_c^\sigma$. Hence $\Delta \subseteq \mathrm{Cn}_0(\Gamma_c^\sigma)$, so

$$[B_c^\sigma] = \mathrm{Cn}_0(\Delta) \subseteq \mathrm{Cn}_0(\mathrm{Cn}_0(\Gamma_c^\sigma)) = \mathrm{Cn}_0(\Gamma_c^\sigma)$$

as required for *Boundedness*. □

[TODO: flow.]

This characterisation in Theorem 8.6.1 allows us to easily analyse when conditioning and score-based operators are selective. In the case of conditioning operators with $K^\emptyset = \mathrm{Cn}(\emptyset)$, we in fact have a precise characterisation. First, some terminology: say that a world W *refines* W' *at* c if for all $i \in \mathcal{S}$ we have $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. Intuitively, this means each source is more knowledgable in case c in world W than they are in W'. Write $\mathcal{W}_{c\,:\,v} = \{W \in \mathcal{W} \mid v_c^W = v\}$ for the set of worlds whose c valuation is v. We have the following.

Proposition 8.6.1. *Suppose an elementary conditioning operator satisfies the basic postulates and has $K^\emptyset = \mathrm{Cn}(\emptyset)$. Then it is selective if and only if for all W, c, v there is $W' \in \mathcal{W}_{c\,:\,v}$ such that $W' \leq W$ and W refines W' at all cases $d \neq c$.*

While the condition on $\leq$ in Proposition 8.6.1 is somewhat technical, it is implied by the very natural *partition-equivalence* property from Section 8.2. Consequently, var-based-cond and part-based-cond are selective. For the score-based operator excess-min, one can show *Boundedness* holds directly using a property of the disagreement scoring function d similar to the property of $\leq$ above. Consequently, excess-min is also selective.

We first state some preliminary results.

Lemma 8.6.1. *Suppose W refines W' at c. Then for any $i \in \mathcal{S}$ and $\varphi \in \mathcal{L}_0$,*

$$W, c \models \mathsf{S}_i\varphi \implies W', c \models \mathsf{S}_i\varphi$$

Proof. Suppose $W, c \models \mathsf{S}_i\varphi$. Then $v_c^W \in \Pi_i^W[\varphi]$, i.e. $\mathrm{mod}_0(\varphi) \cap \Pi_i^W[v_c^W] \neq \emptyset$. By refinement, $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. Hence $\mathrm{mod}_0(\varphi) \cap \Pi_i^{W'}[v_c^{W'}] \neq \emptyset$, so $v_c^{W'} \in \Pi_i^{W'}[\varphi]$. That is, $W', c \models \mathsf{S}_i\varphi$. □

Lemma 8.6.2. *For any $W \in \mathcal{W}$ and $c \in \mathcal{C}$, there is a $*$-consistent sequence σ – containing only reports for case c – such that for all $W' \in \mathcal{W}$,*

$$W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}}) \iff W \textit{ refines } W' \textit{ at } c.$$

Proof. For a valuation $v \in \mathcal{V}$, let $\varphi(v)$ be a propositional formula such that $\mathrm{mod}_0(\varphi(v)) = \{v\}$. Take σ to be any enumeration of reports of the form

$$\langle i, c, \varphi(v) \rangle$$

where $i \in \mathcal{S}$ and $v \in \Pi_i^W[v_c^W]$. Note that such a sequence exists since there are only finitely many sources and valuations. Clearly σ contains only c-reports. Since Π_*^W is the unit partition, the only report from $*$ is $\langle *, c, \varphi(v_c^W) \rangle$. Hence σ is $*$-consistent. We show the desired equivalence.

$\implies$: Suppose $W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$. Take any $i \in \mathcal{S}$. We need to show $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. Take $v \in \Pi_i^W[v_c^W]$. By construction of σ, $\langle i, c, \varphi(v) \rangle \in \sigma$. Hence $W', c \models \mathsf{S}_i\varphi(v)$, i.e. $v_c^{W'} \in \Pi_i^{W'}[\varphi(v)] = \Pi_i^{W'}[v]$. This shows $v \in \Pi_i^{W'}[v_c^{W'}]$ as required.

$\impliedby$: Suppose W refines W' at c. Take any $\langle i, c, \varphi(v) \rangle \in \sigma$. Then $v \in \Pi_i^W[v_c^W]$, so $v_c^W \in \Pi_i^W[v] = \Pi_i^W[\varphi(v)]$. This shows $W, c \models \mathsf{S}_i\varphi(v)$, and Lemma 8.6.1 gives $W', c \models \mathsf{S}_i\varphi(v)$. Hence $W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$.

□

Proof of Proposition 8.6.1. Take an elementary conditioning operator with the basic postulates and $K^\emptyset = \mathrm{Cn}(\emptyset)$.

"if": Suppose the stated property holds. Since all conditioning operators are model-based, by Theorem 8.6.1 it suffices to show *Boundedness*. To that end, let σ be $*$-consistent and take $c \in \mathcal{C}$. We need $[B_c^\sigma] \subseteq \mathrm{Cn}_0(\Gamma_c^\sigma)$; or equivalently, by *Closure*, $\mathrm{mod}_0\,[B_c^\sigma] \supseteq \mathrm{mod}_0(\Gamma_c^\sigma)$.

Take any $v \in \mathrm{mod}_0(\Gamma_c^\sigma)$. Since σ is $*$-consistent, B^σ is consistent by *Consistency*. Hence $\mathcal{Y}_\sigma \neq \emptyset$. Take any $W \in \mathcal{Y}_\sigma$. By the property in the statement of the result, there is $W' \in \mathcal{W}_{c\,:\,v}$ such that $W' \leq W$ and W refines W' at all cases $d \neq c$.

We claim $W' \in \mathcal{X}_\sigma$. By Proposition 8.3.1, elementariness and Lemma 8.4.1, we have $\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma) = \mathrm{mod}(G^\sigma_{\mathsf{snd}})$. Take any $\langle i, d, \varphi \rangle \in \sigma$. We consider cases.

- **Case 1** ($d = c$). Here $\langle i, \varphi \rangle \in \sigma \upharpoonright c$, so $\varphi \in \Gamma_c^\sigma$. Hence $v \in \mathrm{mod}_0(\Gamma_c^\sigma) \subseteq \mathrm{mod}_0(\varphi)$. Since $W' \in \mathcal{W}_{c\,:\,v}$, v is the c-valuation of W'. Hence $W', c \models \varphi$, and $W', c \models \mathsf{S}_i\varphi$ follows.

- **Case 2** ($d \neq c$). By assumption, W refines W' at d. Since $W \in \mathcal{Y}_\sigma \subseteq \mathcal{X}_\sigma$, we have $W, d \models \mathsf{S}_i\varphi$. By Lemma 8.6.1, $W', d \models \mathsf{S}_i\varphi$ also.

We have shown $W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}}) = \mathcal{X}_\sigma$. Now recall that $W \in \mathcal{Y}_\sigma$ – so W is $\leq$-minimal in $\mathcal{X}_\sigma$ – and $W' \leq W$. Thus W' is also $\leq$-minimal in $\mathcal{X}_\sigma$, i.e. $W' \in \mathcal{Y}_\sigma$. Since $W' \in \mathcal{W}_{c\,:\,v}$ also, we have by Lemma 8.5.1 that $v \in \mathrm{mod}_0\,[B_c^\sigma]$, as required.

"only if": Suppose our operator is selective, i.e. satisfies *Boundedness*. To show the desired property holds, take any W, c and v. Enumerate $\mathcal{C} \setminus \{c\}$ as $\{d_1, \ldots, d_N\}$. By Lemma 8.6.2, for each $1 \leq n \leq N$ there is a $*$-consistent sequence σ_n such that

$$\mathrm{mod}(G^{\sigma_n}_{\mathsf{snd}}) = \{W' \in \mathcal{W} \mid W \text{ refines } W' \text{ at } d_n\}.$$

Now, let φ and ψ be formulas with $\mathrm{mod}_0(\varphi) = \{v\}$ and $\mathrm{mod}_0(\psi) = \{v_c^W\}$. Let ρ be the concatenation

$$\rho = \sigma_1 \cdots \sigma_n \cdot \langle *, c, \varphi \vee \psi \rangle.$$

Note that ρ is $*$-consistent, since each σ_n is (and only refers to case d_n). We may therefore apply *Boundedness* for case c. Taking models of both sides yields

$$\mathrm{mod}_0\,[B_c^\rho] \supseteq \mathrm{mod}_0(\Gamma_c^\rho) = \mathrm{mod}_0(\varphi \vee \psi) = \{v, v_c^W\}.$$

In particular, $v \in \mathrm{mod}_0\,[B_c^\rho]$. By Lemma 8.5.1, there is some $W' \in \mathcal{Y}_\rho \cap \mathcal{W}_{c\,:\,v}$.

We show W' has the required properties. First note that since W refines itself at each d_n, we have $W \in \mathrm{mod}(G^{\sigma_n}_{\mathsf{snd}})$. Clearly $W, c \models \psi$, so $W, c \models \mathsf{S}_*(\varphi \vee \psi)$ too. Thus $W \in \mathrm{mod}(G^\rho_{\mathsf{snd}}) = \mathcal{X}_\rho$ (using $K^\emptyset = \mathrm{Cn}(\emptyset)$). Since $W' \in \mathcal{Y}_\rho = \min_\leq \mathcal{X}_\rho$, we get $W' \leq W$ as required.

Next, take any case $d \neq c$. Then there is some n such that $d = d_n$. Since $W' \in \mathcal{Y}_\rho \subseteq \mathcal{X}_\rho = \mathrm{mod}(G^\rho_{\mathsf{snd}}) \subseteq \mathrm{mod}(G^{\sigma_n}_{\mathsf{snd}})$, we get that W refines W' at d. This completes the proof.

$\square$

8.6.1 Case Independence

In the definition of a selection scheme, we allow $f_\sigma(i, c, \varphi)$ to depend on the case c. If one views $f_\sigma(i, c, \varphi)$ as a weakening of φ which accounts for the lack of expertise of i, this is somewhat at odds with other aspects of the framework, where expertise is independent of case. For this reason it is natural to consider *case independent* selective schemes.

Definition 8.6.2. *A selection scheme f is* case independent *if $f_\sigma(i, c, \varphi) \equiv f_\sigma(i, d, \varphi)$ for all $*$-consistent σ and $i \in \mathcal{S}$, $c, d \in \mathcal{C}$ and $\varphi \in \mathcal{L}_0$.*

Say an operator is *case-independent-selective* if it is selective according to some case independent scheme. This stronger notion of selectivity can again be characterised by a postulate which bounds propositional beliefs. For any set of cases $H \subseteq \mathcal{C}$, sequence σ and $c \in \mathcal{C}$, write

$$\Gamma_c^{\sigma,H} = \{\varphi \in \mathcal{L}_0 \mid \exists i \in \mathcal{S} : \langle i, \varphi\rangle \in \sigma \restriction c \\ \text{and } \forall d \in H : \langle i, \varphi\rangle \notin \sigma \restriction d\}.$$

H-Boundedness *For any $*$-consistent σ, $H \subseteq \mathcal{C}$ and $c \in \mathcal{C}$,*

$$[B_c^\sigma] \subseteq \mathrm{Cn}_0\left(\Gamma_c^{\sigma,H} \cup \bigcup_{d \in H} [B_d^\sigma]\right)$$

Note that *Boundedness* is obtained as the special case where $H = \emptyset$. We illustrate with an example.

