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

Acknowledgements

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 references beneath each citation?]

- Joseph Singleton and Richard Booth. "Towards an axiomatic approach to truth discovery". In: *Autonomous Agents and Multi-Agent Systems* 36.2 (2022), pp. 1–49
- 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 [**TODO:** replace with synthese]
- Joseph Singleton and Richard Booth. "Who's the Expert? On Multi-source Belief Change". In: *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*. Aug. 2022, pp. 331–340. DOI: 10.24963/kr.2022/33. URL: https://doi.org/10.24963/kr.2022/33
- [**TODO:** NMR]

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.

1.1 Social Choice Perspectives

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

1.2 Logic-based Perspectives

1.3 Overview

Chapter-by-chapter breakdown of the thesis.

2 Truth Discovery

[**TODO:** Introduction]

2.1 Preliminaries

In this section we give the basic definitions which form our formal framework.

Input. Intuitively, a truth discovery problem consists of a number of *sources* and a number of *objects* of interest. Each source provides a number of *claims*, where a claim is comprised of an object and a *value*. Different sources may give conflicting claims by providing different values for the same object. For simplicity, we only consider categorical values in this work. Note that while this restriction is made in some approaches in the literature [73, 100, 91, 30, 103], in general truth discovery methods also handle continuous values [64, 95].

To formalise this, let $\mathbb{S}$, $\mathbb{O}$ and $\mathbb{V}$ be infinite, disjoint sets, representing the possible sources, objects and values. The input to the truth discovery problem is a *network*, defined as follows.

Definition 2.1.1. *A* truth discovery network *is a tuple* $N = (S, O, D, R)$*, where*

- $S \subseteq \mathbb{S}$ *is a finite set of* sources*.*
- $O \subseteq \mathbb{O}$ *is a finite set of* objects*.*

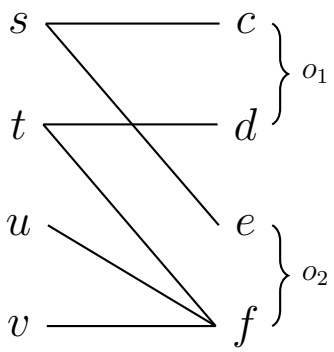


Figure 2.1: Illustrative example of a truth discovery problem, with sources s, t, u, v*, object* o_1 *with associated claims* c *and* d*, and* o_2 *with claims* e *and* f*.*

- $D = \{D_o\}_{o \in O}$ *are the* domains *of the objects, where each* $D_o \subseteq \mathbb{V}$ *is a finite set of values. We write* $V = \bigcup_{o \in O} D_o$.
- $R \subseteq S \times O \times V$ *is a set of* reports.

such that

1. *For each* $(s, o, v) \in R$, *we have* $v \in D_o$.
2. *If* $(s, o, v) \in R$ *and* $(s, o, v') \in R$, *then* $v = v'$.

Note that while $\mathbb{S}$, $\mathbb{O}$ and $\mathbb{V}$ are infinite, each network is finite. The set R is the core data associated with the network: we interpret $(s, o, v) \in R$ as source s claiming that v is the true value for object o. Constraint (1) says that all claimed values are in the domain of the relevant object. Constraint (2) is a basic consistency requirement: a source cannot provide distinct values for a single object. That is, a source provides *at most one value* per object. Thus, while sources may be in conflict with *other sources*, they are not in conflict with themselves. While this is a simplifying assumption, we argue the truth discovery problem is still rich enough when conflicts only arise between distinct sources.

When a network N is understood, we often write S, O, D and R to implicitly refer to the components of N. Any decoration applied to N will also be applied to its components (e.g. N' has sources S', $\hat{N}$ has sources $\hat{S}$ etc…). If necessary, we write S_N, O_N, D_N and R_N to make the dependence on N explicit.

A *claim* is a pair $c = (o, v)$, where $o \in O$ and $v \in D_o$. We write $\mathrm{obj}(c) = o$ in this case, and let C denote the set of all possible claims in a network N, i.e.

$$C = \{(o, v) \mid o \in O, v \in D_o\}.$$

Note that not every claim is necessarily reported by some source. With slight abuse of notation, we write (s, c) for the report (s, o, v). Then R can be viewed as a subset of $S \times C$, i.e. a relation between sources and claims. In fact, we will take this claim-centric view in the remainder of the chapter, with objects and values only playing a role insofar as they tell us which claims are in conflict with one another.

Example 2.1.1. *The network illustrated in Fig. 2.1 is given by* $S = \{s, t, u, v\}$, $O = \{o_1, o_2\}$ *and* $D_{o_1} = D_{o_2} = \{\mathsf{true}, \mathsf{false}\}$. *We label the claims* $c = (o_1, \mathsf{true})$, $d = (o_1, \mathsf{false})$, $e = (o_2, \mathsf{true})$ *and* $f = (o_2, \mathsf{false})$. *Then* $R = \{(s, c), (s, e), (t, d), (t, f), (u, f), (v, f)\}$.

Example 2.1.1 highlights a special case of our framework: the "binary" case in which the domain of each object consists of two values $D_o = \{\mathsf{true}, \mathsf{false}\}$. In this case we can think of each object as a propositional variable. This brings us close to the setting studied in *judgment aggregation* [34] and, specifically (since sources do not necessarily provide a claim for each object) to the setting of *binary aggregation with abstentions* [18, 29]. An important difference, however, is that for simplicity we do not assume any *constraints* on the possible configurations of true claims across objects. That is, any combination of truth values is feasible. In judgment aggregation such an assumption has the effect of neutralising the impossibility results that arise in that domain (see e.g., [18]). We shall see later that that is not the case in our setting.

Notation. We introduce some notation to extract information about a network. For $c \in C$ and $s \in S$, write

$$\begin{aligned} \mathrm{src}_N(c) &= \{s \in S \mid (s,c) \in R\}, \\ \mathrm{cl}_N(s) &= \{c \in C \mid (s,c) \in R\