Example 8.6.1. *Consider case c in the following sequence:*

$$\sigma = (\langle i, c, p\rangle, \langle j, c, q\rangle, \langle j, d, q\rangle, \langle k, d, r\rangle)$$

Boundedness *requires that $[B_c^\sigma] \subseteq \mathrm{Cn}_0(\{p, q\})$. However, the instance of* H-Boundedness *with $H = \{d\}$ makes use of the fact that j reports q in both cases c and d, and requires $[B_c^\sigma] \subseteq \mathrm{Cn}_0(\{p\} \cup [B_d^\sigma])$. This also has an interesting implication for case d: if $\varphi \in [B_c^\sigma]$, then $p \to \varphi \in [B_d^\sigma]$. This follows since $\beta \in \mathrm{Cn}_0(\{\alpha\} \cup \Gamma)$ iff $\alpha \to \beta \in \mathrm{Cn}_0(\Gamma)$ for $\alpha, \beta \in \mathcal{L}_0$. Intuitively, this says that if p (from i) and q (from j) is enough to accept φ in case c, then φ is accepted in case d* if p is, *given that the report of q from j is repeated for d.*

The characterisation is as follows.

Theorem 8.6.2. *A model-based operator is case-independent-selective if and only if it satisfies* H-Boundedness*.*

Proof. "only if": Suppose a model-based operator is selective according to some case-independent scheme f. Take any $*$-consistent σ, $H \subseteq \mathcal{C}$ and $c \in \mathcal{C}$. For any case d, write $M_d = \mathrm{mod}_0\,[B_d^\sigma]$. Note that with c_0 an arbitrary fixed case, and writing $F_\sigma(i,\varphi) = \mathrm{mod}_0(f_\sigma(i,c_0,\varphi))$, we have by case-independent-selectivity that

$$M_d = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction d} F_\sigma(i,\varphi).$$

By closure, it is sufficient for *H-Boundedness* to show that

$$M_c \supseteq \mathrm{mod}_0(\Gamma_c^{\sigma,H}) \cap \bigcap_{d\in H} M_d. \tag{8.1}$$

Take any v in the set on the right-hand side. To show $v \in M_c$, take any $\langle i,\varphi\rangle \in \sigma \restriction c$. If $\varphi \in \Gamma_c^{\sigma,H}$, then clearly

$$\begin{aligned} v &\in \mathrm{mod}_0(\Gamma_c^{\sigma,H}) \\ &\subseteq \mathrm{mod}_0(\varphi) \\ &\subseteq \mathrm{mod}_0(f_\sigma(i,c,\varphi)) \\ &= F_\sigma(i,\varphi) \end{aligned}$$

(where we use $f_\sigma(i,c,\varphi) \in \mathrm{Cn}_0(\varphi)$). Otherwise, $\varphi \notin \Gamma_c^{\sigma,H}$. Since $\langle i,\varphi\rangle \in \sigma \restriction c$, this means there is $d \in H$ such that $\langle i,\varphi\rangle \in \sigma \restriction d$. Hence $v \in M_d$ gives $v \in F_\sigma(i,\varphi)$. This shows the inclusion in (8.1), and we are done.

"if": Suppose a model-based operator satisfies *H-Boundedness*. Let σ be a $*$-consistent sequence. As before, write M_c for $\mathrm{mod}_0\,[B_c^\sigma]$. For $i \in \mathcal{S}$ and $c \in \mathcal{C}$, write

$$\mathcal{C}(i,\varphi) = \{c \in \mathcal{C} \mid \langle i,\varphi\rangle \in \sigma \restriction c\},$$

and set

$$F_\sigma(i,\varphi) = \mathrm{mod}_0(\varphi) \cup \bigcup_{c\in\mathcal{C}(i,\varphi)} M_c.$$

Define f by letting $f_\sigma(i,c,\varphi)$ be any propositional formula with $\mathrm{mod}_0(f_\sigma(i,c,\varphi)) = F_\sigma(i,\varphi)$. Then f is a case-independent selection scheme. We show our operator is selective according to f; by closure of $[B_c^\sigma]$ for each c, it suffices to show

$$M_c = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i,\varphi).$$

Fix c. For the left-to-right inclusion, suppose $v \in M_c$. Take any $\langle i,\varphi\rangle \in \sigma \restriction c$. Then $c \in \mathcal{C}(i,\varphi)$, so $F_\sigma(i,\varphi) \supseteq M_c$ and thus $v \in F_\sigma(i,\varphi)$ as required.

For the right-to-left inclusion, suppose v lies in the intersection. Set

$$H = \{d \in \mathcal{C} \mid v \in M_d\}.$$

Apply *H-Boundedness* and taking the models of both sides, we obtain

$$M_c \supseteq \mathrm{mod}_0(\Gamma_c^{\sigma,H}) \cap \bigcap_{d\in H} M_d. \tag{8.2}$$

Clearly $v \in \bigcap_{d \in H} M_d$ by definition of H. Let $\varphi \in \Gamma_c^{\sigma,H}$. Then there is $i \in \mathcal{S}$ such that $\langle i, \varphi \rangle \in \sigma \restriction c$, and consequently $v \in F_\sigma(i, \varphi)$. We claim $v \in \mathrm{mod}_0(\varphi)$. If not, by definition of $F_\sigma(i, \varphi)$ we must have $v \in \bigcup_{d \in \mathcal{C}(i,\varphi)} M_d$, i.e. there is $d \in \mathcal{C}$ such that $\langle i, \varphi \rangle \in \sigma \restriction d$ and $v \in M_d$. On the one hand, $\varphi \in \Gamma_c^{\sigma,H}$ implies $d \notin H$. On the other, $v \in M_d$ gives $d \in H$ directly by the definition of H: contradiction. This shows $v \in \mathrm{mod}_0(\varphi)$. Since φ was arbitrary, we have $v \in \mathrm{mod}_0(\Gamma_c^{\sigma,H})$. By (8.2) we get $v \in M_c$, and the proof is complete.

□

The question of whether our concrete operators satisfy *H-Boundedness* (equivalently, whether they are case-independent-selective) is still open.

8.6.2 Expertise and Selectivity

In the existing literature on selective belief change (e.g. [41, 13]), the selection function typically acts as a means to separate out the part of new information on which the reporting sources is *credible*, or *trusted*. In our framework, this property of the selection scheme f can be captured as follows.

Definition 8.6.3. *A selection scheme f is* expertise-compatible *(EC) with an operator $\sigma \mapsto \langle B^\sigma, K^\sigma \rangle$ if for all $*$-consistent σ and $\langle i, c, \varphi \rangle \in \sigma$,*

$$\mathsf{E}_i f_\sigma(i, c, \varphi) \in B_c^\sigma$$

Say an operator is *EC-selective* if it is selective according to some expertise-compatible scheme. While EC-selectivity may appear natural on first glance, we argue that it can be overly restrictive when expertise is derived from the input sequence itself. For example, consider the sequence

$$\sigma = (\langle i, c, p \rangle, \langle j, c, p \rangle, \langle i, d, p \rangle, \langle i, d, \neg p \rangle)$$

By *Soundness* and *Closure*, we cannot have both $\mathsf{E}_i p$ and $\mathsf{E}_j p$ in B_c^σ. Ideas of symmetry ([**TODO:** check for typos: there is no symmetry in the sequence as written]) suggest that neither can we pick one of i or j over the other, so that in fact it is reasonable to have neither $\mathsf{E}_i p$ nor $\mathsf{E}_j p$ in B_c^σ. Consequently – assuming p is the only propositional variable – the only formulas weaker than p on which i and j are believed to have expertise are tautologies. Any EC scheme f must therefore have $f_\sigma(i, c, p) \equiv f_\sigma(j, c, p) \equiv \top$. Consequently, EC-selectivity would imply $[B_c^\sigma] = \mathrm{Cn}_0(\top)$. This is a very conservative stance: while there is total consensus for p in case c, p cannot be believed due to disagreement elsewhere. This also conflicts with the "optimistic" attitude described in Example 8.3.1. According to that view we should have $\mathsf{E}_i p \vee \mathsf{E}_j p \in B_c^\sigma$, but this implies $p \in B_c^\sigma$ by *Soundness*, *Containment* and *Closure*.

The core issue here is that for $\mathsf{E}_i\varphi$ to be believed, i needs to be trusted on φ in *every* maximally plausible world. If these worlds look very different – e.g. W_1 trusts i but not j, and vice versa in W_2 – then EC-selectivity requires reports to be significantly weakened before expertise is believed in all worlds.

This discussion also hints that propositional beliefs for EC-selective operators are fully determined by the expertise part of B^σ, together with the reports in σ. If the knowledge K^σ is formed according to the soundness constraints of σ (as in our examples so far), we in fact have a characterisation result to this effect. The following lemma, which may also be of independent interest, is required. For a set of worlds $S \subseteq \mathcal{W}$, write $\Pi_i^S = \bigvee_{W \in S} \Pi_i^W$ for the join of the i-partitions of worlds in S.

Lemma 8.6.3. *If a model-based operator is EC-selective and satisfies* Soundness, *then*

$$\mathrm{mod}_0\left[B_c^\sigma\right] = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$$

for all $*$-consistent σ and $c \in \mathcal{C}$.

Now write G_σ^{snd} for the collection with $(G_\sigma^{\mathsf{snd}})_c = \{\mathsf{S}_i\varphi \mid \langle i, \varphi\rangle \in \sigma \restriction c\}$. For any collection G, write $E\left(G\right)$ for the sub-collection of formulas of the form $\mathsf{E}_i\varphi$.

Theorem 8.6.3. *Suppose a model-based operator satisfies* Consistency *and* $K^\sigma = \mathrm{Cn}(G_\sigma^{\mathsf{snd}})$ *for all σ. Then it is EC-selective if and only if*

$$[B_c^\sigma] = [\mathrm{Cn}_c(K^\sigma \sqcup E\left(B^\sigma\right))]$$

for all $*$-consistent σ and $c \in \mathcal{C}$.

[TODO: Write up from online notes.]

Finally, note that Lemma 8.6.3 immediately implies selectivity with respect to any scheme f such that $\mathrm{mod}_0(f_\sigma(i, c, \varphi)) = \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. Since the right-hand side does not depend on the case c, we get the following corollary.

Corollary 8.6.1. *If a model-based operator is EC-selective and satisfies* Soundness, *then it is case-independent-selective.*

8.7 Related Work

In this section we discuss related work.

Belief Merging. In the framework of Konieczny and Pérez [64], a merging operator Δ maps a multiset of propositional formulas $\Phi = \{\varphi_1, \ldots, \varphi_n\}$ and an integrity constraint μ to a formula $\Delta_\mu(\Phi)$. Here φ_i represents the input from source i, and $\Delta_\mu(\Phi)$ represents the merged result. Various operators and postulates have been proposed in the literature; see [63] for a review.

This can be seen as the special case of our framework with a single case c: for Φ, μ we consider the sequence $\sigma_{\Phi,\mu}$ where $*$ reports μ and each source i reports φ_i. Any operator then gives rise to a merging operator $\Delta_\mu(\Phi) = \bigwedge \left[B_c^{\sigma_{\Phi,\mu}}\right]$. Note that our basic postulates imply $\Delta_\mu(\Phi) \vdash \mu$ – the *IC0* postulate of Konieczny and Pérez [64]. We leave it to future work to determine which other merging postulates hold.

We go beyond this setting by considering multiple cases and explicitly modelling expertise (and trust, via beliefs about expertise). While it may be possible to model expertise *implicitly* in belief merging (for example, say i is not trusted on ψ if $\Delta_\mu(\Phi) \nvdash \psi$ when $\varphi_i \vdash \psi$), bringing expertise to the object level allows us to express more complex beliefs about expertise, such as $\mathsf{E}_a x \vee \mathsf{E}_b x$ in Example 8.3.1. It also facilitates postulates which refer directly to expertise, such as the weakenings of *Success* in Section 8.5.

However, our problem is more specialised than merging, since we focus specifically on conflicting information due to lack of expertise. Belief merging may be applied more broadly to other types of *information fusion*, e.g. subjective beliefs or goals [49], where notions of objective expertise do not apply. While our framework *could* be applied in these settings, our postulates may no longer be desirable.

Epistemic Logic. Our notions of expertise and soundness are related to *S5 knowledge* from epistemic logic [96]. In such logics, an agent *knows* φ at a state x if φ holds at all states y “accessible" from x. Knowledge is thus determined by an *epistemic accessibility relation*, which describes the distinctions between states the agent can make. The logic of S5 arises when this relation is an equivalence relation (or equivalently, a partition).

Our previous work [87] – in which expertise and soundness were introduced in a modal logic framework – showed that “expertise models" are in 1-to-1 correspondence with S5 models, such that $E(\varphi)$ holds iff $A(\varphi \to K\varphi)$ holds in the S5 model, where A is the universal modality. By symmetry of expertise, we can also replace φ with its negation. Thus, expertise has a precise epistemic interpretation: it is the ability to *know whether* φ holds in *any possible state*. Similarly, $S(\varphi)$ translates to $\neg K \neg \varphi$. That is, φ is sound exactly when the source does not *know* φ is false.

In the present framework, if we set $W, c \models K_i(\varphi)$ iff $\Pi_i[v_c] \subseteq \mathrm{mod}_0(\varphi)$ and $W, c \models A\Phi$ iff $\forall v : W_{c=v}, c \models \Phi$, where $W_{c=v}$ is the world obtained from W by setting $v'_c = v$, then we have $\mathsf{E}_i\varphi \equiv A(\varphi \to K_i(\varphi))$ and $\mathsf{S}_i\varphi \equiv \neg K_i(\neg\varphi)$. While K_i is not quite an S5 modality (the **5** axiom requires iterating K_i, which is not possible in our framework), this shows the fundamental link between expertise, soundness and knowledge.

8.8 Conclusion

Summary. In this chapter we studied a belief change problem – extending the classical AGM framework – in which beliefs about the state of the world in multiple cases, as well as expertise of multiple sources, must be inferred from a sequence of reports. This allowed us to take a fresh look at the interaction between trust (seen as *belief in expertise*) and belief. By inferring the expertise of the sources from the reports, we have generalised some earlier approaches to non-prioritised revision which assume expertise (or reliability, credibility, priority etc) is known up-front (e.g. [41, 53, 13, 25]). We went on to propose some concrete belief change operators, and explored their properties through examples and postulates.

We saw that conditioning operators satisfy some desirable properties, and our concrete instances make useful inferences that go beyond weak-mb. However, we have examples in which intuitively plausible inferences are blocked, and conditioning is largely incompatible with *Strong-cond-success*. Score-based operators, and in particular excess-min, offer a way around these limitations, but may come at the expense of some other seemingly reasonable postulates, such as *Duplicate-removal*.

Future Work. There are many possibilities for future work. Firstly, we have a representation result only for conditioning operators. A characterisation of score-based operators – either the class in general or the specific operator excess-min – remains to be found. This would help to further clarify the differences between conditioning and score-based operators. We have also not considered any computational issues. Determining the complexity of calculating the results of our example operators, and the complexity for conditioning and score-based operators more broadly, is left to future work. Secondly, there is scope for deeper postulate-based analysis. For example, there should be postulates governing how beliefs change in case c in response to reports in case d. We could also consider more postulates relating trust and belief, and compare these postulates with those of Yasser and Ismail [107]. Moreover, there are many weaker version of *Success* which have been considered in the literature (e.g. in [41, 53, 13]); we should compare these against our *Cond-success* and *Strong-cond-success* in future work.

Finally, our framework only deals with three levels of trust on a proposition: we can believe $\mathsf{E}_i\varphi$, believe $\neg\mathsf{E}_i\varphi$, or neither. Future work could investigate how to extend our semantics to talk about *graded expertise*, and thereby permit more fine-grained *degrees of trust* [57, 107, 25].

9 Truth-Tracking

9.1 Introduction

Part IV

Conclusions

10 Conclusion

10.1 Summary

10.2 Future Work

Bibliography

[1] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. "On the logic of theory change: Partial meet contraction and revision functions". In: *Journal of symbolic logic* (1985), pp. 510–530 (cited on pages 136, 161).

[2] Alon Altman and Moshe Tennenholtz. "Axiomatic Foundations for Ranking Systems". In: *J. Artif. Int. Res.* 31.1 (Mar. 2008), pp. 473–495. ISSN: 1076-9757. URL: http://dl.acm.org/citation.cfm?id=1622655.1622669 (cited on pages 8, 15, 16, 43–45).

[3] Alon Altman and Moshe Tennenholtz. "Ranking systems: the PageRank axioms". In: *Proceedings of the 6th ACM conference on Electronic commerce*. ACM. 2005, pp. 1–8 (cited on pages 8, 45).

[4] Milica Aneli et al. "Some new considerations about double nested graphs". In: *Linear Algebra and its Applications* 483 (2015), pp. 323–341 (cited on page 93).

[5] Reid Andersen et al. "Trust-based Recommendation Systems: An Axiomatic Approach". In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08. New York, NY, USA: ACM, 2008, pp. 199–208. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367525. URL: http://doi.acm.org/10.1145/1367497.1367525 (cited on page 8).

[6] Kenneth J. Arrow. "Social Choice and Individual Values". In: *Ethics* 62.3 (1952), pp. 220–222 (cited on page 15).

[7] Sheldon Axler. *Linear Algebra Done Right*. Springer International Publishing, Nov. 2014. DOI: 10.1007/978-3-319-11080-6. URL: https://doi.org/10.1007/978-3-319-11080-6 (cited on page 34).

[8] Raju Balakrishnan and Subbarao Kambhampati. "Sourcerank: Relevance and trust assessment for deep web sources based on inter-source agreement". In: *Proceedings of the 20th international conference on World wide web*. 2011, pp. 227–236 (cited on page 18).

[9] Alexandru Baltag and Sonja Smets. "A qualitative theory of dynamic interactive belief revision". In: *Logic and the foundations of game and decision theory (LOFT 7)* 3 (2008), pp. 9–58 (cited on page 130).

[10] Johan van Benthem and Guram Bezhanishvili. "Modal logics of space". In: *Handbook of spatial logics*. Springer, 2007, pp. 217–298 (cited on pages 107, 119).

[11] Laure Berti-Equille and Javier Borge-Holthoefer. "Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics". In: *Synthesis Lectures on Data Management* 7.3 (2015), pp. 1–155 (cited on page 7).

[12] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*. Vol. 53. Cambridge University Press, 2002 (cited on page 119).

[13] Richard Booth and Aaron Hunter. "Trust as a Precursor to Belief Revision". In: *JAIR* 61 (2018), pp. 699–722 (cited on pages 135–137, 140, 175, 178).

[14] Denis Bouyssou. "Monotonicity of ranking by choosing: A progress report". In: *Social Choice and Welfare* 23.2 (2004), pp. 249–273 (cited on pages 76, 81).

[15] Sándor Bozóki, László Csató, and József Temesi. "An application of incomplete pairwise comparison matrices for ranking top tennis players". In: *European Journal of Operational Research* 248.1 (2016), pp. 211–218 (cited on page 47).

[16] Ralph Allan Bradley and Milton E. Terry. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons". In: *Biometrika* 39.3/4 (1952), pp. 324–345. ISSN: 00063444 (cited on page 53).

[17] Felix Brandt et al. "Introduction to Computational Social Choice". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 1 (cited on pages 15, 17).

[18] Felix Brandt et al. "Tournament Solutions". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. Cambridge University Press, 2016, pp. 57–84 (cited on pages 47, 50, 51, 93).

[19] Michelene TH Chi, Robert Glaser, and Marshall J Farr. *The nature of expertise*. Psychology Press, 2014 (cited on page 98).

[20] Zoé Christoff and Davide Grossi. "Binary Voting with Delegable Proxy: An Analysis of Liquid Democracy". In: *Proc. TARK 2017*. 2017 (cited on pages 11, 19).

[21] Harry Collins and Robert Evans. *Rethinking expertise*. University of Chicago Press, 2008 (cited on page 98).

[22] David Cook II. "Nested colorings of graphs". In: *Australasian Journal of Combinatorics* 62.1 (2015), pp. 100–127 (cited on page 93).

[23] László Csató. “An impossibility theorem for paired comparisons”. In: *Central European Journal of Operations Research* 27.2 (2019), pp. 497–514 (cited on pages 48, 49).

[24] Mehdi Dastani et al. “Inferring trust”. In: *CLIMA*. Springer. 2004, pp. 144–160 (cited on page 135).

[25] James P Delgrande, Didier Dubois, and Jérôme Lang. “Iterated Revision as Prioritized Merging.” In: *KR* 6 (2006), pp. 210–220 (cited on pages 145, 178).

[26] James P Delgrande, Pavlos Peppas, and Stefan Woltran. “General belief revision”. In: *Journal of the ACM (JACM)* 65.5 (2018), pp. 1–34 (cited on pages 143, 151, 152).

[27] Paul Denny et al. “PeerWise: Students Sharing Their Multiple Choice Questions”. In: *Proceedings of the Fourth International Workshop on Computing Education Research*. ICER ’08. Sydney, Australia: ACM, 2008, pp. 51–58 (cited on page 47).

[28] Hu Ding, Jing Gao, and Jinhui Xu. “Finding global optimum for truth discovery: Entropy based geometric variance”. In: *Proc. 32nd International Symposium on Computational Geometry (SoCG 2016)*. 2016 (cited on page 12).

[29] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer Netherlands, 2008. DOI: 10.1007/978-1-4020-5839-4 (cited on page 131).

[30] Elad Dokow and Ron Holzman. “Aggregation of binary evaluations with abstentions”. In: *Journal of Economic Theory* 145 (2010), pp. 544–561 (cited on page 11).

[31] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. “Integrating conflicting data: the role of source dependence”. In: *Proceedings of the VLDB Endowment* 2.1 (2009), pp. 550–561 (cited on page 18).

[32] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. “Truth Discovery and Copying Detection in a Dynamic World”. In: *Proc. VLDB Endow.* 2.1 (Aug. 2009), pp. 562–573. ISSN: 2150-8097. DOI: 10.14778/1687627.1687691. URL: https://doi.org/10.14778/1687627.1687691 (cited on page 18).

[33] Pål Grønås Drange et al. “On the threshold of intractability”. In: *Algorithms-ESA 2015*. Springer, 2015, pp. 411–423 (cited on pages 52, 93).

[34] Yang Du et al. “Bayesian Co-Clustering Truth Discovery for Mobile Crowd Sensing Systems”. In: *IEEE Transactions on Industrial Informatics* (2019), pp. 1–1. ISSN: 1551-3203. DOI: 10.1109/TII.2019.2896287 (cited on pages 7, 14).

[35] Edith Elkind and Arkadii Slinko. “Rationalizations of voting rules”. In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. Cambridge University Press, 2016, pp. 169–196 (cited on pages 53, 93).

[36] Ulle Endriss. “Judgment Aggregation”. In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 17 (cited on pages 8, 11, 15).

[37] K Anders Ericsson and Tyler J Towne. “Expertise.” In: *WIREs Cognitive Science* (2010) (cited on page 98).

[38] Patricia Everaere, Sebastien Konieczny, and Pierre Marquis. “Belief Merging Operators as Maximum Likelihood Estimators”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. Ed. by Christian Bessiere. International Joint Conferences on Artificial Intelligence Organization, July 2020, pp. 1763–1769 (cited on page 53).

[39] Patricia Everaere, Sébastien Konieczny, Pierre Marquis, et al. “The Epistemic View of Belief Merging: Can We Track the Truth?.” In: *ECAI*. 2010, pp. 621–626 (cited on page 46).

[40] Ronald Fagin et al. *Reasoning about knowledge*. MIT press, 2003 (cited on pages 99, 123, 124, 126).

[41] Eduardo L Fermé and Sven Ove Hansson. “Selective revision”. In: *Studia Logica* 63.3 (1999), pp. 331–342 (cited on pages 136, 138, 168, 175, 178).

[42] Alban Galland et al. “Corroborating Information from Disagreeing Views”. In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. WSDM ’10. New York, New York, USA: ACM, 2010, pp. 131–140. ISBN: 978-1-60558-889-6. DOI: `10.1145/1718487.1718504`. URL: `http://doi.acm.org/10.1145/1718487.1718504` (cited on pages 8, 11, 14).

[43] Alban Galland et al. “Corroborating Information from Disagreeing Views”. In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. WSDM ’10. New York: ACM, 2010, pp. 131–140. ISBN: 978-1-60558-889-6 (cited on page 48).

[44] Ebrahim Ghorbani. *Some spectral properties of chain graphs*. 2017. arXiv: `1703.03581 [math.CO]` (cited on page 93).

[45] Arpita Ghosh, Satyen Kale, and Preston McAfee. “Who moderates the moderators?” In: *Proceedings of the 12th ACM conference on Electronic commerce - EC 11*. ACM Press, 2011. DOI: `10.1145/1993574.1993599`. URL: `https://doi.org/10.1145%2F1993574.1993599` (cited on page 8).

[46] Alvin I Goldman. “Expertise”. In: *Topoi* 37.1 (2018), pp. 3–10 (cited on page 98).

[47] Julio González-Daz, Ruud Hendrickx, and Edwin Lohmann. "Paired comparisons analysis: an axiomatic approach to ranking methods". In: *Social Choice and Welfare* 42.1 (2014), pp. 139–169 (cited on pages 47–49, 53, 59, 93).

[48] Valentin Goranko and Solomon Passy. "Using the Universal Modality: Gains and Questions". In: *Journal of Logic and Computation* 2.1 (1992), pp. 5–30 (cited on page 100).

[49] Eric Grégoire and Sébastien Konieczny. "Logic-based approaches to information fusion". In: *Information Fusion* 7.1 (2006). Logic-based Approaches to Information Fusion, pp. 4–18. ISSN: 1566-2535. DOI: `https://doi.org/10.1016/j.inffus.2005.08.001`. URL: `http://www.sciencedirect.com/science/article/pii/S1566253505000771` (cited on page 177).

[50] Adam Grove. "Two modellings for theory change". In: *Journal of philosophical logic* (1988), pp. 157–170 (cited on page 161).

[51] Manish Gupta and Jiawei Han. "Heterogeneous Network-based Trust Analysis: A Survey". In: *SIGKDD Explor. Newsl.* 13.1 (Aug. 2011), pp. 54–71. ISSN: 1931-0145. DOI: `10.1145/2031331.2031341`. URL: `http://doi.acm.org/10.1145/2031331.2031341` (cited on pages 7, 8, 10).

[52] Sven Ove Hansson. "A survey of non-prioritized belief revision". In: *Erkenntnis* 50.2 (1999), pp. 413–427 (cited on page 136).

[53] Sven Ove Hansson et al. "Credibility limited revision". In: *Journal of Symbolic Logic* 66.4 (Dec. 2001), pp. 1581–1596. DOI: `10.2307/2694963`. URL: `https://doi.org/10.2307%2F2694963` (cited on pages 136, 178).

[54] Stephan Hartmann and Jan Sprenger. "Judgment aggregation and the problem of tracking the truth". In: *Synthese* (2012), pp. 209–221 (cited on page 46).

[55] Andreas Herzig et al. "A logic of trust and reputation". In: *Logic Journal of the IGPL* 18.1 (2010), pp. 214–244 (cited on page 135).

[56] Aaron Hunter. "Building Trust for Belief Revision". In: *PRICAI 2021: Trends in Artificial Intelligence*. Ed. by Duc Nghia Pham et al. Cham: Springer International Publishing, 2021, pp. 543–555. ISBN: 978-3-030-89188-6 (cited on page 135).

[57] Aaron Hunter. "Building Trust for Belief Revision". In: *PRICAI 2021: Trends in Artificial Intelligence*. Ed. by Duc Nghia Pham et al. Cham: Springer International Publishing, 2021, pp. 543–555. ISBN: 978-3-030-89188-6 (cited on pages 138, 178).

[58] Yang Jiao, R Ravi, and Wolfgang Gatterbauer. "Algorithms for automatic ranking of participants and tasks in an anonymized contest". In: *International Workshop on Algorithms and Computation*. Springer. 2017, pp. 335–346 (cited on pages 47, 48, 51, 52, 58, 93, 94).

[59] Hirofumi Katsuno and Alberto O. Mendelzon. "Propositional knowledge base revision and minimal change". In: *Artificial Intelligence* 52.3 (Dec. 1991), pp. 263–294. DOI: 10.1016/0004-3702(91)90069-v. URL: https://doi.org/10.1016%2F0004-3702%2891%2990069-v (cited on page 161).

[60] Daniel Kilov. "The brittleness of expertise and why it matters". In: *Synthese* 199.1 (2021), pp. 3431–3455 (cited on page 98).

[61] Jon M. Kleinberg. "Authoritative Sources in a Hyperlinked Environment". In: *J. ACM* 46.5 (Sept. 1999), pp. 604–632. ISSN: 0004-5411. DOI: 10.1145/324133.324140. URL: http://doi.acm.org/10.1145/324133.324140 (cited on pages 13, 45).

[62] Sébastien Konieczny and Ramón Pino Pérez. "Improvement Operators". In: *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*. Ed. by Gerhard Brewka and Jérôme Lang. AAAI Press, 2008, pp. 177–187. URL: http://www.aaai.org/Library/KR/2008/kr08-018.php (cited on page 39).

[63] Sébastien Konieczny and Ramón Pino Pérez. "Logic Based Merging". In: *Journal of Philosophical Logic* 40.2 (Mar. 2011), pp. 239–270. DOI: 10.1007/s10992-011-9175-5. URL: https://doi.org/10.1007%2Fs10992-011-9175-5 (cited on page 177).

[64] Sébastien Konieczny and Ramón Pino Pérez. "Merging information under constraints: a logical framework". In: *Journal of Logic and computation* 12.5 (2002), pp. 773–808 (cited on pages 39, 136, 177).

[65] Neema Kotonya and Francesca Toni. "Explainable Automated Fact-Checking for Public Health Claims". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 7740–7754 (cited on page 8).

[66] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. "Nonmonotonic reasoning, preferential models and cumulative logics". In: *Artificial intelligence* 44.1-2 (1990), pp. 167–207 (cited on page 151).

[67] Justin Kruger et al. "Axiomatic Analysis of Aggregation Methods for Collective Annotation". In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. AAMAS '14. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 2014,

pp. 1185–1192. ISBN: 978-1-4503-2738-1. URL: http://dl.acm.org/citation.cfm?id=2617388.2617437 (cited on page 17).

[68] Jean-François Laslier and M Remzi Sanver. *Handbook on approval voting*. Springer Science & Business Media, 2010 (cited on page 43).

[69] Yaliang Li et al. "A Survey on Truth Discovery". In: *SIGKDD Explor. Newsl.* 17.2 (2016), pp. 1–16. ISSN: 1931-0145. DOI: 10.1145/2897350.2897352. URL: http://doi.acm.org/10.1145/2897350.2897352 (cited on pages 7, 47, 137).

[70] Yaliang Li et al. "Conflicts to Harmony: A Framework for Resolving Conflicts in Heterogeneous Data by Truth Discovery". In: *IEEE Transactions on Knowledge and Data Engineering* 28.8 (Aug. 2016), pp. 1986–1999. ISSN: 1041-4347. DOI: 10.1109/TKDE.2016.2559481 (cited on pages 8, 12, 14).

[71] Churn-Jung Liau. "Belief, information acquisition, and trust in multi-agent systems – A modal logic formulation". In: *Artificial Intelligence* 149.1 (2003), pp. 31–60 (cited on page 135).

[72] Emiliano Lorini, Guifei Jiang, and Laurent Perrussel. "Trust-Based Belief Change". In: *Proc. ECAI*. 2014, pp. 549–554 (cited on page 135).

[73] Fenglong Ma et al. "FaitCrowd: Fine Grained Truth Discovery for Crowdsourced Data Aggregation". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. event-place: Sydney, NSW, Australia. New York, NY, USA: ACM, 2015, pp. 745–754. ISBN: 978-1-4503-3664-2. DOI: 10.1145/2783258.2783314. URL: http://doi.acm.org/10.1145/2783258.2783314 (cited on page 7).

[74] Fenglong Ma et al. "Unsupervised Discovery of Drug Side-Effects from Heterogeneous Data Sources". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. Halifax, NS, Canada: ACM, 2017, pp. 967–976. ISBN: 978-1-4503-4887-4. DOI: 10.1145/3097983.3098129. URL: http://doi.acm.org/10.1145/3097983.3098129 (cited on page 7).

[75] Jermaine Marshall, Arturo Argueta, and Dong Wang. "A neural network approach for truth discovery in social sensing". In: *2017 IEEE 14th international conference on mobile Ad Hoc and sensor systems (MASS)*. IEEE. 2017, pp. 343–347 (cited on page 8).

[76] Reshef Meir et al. "Truth Discovery via Proxy Voting". In: *CoRR* abs/1905.00629 (2019). arXiv: 1905.00629. URL: http://arxiv.org/abs/1905.00629 (cited on page 46).

[77] Assaf Natanzon, Ron Shamir, and Roded Sharan. "A polynomial approximation algorithm for the minimum fill-in problem". In: *SIAM Journal on Computing* 30.4 (2000), pp. 1067–1079 (cited on page 94).

[78] Assaf Natanzon, Ron Shamir, and Roded Sharan. "Complexity classification of some edge modification problems". In: *Discrete Applied Mathematics* 113.1 (2001), pp. 109–128 (cited on page 93).

[79] Catherine Olsson et al. *Skill Rating for Generative Models*. 2018. arXiv: `1808.04888 [stat.ML]` (cited on page 48).

[80] Aybüke Özgün. "Evidence in Epistemic Logic : A Topological Perspective". Theses. Université de Lorraine, Oct. 2017 (cited on page 107).

[81] Eric Pacuit. *Neighborhood semantics for modal logic*. Springer International Publishing, 2017 (cited on page 101).

[82] Lawrence Page et al. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66. Stanford InfoLab, Nov. 1999. URL: `http://ilpubs.stanford.edu:8090/422/` (cited on page 45).

[83] Jeff Pasternack and Dan Roth. "Knowing What to Believe (when You Already Know Something)". In: *Proceedings of the 23rd International Conference on Computational Linguistics*. COLING '10. Beijing, China: Association for Computational Linguistics, 2010, pp. 877–885. URL: `http://dl.acm.org/citation.cfm?id=1873781.1873880` (cited on pages 8, 9, 11, 13, 15).

[84] Jan Plaza. "Logics of public communications". In: *Synthese* 158.2 (2007), pp. 165–179 (cited on page 131).

[85] Ariel Rubinstein. "Ranking the participants in a tournament". In: *SIAM Journal on Applied Mathematics* 38.1 (1980), pp. 108–111 (cited on pages 50, 58–60).

[86] Nicolas Schwind and Sébastien Konieczny. "Non-Prioritized Iterated Revision: Improvement via Incremental Belief Merging". In: *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning*. Sept. 2020, pp. 738–747. DOI: `10.24963/kr.2020/76`. URL: `https://doi.org/10.24963/kr.2020/76` (cited on page 145).

[87] Joseph Singleton. "A Logic of Expertise". In: *ESSLLI 2021 Student Session* (2021). URL: `https://arxiv.org/abs/2107.10832` (cited on pages vi, 98, 99, 135, 137, 177).

[88] Joseph Singleton and Richard Booth. "An Axiomatic Approach to Truth Discovery". In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS 20. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 2011–2013. ISBN: 9781450375184 (cited on pages vi, 47).

[89] Joseph Singleton and Richard Booth. "Rankings for Bipartite Tournaments via Chain Editing". In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '21. Virtual Event, United Kingdom: International Foundation for Autonomous Agents and Multiagent Systems, 2021, pp. 1236–1244. ISBN: 9781450383073 (cited on page vi).

[90] Joseph Singleton and Richard Booth. *Who's the Expert? On Multi-source Belief Change*. 2022. DOI: 10.48550/ARXIV.2205.00077. URL: https://arxiv.org/abs/2205.00077 (cited on page vi).

[91] Giora Slutzki and Oscar Volij. "Ranking participants in generalized tournaments". In: *International Journal of Game Theory* 33.2 (2005), pp. 255–270 (cited on pages 48, 49, 94).

[92] Giora Slutzki and Oscar Volij. "Scoring of web pages and tournamentsaxiomatizations". In: *Social Choice and Welfare* 26.1 (2006), pp. 75–92 (cited on page 47).

[93] Moshe Tennenholtz. "Reputation Systems: An Axiomatic Approach". In: *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*. UAI '04. Banff, Canada: AUAI Press, 2004, pp. 544–551. ISBN: 0-9749039-0-6. URL: http://dl.acm.org/citation.cfm?id=1036843.1036909 (cited on pages 8, 15, 16, 45).

[94] Johan van Benthem, David Fernández-Duque, and Eric Pacuit. "Evidence and plausibility in neighborhood structures". In: *Annals of Pure and Applied Logic* 165.1 (2014), pp. 106–133 (cited on page 135).

[95] Johan van Benthem and Eric Pacuit. "Dynamic logics of evidence-based beliefs". In: *Studia Logica* 99.1 (2011), pp. 61–92 (cited on page 135).

[96] Hans van Ditmarsch et al., eds. *Handbook of Epistemic Logic*. English. College Publications, 2015. ISBN: 978-1-84890-158-2 (cited on page 177).

[97] Dalia Attia Waguih and Laure Berti-Equille. "Truth discovery algorithms: An experimental evaluation". In: *arXiv preprint arXiv:1409.6428* (2014) (cited on page 8).

[98] Dong Wang et al. "On Truth Discovery in Social Sensing: A Maximum Likelihood Estimation Approach". In: *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*. IPSN '12. event-place: Beijing, China. New York, NY, USA: ACM, 2012, pp. 233–244. ISBN: 978-1-4503-1227-1. DOI: 10.1145/2185677.2185737. URL: http://doi.acm.org/10.1145/2185677.2185737 (cited on pages 7, 53).

[99] Yaqing Wang et al. "Eann: Event adversarial neural networks for multimodal fake news detection". In: *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. 2018, pp. 849–857 (cited on page 8).

[100] Tomasz Ws and Oskar Skibski. *Axiomatic Characterization of PageRank*. 2020. DOI: 10.48550/ARXIV.2010.08487. URL: https://arxiv.org/abs/2010.08487 (cited on page 45).

[101] Kyle Powys Whyte and Robert P Crease. "Trust, expertise, and the philosophy of science". In: *Synthese* 177.3 (2010), pp. 411–425 (cited on page 98).

[102] Houping Xiao. "Multi-sourced Information Trustworthiness Analysis: Applications and Theory". PhD thesis. University at Buffalo, State University of New York, 2018 (cited on page 8).

[103] Houping Xiao et al. "A Truth Discovery Approach with Theoretical Guarantee". In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 1925–1934. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939816. URL: http://doi.acm.org/10.1145/2939672.2939816 (cited on page 8).

[104] Yi Yang, Quan Bai, and Qing Liu. "A probabilistic model for truth discovery with object correlations". In: *Knowledge-Based Systems* 165 (2019), pp. 360–373. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2018.12.004. URL: http://www.sciencedirect.com/science/article/pii/S0950705118305914 (cited on pages 14, 46).

[105] Yi Yang, Quan Bai, and Qing Liu. "On the Discovery of Continuous Truth: A Semi-supervised Approach with Partial Ground Truths". In: *Web Information Systems Engineering WISE 2018*. Springer International Publishing, 2018, pp. 424–438. DOI: 10.1007/978-3-030-02922-7_29. URL: https://doi.org/10.1007%2F978-3-030-02922-7_29 (cited on page 12).

[106] Mihalis Yannakakis. "Computing the minimum fill-in is NP-complete". In: *SIAM Journal on Algebraic Discrete Methods* 2.1 (1981), pp. 77–79 (cited on pages 50, 93).

[107] Ammar Yasser and Haythem O. Ismail. "Trust Is All You Need: From Belief Revision to Information Revision". In: *Logics in Artificial Intelligence*. Ed. by Wolfgang Faber et al. Cham: Springer International Publishing, 2021, pp. 50–65. ISBN: 978-3-030-75775-5 (cited on pages 137, 178).

[108] Xiaoxin Yin, Jiawei Han, and Philip S. Yu. "Truth Discovery with Multiple Conflicting Information Providers on the Web". In: *IEEE Transactions on Knowledge and Data Engineering* 20.6 (June 2008), pp. 796–808. ISSN: 1041-4347. DOI: 10.1109/TKDE.2007.190745 (cited on pages 8, 10, 11, 15).

[109] Xiaoxin Yin and Wenzhao Tan. "Semi-supervised Truth Discovery". In: *Proceedings of the 20th International Conference on World Wide Web*. WWW '11. event-place: Hyderabad, India. New York, NY, USA: ACM, 2011, pp. 217–226. ISBN: 978-1-4503-0632-4. DOI: 10.1145/1963405.1963439. URL: http://doi.acm.org/10.1145/1963405.1963439 (cited on page 8).

[110] Daniel Yue Zhang et al. "On robust truth discovery in sparse social media sensing". In: *2016 IEEE International Conference on Big Data (Big Data)*. 2016-12, pp. 1076–1081. DOI: 10.1109/BigData.2016.7840710 (cited on pages 7, 11).

[111] Liyan Zhang et al. "Latent Dirichlet Truth Discovery: Separating Trustworthy and Untrustworthy Components in Data Sources". In: *IEEE Access* 6 (2018), pp. 1741–1752. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2780182 (cited on pages 11, 14, 15).

[112] Shi Zhi et al. "Modeling Truth Existence in Truth Discovery". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. Sydney, NSW, Australia: ACM, 2015, pp. 1543–1552. ISBN: 978-1-4503-3664-2. DOI: 10.1145/2783258.2783339. URL: http://doi.acm.org/10.1145/2783258.2783339 (cited on pages 11, 14).

[113] William S. Zwicker. "Introduction to the Theory of Voting". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 2 (cited on pages 8, 17).

A Proofs for Chapter 4

A.1 Proof of Theorem 4.4.1

The following lemma is required before the proof.

Lemma A.1.1. *Suppose a network $N = (V, E)$ contains claims only for a single object $o \in \mathcal{O}$; that is, there exists $o \in \mathcal{O}$ such that $(s, f) \in E$ implies $obj_N(f) = o$ for all $s \in \mathcal{S}, f \in \mathcal{F}$. Then for any Symmetric operator T and $f_1, f_2 \in \mathcal{F}$, $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)| > 0$ implies $f_1 \approx_N^T f_2$.*

Proof. Suppose N has the stated property, T satisfies symmetry, and $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)| > 0$. Then there is a bijection $\varphi : \mathsf{src}_N(f_1) \to \mathsf{src}_N(f_2)$. Note that since f_1 and f_2 are for the same object no source can claim both facts, i.e. $\mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2) = \emptyset$.

Define a permutation π by

$$\pi(s) = \begin{cases} \varphi(s) & \text{if } s \in \mathsf{src}_N(f_1) \\ \varphi^{-1}(s) & \text{if } s \in \mathsf{src}_N(f_2) \\ s & \text{otherwise} \end{cases}$$

$$\pi(f) = \begin{cases} f_2 & \text{if } f = f_1 \\ f_2 & \text{if } f = f_2 \\ f & \text{otherwise} \end{cases}$$

and $\pi(o) = o$ for all $o \in \mathcal{O}$. That is, π swaps facts f_1 and f_2, and swaps the sources of f_1 with their counterparts in f_2. Note that $\pi = \pi^{-1}$.

Write $N' = \pi(N)$. We claim that $N' = N$. Write E, E' for the edges in N and N' respectively. First we will show $E \subseteq E'$. Suppose $(s, f) \in E$. There are three cases.

Case 1: $f = f_1$. Here we have $(s, f_1) \in E$, so $s \in \mathsf{src}_N(f_1)$. Consequently $\pi(s) = \varphi(s) \in \mathsf{src}_N(f_2)$, i.e. $(\pi(s), f_2) \in E$. By the definition of a graph isomorphism we get $(\pi(\pi(s)), \pi(f_2)) \in E'$. Noting that $\pi(f_2) = f_1 = f$ and $\pi(\pi(s)) = s$ (since $\pi = \pi^{-1}$), we have $(s, f) \in E'$ as desired.

Case 2: $f = f_2$. Similar to the above case, here we have $s \in \mathsf{src}_N(f_2)$ and so $\pi(s) = \varphi^{-1}(s) \in \mathsf{src}_N(f_1)$, i.e. $(\pi(s), f_1) \in E$. As before, applying the definition of a graph isomorphism and using $\pi = \pi^{-1}$, we get $(s, f) \in E'$.

Case 3: $f \notin \{f_1, f_2\}$. By hypothesis f relates to the same object as f_1 and f_2. This means $s \notin \mathsf{src}_N(f_1)$ and $s \notin \mathsf{src}_N(f_2)$, since otherwise s would make claims for multiple facts for a single object. Hence we have $\pi(s) = s$ and $\pi(f) = f$. This means $(s, f) = (\pi(s), \pi(f)) \in E'$ as required.

To complete the claim $E \subseteq E'$, suppose $(f, o) \in E$. There are again three cases: $f = f_1$, $f = f_2$, or $f \notin \{f_1, f_2\}$. In each case the definition of π and $\pi(N)$ easily yield $(f, o) \in E'$. Hence $E \subseteq E'$.

Now for the reverse direction: we must show $E' \subseteq E$. Let $(x, y) \in E'$. By definition of a graph isomorphism, we have $(\pi^{-1}(x), \pi^{-1}(y)) \in E$. Using $\pi^{-1} = \pi$ and the first part we get $(\pi(x), \pi(y)) = (\pi^{-1}(x), \pi^{-1}(y)) \in E \subseteq E'$. The definition of a graph isomorphism then yields $(x, y) \in E$ and so $E' \subseteq E$. Hence $E = E'$ and $N = N'$.

To conclude the proof, we apply Symmetry of T to get

$$\begin{aligned} f_1 \preceq_N^T f_2 &\iff \pi(f_1) \preceq_{N'}^T \pi(f_2) \\ &\iff f_2 \preceq_{N'}^T f_1 \\ &\iff f_2 \preceq_N^T f_1 \end{aligned}$$

and so $f_1 \approx_N^T f_2$ as required. □

Proof of Theorem 4.4.1. Suppose T is an operator satisfying Symmetry, Monotonicity and POI. Let $N \in \mathcal{N}$, $o \in \mathcal{O}$ and $f_1, f_2 \in \mathsf{obj}_N^{-1}(o)$. We need to show that $f_1 \preceq_N^T f_2$ iff $|\mathsf{src}_N(f_1)| \leq |\mathsf{src}_N(f_2)|$.

Let N' be the network obtained from N by removing all claims for facts other than those for object o; that is, $N' = (V, E')$ where E is the set of edges in N and

$$E' = (E \cap (\mathcal{S} \times \mathsf{obj}_N^{-1}(o))) \cup (E \cap (\mathcal{F} \times \mathcal{O}))$$

Note that the fact-object affiliations are the same in N' as in N, and the set of sources for each fact in $\mathsf{obj}_N^{-1}(o)$ is the same. Therefore POI applies, and it is sufficient to show that $f_1 \preceq_{N'}^T f_2$ iff $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$.

First suppose $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. If $|\mathsf{src}_{N'}(f_1)| = |\mathsf{src}_{N'}(f_2)|$, then we have $f_1 \approx_{N'}^T f_2$ by Symmetry and Lemma A.1.1; in particular $f_1 \preceq_{N'}^T f_2$. Otherwise $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Removing k sources from f_2 to obtain a new network N'', we can apply the lemma to get $f_1 \approx_{N''}^T f_2$. We may then add these sources *back* to obtain N' again; k applications of Monotonicity then give $f_1 \prec_{N'}^T f_2$ as required.

To complete the proof we show that $f_1 \preceq^T_{N'} f_2$ implies $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. Indeed, suppose $f_1 \preceq^T_{N'} f_2$ but $|\mathsf{src}_{N'}(f_1)| > |\mathsf{src}_{N'}(f_2)|$. Then the argument above gives $f_1 \succ^T_{N'} f_2$, which is clearly a contradiction. Hence the proof is complete. □

A.2 Proof of Theorem 4.4.3

The proof of this theorem is similar in spirit to that of Theorem 4.4.1. Like in that case, a preliminary result is required first.

Lemma A.2.1. *Let N be a network and $f_1, f_2 \in \mathcal{F}$. Write $o_1 = \mathsf{obj}_N(f_1)$, $o_2 = \mathsf{obj}_N(f_2)$. Suppose N has the following properties:*

1. *There is $o^* \in \mathcal{O} \setminus \{o_1, o_2\}$ such that $f \in \mathcal{F} \setminus \{f_1, f_2\} \implies \mathsf{obj}_N(f) = o^*$; and*
2. *$\mathsf{src}_N(f) = \emptyset$ for all $f \in \mathcal{F} \setminus \{f_1, f_2\}$.*

Then for any operator T satisfying Symmetry, $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)|$ implies $f_1 \approx^T_N f_2$.

Proof. The proof is similar to that of Lemma A.1.1. Suppose $|src_N(f_1)| = |\mathsf{src}_N(f_2)|$. Write

$$Q_1 = \mathsf{src}_N(f_1) \setminus \mathsf{src}_N(f_2)$$
$$Q_2 = \mathsf{src}_N(f_2) \setminus \mathsf{src}_N(f_1)$$

Then $|Q_1| = |Q_2|$, so there exists a bijection $\varphi : Q_1 \to Q_2$. Define a permutation π as follows:

$$\pi(s) = \begin{cases} \varphi(s) & \text{if } s \in Q_1 \\ \varphi^{-1}(s) & \text{if } s \in Q_2 \\ s & \text{otherwise} \end{cases}$$

$$\pi(f) = \begin{cases} f_2 & \text{if } f = f_1 \\ f_1 & \text{if } f = f_2 \\ f & \text{otherwise} \end{cases}$$

$$\pi(o) = \begin{cases} o_2 & \text{if } o = o_1 \\ o_1 & \text{if } o = o_2 \\ o & \text{otherwise} \end{cases}$$

That is, π swaps f_1 and f_2, swaps o_1 and o_2, and swaps sources in Q_1 with their counterparts in Q_2. Note that $\pi = \pi^{-1}$. Write $N' = \pi(N)$. We claim that $N' = N$. Write E, E' for the edges in N and N' respectively. First we show that $E \subseteq E'$. For this, first suppose $(s, f) \in E$ for some $s \in \mathcal{S}$, $f \in \mathcal{F}$. By definition of E, either $f = f_1$ or $f = f_2$.

Case 1: $f = f_1$. Here $\pi(f) = f_2$, and we have either $s \in Q_1$ or $s \in \mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2)$. In the first case, $\pi(s) = \varphi(s) \in Q_2 \subseteq \mathsf{src}_N(f_2) = \mathsf{src}_N(\pi(f))$. In the second case $\pi(s) = s \in \mathsf{src}_N(f_2) = \mathsf{src}_N(\pi(f))$. In either case, $(\pi(s), \pi(f)) \in E$.

Applying the definition of a graph isomorphism we get $(\pi(\pi(s)), \pi(\pi(f))) \in E'$. But $\pi = \pi^{-1}$, so this means $(s, f) \in E'$ as desired.

Case 2: $f = f_2$. This case is similar. Here $\pi(f) = f_1$. If $s \in Q_2$, then $\pi(s) = \varphi^{-1}(s) \in Q_1 \subseteq \mathsf{src}_N(f_1) = \mathsf{src}_N(\pi(f))$. Otherwise $s \in \mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2)$ and $\pi(s) = s \in \mathsf{src}_N(f_1) = \mathsf{src}_N(\pi(f))$. Again, we have $(\pi(s), \pi(f)) \in E$ in either case, so $(s, f) \in E'$.

Note that these two cases cover all possibilities since by hypothesis $\mathsf{src}_N(f) = \emptyset$ if $f \notin \{f_1, f_2\}$.

Next, suppose $(f, o) \in E$. If $f = f_1$ then $o = o_1$, so $(\pi(f), \pi(o)) = (f_2, o_2) \in E$. Similarly if $f = f_2$ then $o = o_2$ and $(\pi(f), \pi(o)) = (f_1, o_1) \in E$. If $f \notin \{f_1, f_2\}$ then $\pi(f) = f$ and $o = o^*$, so $\pi(o) = o$. We see that in all cases, $(\pi(f), \pi(f)) \in E$. Applying the same argument as in case 1 above, we see that $(f, o) \in E'$. This shows $E \subseteq E'$.

To complete the claim that $N = N'$ we must show $E' \subseteq E$. This can be shown using the same argument used in Lemma A.1.1. Indeed, suppose $(x, y) \in E'$. Then by definition of a graph isomorphism, $(\pi^{-1}(x), \pi^{-1}(y)) \in E$. Using the fact that $\pi = \pi^{-1}$ and $E \subseteq E'$ we get $(\pi(x), \pi(y)) \in E'$, which then yields $(x, y) \in E$ as required. Hence $E = E'$ and $N = N'$.

Finally, using Symmetry of T we have

$$\begin{aligned} f_1 \preceq^T_N f_2 &\iff \pi(f_1) \preceq^T_{\pi(N)} \pi(f_2) \\ &\iff f_2 \preceq^T_{N'} f_1 \\ &\iff f_2 \preceq^T_N f_1 \end{aligned}$$

Consequently $f_1 \approx^T_N f_2$. □

Proof of Theorem 4.4.3. The 'if' direction is clear since *Voting* satisfies Strong Independence, Monotonicity and Symmetry (see Theorem 4.5.1). For the other direction, suppose T satisfies the stated axioms. Let N be a network and $f_1, f_2 \in \mathcal{F}$. We will construct a network N' where all claims for facts other than f_1, f_2 are removed, and these facts are grouped under a single object. To do so, let $o_1 = \mathsf{obj}_N(f_1)$, $o_2 = \mathsf{obj}_N(f_2)$ and take $o^* \in \mathcal{O} \setminus \{o_1, o_2\}$. Define an edge set E' by

$$(s, f) \in E' \iff f \in \{f_1, f_2\} \text{ and } s \in \mathsf{src}_N(f)$$

$$(f, o) \in E' \iff (f \in \{f_1, f_2\} \text{ and } o = \mathsf{obj}_N(f)) \text{ or } (f \notin \{f_1, f_2\} \text{ and } o = o^*)$$

Then let N' be the network with edge set E'. Note that $\mathsf{src}_{N'}(f_j) = \mathsf{src}_N(f_j)$. By Strong Independence it is therefore sufficient to show that $f_1 \preceq^T_{N'} f_2$ iff $|src_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. Note also that N' satisfies the hypothesis of Lemma A.2.1.

Now, suppose $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. If $|\mathsf{src}_{N'}(f_1)| = |\mathsf{src}_{N'}(f_2)|$ then by Lemma A.2.1 $f_1 \approx^T_{N'} f_2$, and in particular $f_1 \preceq^T_{N'} f_2$.

Otherwise, $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Consider N'' where k sources from $\mathsf{src}_{N'}(f_2)$ are removed, and all other claims remain. By the lemma, $f_1 \approx^T_{N''} f_2$. Applying Monotonicity k times we can produce N' from N'' and get $f_1 \prec^T_{N'} f_2$ as desired.

For the other implication, suppose $f_1 \preceq^T_{N'} f_2$ and, for contradiction, $|\mathsf{src}_{N'}(f_1)| > |\mathsf{src}_{N'}(f_2)|$. Applying Monotonicity again as above gives $f_1 \succ^T_{N'} f_2$ and the required contradiction. □

A.3 Proof of Theorem 4.5.1

Proof. We will show that *Voting* satisfies Symmetry, Unanimity, Groundedness, Monotonicity, POI, Strong Independence and PCI, and that Coherence is *not* satisfied. For Symmetry and PCI we use the (stronger) numerical variants *numerical Symmetry* and *numerical PCI*, introduced in Section 4.5.2. T will denote the (numerical) *Voting* operator in what follows.

Symmetry. Suppose N and $\pi(N)$ are equivalent networks. Let $f \in \mathcal{F}$. By definition of equivalent networks we have $s \in \mathsf{src}_N(f)$ iff $\pi(s) \in \mathsf{src}_{\pi(N)}(\pi(f))$ for all $s \in \mathcal{S}$. Consequently π restricted to $\mathsf{src}_N(f)$ is a bijection into $\mathsf{src}_{\pi(N)}(\pi(f))$, and hence

$$T_N(f) = |\mathsf{src}_N(f)| = |\mathsf{src}_{\pi(N)}(\pi(f))| = T_{\pi(N)}(\pi(f))$$

Now let $s \in \mathcal{S}$. Clearly we have $T_N(s) = 1 = T_{\pi(N)}(\pi(s))$. Hence T satisfies numerical Symmetry and therefore Symmetry.

Unanimity and Groundedness. Suppose $N \in \mathcal{N}$ and $f \in \mathcal{F}$. If $\mathsf{src}_N(f) = \mathcal{S}$ then for any $g \in \mathcal{F}$,

$$T_N(g) = |\mathsf{src}_N(g)| \leq |\mathcal{S}| = |\mathsf{src}_N(f)| = T_N(f)$$

so $g \preceq^T_N f$ and Unanimity is satisfied. If instead $\mathsf{src}_N(f) = \emptyset$, we have

$$T_N(g) = |\mathsf{src}_N(g)| \geq 0 = |\mathsf{src}_N(f)| = T_N(f)$$

so $f \preceq^T_N g$ and Groundedness is satisfied.

Monotonicity. Let N, N', s and f be as given in the statement of Monotonicity. It is clear that $|\mathsf{src}_{N'}(f)| = |\mathsf{src}_N(f)| + 1$. Also, for any $g \in \mathcal{F}$, $g \neq f$, the set of sources in N' is the same as in N but with s possibly removed. Hence $|\mathsf{src}_{N'}(g)| \leq |\mathsf{src}_N(g)$. Therefore $g \preceq^T_N f$ implies

$$|\mathsf{src}_{N'}(g)| \leq |\mathsf{src}_N(g)| \leq |\mathsf{src}_N(f)| < |\mathsf{src}_{N'}(f)|$$

and so $g \prec^T_{N'} f$ as required.

Independence axioms. Next we show Strong Independence, which implies POI. Suppose $N_1, N_2 \in \mathcal{N}$, $f_1, f_2 \in \mathcal{F}$ and $\mathsf{src}_{N_1}(f_j) = \mathsf{src}_{N_2}(f_j)$ for each $j \in \{1, 2\}$. Clearly we have

$$T_{N_1}(f_j) = |\mathsf{src}_{N_1}(f_j)| = |\mathsf{src}_{N_2}(f_j)| = T_{N_2}(f_j)$$

Consequently

$$\begin{aligned} f_1 \preceq^T_{N_1} f_2 &\iff T_{N_1}(f_1) \leq T_{N_1}(f_2) \\ &\iff T_{N_2}(f_1) \leq T_{N_2}(f_2) \\ &\iff f_1 \preceq^T_{N_2} f_2 \end{aligned}$$

as required for Strong Independence.

For PCI we proceed as with Symmetry by showing numerical PCI. Let N_1, N_2 have a common connected component G. Let $f \in G \cap \mathcal{F}$. By definition of a connected component, $s \in \mathsf{src}_{N_1}(f)$ iff $s \in \mathsf{src}_{N_2}(f)$, so $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$. Hence

$$T_{N_1}(f) = |\mathsf{src}_{N_1}(f)| = |\mathsf{src}_{N_2}(f)| = T_{N_2}(f)$$

For $s \in G \cap \mathcal{S}$, we trivially have $T_{N_1}(s) = 1 = T_{N_1}(s)$. Hence numerical PCI is satisfied.

Coherence. The violation of Coherence follows from Theorem 4.4.2, since we have already shown that Symmetry, Monotonicity and POI are satisfied. □

A.4 Proof of Lemma 4.5.2

Proof. The first statement follows easily from the definition of the limit. We shall prove only the second one.

First we prove the 'if' direction. Write $D = T^*_N(f_1) - T^*_n(f_2)$. We need to show that $D < 0$. Write $d_n = T^n_N(f_1) - T^n_N(f_2)$ so that $D = \lim_{n\to\infty} d_n$. Take $\varepsilon = \rho/2 > 0$. Then for sufficiently large n we have $d_n \leq -\rho/2 < 0$. Taking $n \to \infty$, we have $D = \lim_{n\to\infty} d_n \leq -\rho/2 < 0$ as required.

For the 'only if' direction, suppose $D < 0$. Let $\rho = -D$. Then for any $\varepsilon > 0$, by the definition of the limit there is $K \in \mathbb{N}$ such that $|d_n - D| < \varepsilon$ for $n \geq K$; in particular, $d_n < \varepsilon + D = \varepsilon - \rho$ as required. □

A.5 Proof of Theorem 4.5.2

The following results will be helpful to simplify the proof of Theorem 4.5.2.

Lemma A.5.1. norm *has the following properties.*

1. norm *preserves numerical Symmetry, in the sense that* $\mathsf{norm}(T)$ *satisfies numerical Symmetry whenever* T *does.*

2. norm *leaves rankings unchanged, in the following sense. For* $T \in \mathcal{T}_{Num}$*,* $N \in \mathcal{N}$*,* $s_1, s_2 \in \mathcal{S}$*,* $f_1, f_2 \in \mathcal{F}$*,*

$$s_1 \sqsubseteq_N^T s_2 \iff s_1 \sqsubseteq_N^{\mathsf{norm}(T)} s_2$$
$$f_1 \preceq_N^T f_2 \iff f_1 \preceq_N^{\mathsf{norm}(T)} f_2$$

Proof. For part (i), suppose T satisfies numerical Symmetry, and write $T' = U(T)$. Let N and $\pi(N)$ be equivalent networks. First note that

$$\max_{x\in\mathcal{S}} |T_N(x)| = \max_{x\in\mathcal{S}} |T_{\pi(N)}(\pi(x))| = \max_{x\in\mathcal{S}} |T_{\pi(N)}(x)|$$

where the second equality follows since π restricted to $\mathcal{S}$ is a surjection into $\mathcal{S}$ by the definition of equivalent networks. If this maximum is 0, then $T'_N(s) = 0 = T'_{\pi(N)}(s)$ for all $s \in \mathcal{S}$. Otherwise,

$$T'_N(s) = \frac{T_N(s)}{\max\limits_{x\in\mathcal{S}} |T_N(x)|} = \frac{T_{\pi(N)}(\pi(s))}{\max\limits_{x\in\mathcal{S}} |T_{\pi(N)}(x)|} = T'_{\pi(N)}(\pi(s))$$

One can show that $T'_N(f) = T'_{\pi(N)}(\pi(f))$ by an identical argument. Hence $T' = U(T)$ satisfies numerical Symmetry also.

Now we prove part (ii). First suppose $s_1 \sqsubseteq_N^T s_2$. Write $T' = \mathsf{norm}(T)$. We have $T'_N(x) = \alpha T_N(x)$ for some $\alpha \geq 0$ and all $x \in \mathcal{S}$ (either $\alpha = 1/\max_{x\in\mathcal{S}} |T_N(x)|$ or $\alpha = 0$). We therefore have

$$\begin{aligned} s_1 \sqsubseteq_N^T s_2 &\implies T_N(s_1) \leq T_N(s_2) \\ &\implies \alpha T_N(s_1) \leq \alpha T_N(s_2) \\ &\implies T'_N(s_1) \leq T'_N(s_2) \\ &\implies s_1 \sqsubseteq_N^{T'} s_2 \end{aligned}$$

as desired.

Now suppose $s_1 \sqsubseteq_N^{T'} s_2$, i.e. $\alpha T_N(s_1) \leq \alpha T_N(s_2)$. If $\alpha > 0$ then dividing by α readily gives $s_1 \sqsubseteq_N^T s_2$. Otherwise, $\alpha = 0$. This means $\max_{x\in\mathcal{S}} |T_N(x)| = 0$, and thus $T_N(x) = 0$ for all $x \in \mathcal{S}$. In particular $T_N(s_1) = 0 \leq 0 = T_N(s_2)$ so $s_1 \sqsubseteq_N^T s_2$.

The second statement regarding fact ranking may be shown using an identical argument. □

Corollary A.5.1. norm *preserves Coherence, Unanimity, Groundedness and PCI.*

Proof of Theorem 4.5.2. Throughout this proof, $(T^n)_{n\in\mathbb{N}}$ will denote the iterative operator *Sums*, T^* will denote the limit operator, and $U = \mathsf{norm} \circ U^{\mathsf{Sums}}$ will denote the update function for *Sums*.

Coherence. Source-Coherence was shown in the main text. The proof that Fact-Coherence is satisfied is similar, and uses Lemma 4.5.3. Suppose $N \in \mathcal{N}$, $T = T^n$ for some $n \in \mathbb{N}$, $\varepsilon, \rho > 0$, and $\mathsf{src}_N(f_1)$ is (ε, ρ)-less trustworthy than $\mathsf{src}_N(f_2)$ with respect to N and $\tilde{T}$ under a bijection φ, where $\tilde{T} = U(T)$. Let $\hat{s} \in \mathsf{src}_N(f_1)$ be such that $\tilde{T}_N(s) - \tilde{T}_N(\varphi(s)) \leq \varepsilon - \rho$.

Write $T' = U^{\mathsf{Sums}}(T)$ so that $\tilde{T} = \mathsf{norm}(T')$, and set

$$\alpha = \frac{1}{\max\limits_{x\in\mathcal{S}} |T'_N(x)|}$$

We may assume without loss of generality that $\varepsilon < \frac{1}{|\mathcal{S}|}\rho$. Note that for $s \in \mathcal{S}$, $\tilde{T}_N(s) = \alpha T'_N(s)$ and therefore $T'_N(s) = \frac{1}{\alpha}\tilde{T}_N(s)$. Writing

$$\beta = \frac{1}{\max\limits_{y\in\mathcal{F}} |T'_N(y)|}$$

and applying a similar argument as for showing Source-Coherence, we find

$$\begin{aligned}
\tilde{T}_N(f_1) - \tilde{T}_N(f_2) &= \beta \sum_{s\in\mathsf{src}_N(f_1)} \Big(T'_N(s) - T'_N(\varphi(s))\Big) \\
&= \frac{\beta}{\alpha} \sum_{s\in\mathsf{src}_N(f_1)} \Big(\tilde{T}_N(s) - \tilde{T}_N(\varphi(s))\Big) \\
&= \frac{\beta}{\alpha} \left[\underbrace{\Big(\tilde{T}_N(\hat{s}) - \tilde{T}_N(\varphi(\hat{s}))\Big)}_{\leq \varepsilon-\rho} + \sum_{s\in\mathsf{src}_N(f_1)\setminus\{\hat{s}\}} \underbrace{\Big(\tilde{T}_N(s) - \tilde{T}_N(\varphi(s))\Big)}_{\leq\varepsilon}\right] \\
&\leq \frac{\beta}{\alpha} \cdot \underbrace{\Big(|\mathcal{S}|\varepsilon - \rho\Big)}_{<0}
\end{aligned}$$

Now we need to bound β/α from below. Since we assume $T = T^n$ for some $n \in \mathbb{N}$, for any $y \in \mathcal{F}$ we have

$$|T'_N(y)| = \sum_{s\in\mathsf{src}_N(y)} \underbrace{T'_N(s)}_{\leq|\mathcal{F}|} \leq |\mathsf{src}_N(y)| \cdot |\mathcal{F}| \leq |\mathcal{S}| \cdot |\mathcal{F}|$$

Therefore

$$\beta \geq \frac{1}{|\mathcal{S}| \cdot |\mathcal{F}|}$$

Next, we claim there is some fact $\bar{f} \in \mathcal{F}$ with $T_N(\bar{f}) \geq 1/2$ and $\mathsf{src}_N(\bar{f}) \neq \emptyset$. Indeed, if $T = T^1 = T^{\text{fixed}}$ then take any fact with at least one associated source.[1] Otherwise, since we assume not all scores are 0 in the limit, there is some $\bar{f}$ with $T_N(\bar{f}) = 1$ due to the application of norm. Clearly $\mathsf{src}_N(\bar{f}) \neq \emptyset$, since we would have $T_N(\bar{f}) = 0$ otherwise.

Let $\bar{x} \in \mathsf{src}_N(\bar{f})$. Then

$$|T'_N(\bar{x})| = T'_N(\bar{x}) = \underbrace{T_N(\bar{f})}_{\geq 1/2} + \underbrace{\sum_{f \in \mathsf{facts}_N(\bar{x}) \setminus \{\bar{f}\}} T_N(f)}_{\geq 0} \geq \frac{1}{2}$$

This means

$$\frac{1}{\alpha} = \max_{x \in \mathcal{S}} |T'_N(x)| \geq |T'_N(\bar{x})| \geq \frac{1}{2}$$

and so, finally,

$$\frac{\beta}{\alpha} \geq \frac{1}{|\mathcal{S}| \cdot |\mathcal{F}|} \cdot \frac{1}{2}$$

Combined with what was shown before, this means

$$\tilde{T}_N(f_1) - \tilde{T}_N(f_2) \leq \frac{1}{2 \cdot |\mathcal{S}| \cdot |\mathcal{F}|}\Big(|\mathcal{S}|\varepsilon - \rho\Big)$$

and Fact-Coherence follows from Lemma 4.5.3.

Symmetry. As a consequence of Lemma 4.5.4, to show Symmetry it is sufficient to show that T^{fixed} satisfies numerical Symmetry, and that $U = \mathsf{norm} \circ U^{\text{Sums}}$ preserves numerical Symmetry. Since T^{fixed} is constant with value $1/2$, it is clear that numerical Symmetry is satisfied. Moreover, Lemma A.5.1 part (i) already shows that norm preserves numerical Symmetry, so we only need to show that U^{Sums} does.

To that end, suppose $T \in \mathcal{T}_{Num}$ satisfies numerical symmetry, and write $T' = U^{\text{Sums}}(T)$. Let N and $\pi(N)$ be equivalent networks and $s \in \mathcal{S}$. Then

$$T'_{\pi(N)}(\pi(s)) = \sum_{y \in \mathsf{facts}_{\pi(N)}(\pi(s))} T_{\pi(N)}(y)$$

Note that $f \in \mathsf{facts}_N(s)$ iff $\pi(f) \in \mathsf{facts}_{\pi(N)}(\pi(s))$. Rephrased slightly, we have $y \in \mathsf{facts}_{\pi(N)}(\pi(s))$ iff $\pi^{-1}(y) \in \mathsf{facts}_N(s)$. Hence we may make a 'substitution' $f = \pi^{-1}(y)$ and sum over $\mathsf{facts}_N(s)$, i.e.

$$T'_{\pi(N)}(\pi(s)) = \sum_{f \in \mathsf{facts}_N(s)} T_{\pi(N)}(\pi(f))$$

[1] Note that this is always possible since a truth discovery network contains at least one claim by definition.

Applying numerical symmetry for T, we get

$$\begin{aligned} T'_{\pi(N)}(\pi(s)) &= \sum_{f \in \mathsf{facts}_N(s)} T_N(f) \\ &= T'_N(s) \end{aligned}$$

Following the same tactic, one may also show that $T'_{\pi(N)}(\pi(f)) = T'_N(f)$ for all $f \in \mathcal{F}$. Hence U^{Sums} preserves numerical Symmetry, and we are done.

Unanimity and Groundedness. Unanimity and Groundedness can be proved together using Lemma 4.5.5 and Corollary A.5.1. By these results it is sufficient that T^{fixed} satisfies Unanimity and Groundedness – this is trivial – and that U^{Sums} preserves them.

Suppose T satisfies Unanimity and Groundedness and write $T' = U^{\mathsf{Sums}}(T)$. Assume without loss of generality that $T = T^n$ for some $n \in \mathbb{N}$ so that $T'_N \geq 0$. Suppose $N \in \mathcal{N}$, $f \in \mathcal{F}$ and that $\mathsf{src}_N(f) = \mathcal{S}$. Let $g \in \mathcal{F}$. We must show that $g \preceq^{T'}_N f$. We have

$$T'_N(g) = \sum_{s \in \mathsf{src}_N(g)} T'_N(s) \leq \sum_{s \in \mathcal{S}} T'_N(s) = T'_N(f)$$

i.e. $g \preceq^{T'}_N f$ as required for Unanimity. For Groundedness, suppose $\mathsf{src}_N(f) = \emptyset$. We must show $f \preceq^{T'}_N g$. Indeed, the sum in the expression for $T'_N(f)$ is taken over the empty set, which by convention is 0. Since $T'_N(g) \geq 0$, we are done. □

A.6 Proof of Theorem 4.5.3

Proof. Here we give only the technical details for the argument showing *SC-Voting* satisfies Symmetry, since the results for the other axioms were given in the main text.

Symmetry. Since *Voting* satisfies Symmetry, it is clear that $f_1 \preceq^{T^{SCV}}_N f_2$ iff $\pi(f_1) \preceq^{T^{SCV}}_{\pi(N)} \pi(f_2)$ for any equivalent networks N and $\pi(N)$. We need to show that $s_1 \sqsubseteq^{T^{SCV}}_N s_2$ iff $\pi(s_1) \sqsubseteq^{T^{SCV}}_{\pi(N)} \pi(s_2)$.

First we will show that $\lhd_N$ and $\lhd_{\pi(N)}$ have a similar symmetry property: $s_1 \lhd_N s_2$ iff $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$. Indeed, suppose $s_1 \lhd_N s_2$. Then there is a bijection $\varphi : \mathsf{facts}_N(s_1) \to \mathsf{facts}_N(s_2)$ with $f \preceq^{T^{SCV}}_N \varphi(f)$, and there is some $\hat{f}$ with $\hat{f} \prec^{T^{SCV}}_N \varphi(\hat{f})$.

It can be seen that π restricted to $\mathsf{facts}_N(s_i)$ is a bijection into $\mathsf{facts}_{\pi(N)}(\pi(s_i))$. Let π_1 and π_2 denote these restrictions for $i = 1, 2$ respectively. Set $\theta = \pi_2 \circ \varphi \circ \pi_1^{-1}$, so that θ maps $\mathsf{facts}_{\pi(N)}(\pi(s_1))$ into $\mathsf{facts}_{\pi(N)}(\pi(s_2))$. As a composition of bijections, θ is itself bijective.

Let $g \in \mathsf{facts}_{\pi(N)}(\pi(s_1))$. Write $f = \pi_1^{-1}(g) \in \mathsf{facts}_N(s_1)$. By the property of φ, we have

$$f \preceq_N^{T^{SCV}} \varphi(f)$$

By the symmetry property of the fact-ranking (which follows from symmetry of *Voting*), we can apply π to the above to get

$$\pi(f) \preceq_{\pi(N)}^{T^{SCV}} \pi(\varphi(f))$$

Since $f \in \mathsf{facts}_N(s_1)$ and $\varphi(f) \in \mathsf{facts}_N(s_2)$, we have $\pi(f) = \pi_1(f)$ and $\pi(\varphi(f)) = \pi_2(\varphi(f))$. Using this fact in the above inequality and recalling $f = \pi^{-1}(g)$ we get

$$g = \pi_1(f) = \pi(f) \preceq_{\pi(N)}^{T^{SCV}} \pi(\varphi(f)) = \pi_2(\varphi(f)) = \pi_2(\varphi(\pi_1^{-1}(g))) = \theta(g)$$

i.e. $g \preceq_{\pi(N)}^{T^{SCV}} \theta(g)$. Applying the same argument with $\hat{g} = \pi_1^{-1}(\hat{f})$ we get $\hat{g} \prec_{\pi(N)}^{T^{SCV}} \theta(\hat{g})$.

This shows that $\mathsf{facts}_{\pi(N)}(\pi(s_1))$ is less believable than $\mathsf{facts}_{\pi(N)}(\pi(s_2))$ with respect to *SC-Voting* (whose fact-ranking coincides with *Voting*) in $\pi(N)$ under θ. Hence $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$.

We have shown $s_1 \lhd_N s_2 \implies \pi(s_1) \lhd_{\pi(N)} \pi(s_2)$. For the converse implication, apply the same argument starting from $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$ with the π^{-1}.

Next, we note that for $i = 1, 2$ and any $t \in \mathcal{S}$,

$$\begin{aligned} t \in W_N(s_i) &\iff t \lhd_N s_i \\ &\iff \pi(t) \lhd_{\pi(N)} \pi(s_i) \\ &\iff \pi(t) \in W_{\pi(N)}(\pi(s_i)) \end{aligned}$$

Consequently π restricted to $W_N(s_i)$ is a bijection into $W_{\pi(N)}(\pi(s_i))$, which means $|W_N(s_i)| = |W_{\pi(N)}(\pi(s_i))|$. Finally, this means

$$\begin{aligned} s_1 \sqsubseteq_N^{T^{SCV}} s_2 &\iff |W_N(s_1)| \leq |W_N(s_2)| \\ &\iff |W_{\pi(N)}(\pi(s_1))| \leq |W_{\pi(N)}(\pi(s_2))| \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^{SCV}} \pi(s_2) \end{aligned}$$

as required for Symmetry. □

A.7 Proof of Theorem 4.5.5

Proof. Here we show that *UnboundedSums* satisfies Symmetry, PCI, Unanimity and Groundedness, since the other axioms were dealt with in the main text.

Throughout the proof, let $(T^n)_{n \in \mathbb{N}}$ denote *UnboundedSums*, T^* denote the ordinal limit of *UnboundedSums*, and for a network N let J_N be as in Theorem 4.5.4. Then the rankings in N induced by T^n for $n \geq J_N$ are the same as T^*.

Symmetry. In the proof of Theorem 4.5.2, we saw that the update function U^{Sums} preserves numerical Symmetry, in the sense that if T satisfies numerical Symmetry then $U^{\text{Sums}}(T)$ does also. Since it is clear that the prior operator for *UnboundedSums* satisfies numerical Symmetry, T^n satisfies numerical Symmetry and consequently Symmetry for all $n \in \mathbb{N}$.

Now, let N and $\pi(N)$ be equivalent networks. Let $J, J' \in \mathbb{N}$ be such that $T^*(N)$ and $T^*(\pi(N))$ are given by T_N^J and $T_{\pi(N)}^{J'}$ respectively and take $n \geq \max\{J, J'\}$. For $s_1, s_2 \in \mathcal{S}$ we have by Symmetry of T^n,

$$\begin{aligned} s_1 \sqsubseteq_N^{T^*} s_2 &\iff s_1 \sqsubseteq_N^{T^n} s_2 \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^n} \pi(s_2) \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^*} \pi(s_2) \end{aligned}$$

as required for Symmetry. Using an identical argument, one can show that $f_1 \preceq_N^{T^*} f_2$ iff $\pi(f) \preceq_{\pi(N)}^{T^*} \pi(f_2)$. Hence T^* satisfies Symmetry.

PCI. As with Symmetry, we will show that T^n satisfies numerical PCI, and consequently PCI, for all $n \in \mathbb{N}$. Let N_1, N_2 be networks with a common connected component G. Let $s \in G \cap \mathcal{S}$ and $f \in G \cap \mathcal{F}$. Note that $\mathsf{facts}_{N_1}(s) = \mathsf{facts}_{N_2}(s)$ and $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$ since by definition a source is connected to its facts and vice versa. For $n = 1$ we have

$$T_{N_1}^1(s) = 1 = T_{N_2}^1(s)$$

$$T_{N_1}^1(f) = |\mathsf{src}_{N_1}(f)| = |\mathsf{src}_{N_2}(f)| = T_{N_2}^1(f)$$

so T^1 has numerical PCI. Supposing T^n has numerical PCI for some $n \in \mathbb{N}$, we have

$$T_{N_1}^{n+1}(s) = \sum_{g \in \mathsf{facts}_{N_1}(s)} \underbrace{T_{N_1}^n(g)}_{=T_{N_2}^n(g)} = \sum_{g \in \mathsf{facts}_{N_2}(s)} T_{N_2}^n(g) = T_{N_2}^{n+1}(s)$$

and similarly

$$T_{N+1}^{n+1}(f) = T_{N_2}^{n+1}(f)$$

Hence, by induction, T^n has numerical PCI for all $n \in \mathbb{N}$, and we are done.

Unanimity and Groundedness. For Unanimity, suppose $\mathsf{src}_N(f) = \mathcal{S}$. For any $g \in \mathcal{F}$ and $n \in \mathbb{N}$ we have

$$\begin{aligned} T_N^n(g) &= \sum_{s \in \mathsf{src}_N(g)} T_N^n(s) \\ &\leq \sum_{s \in \mathsf{src}_N(g)} T_N^n(s) + \sum_{s \in \mathcal{S} \setminus \mathsf{src}_N(g)} T_N^n(s) \\ &= \sum_{s \in \mathcal{S}} T_N^n(s) \\ &= \sum_{s \in \mathsf{src}_N(f)} T_N^n(s) \\ &= T_N^n(f) \end{aligned}$$

so $g \preceq_N^{T^n} f$ for all $n \in \mathbb{N}$. Since the ranking of T^* corresponds to T^n for large n, we have $g \preceq_N^{T^*} f$ as required

For Groundedness, suppose $\mathsf{src}_N(f) = \emptyset$. Then $T_N^n(f) = 0$ for all $n \in \mathbb{N}$. For any $g \in \mathcal{F}$, we have $T_N^n(g) \geq 0 = T_N^n(f)$. Consequently $f \preceq_N^{T^n} g$ for all $n \in \mathbb{N}$. As above, this gives $f \preceq_N^{T^*} g$ as required. □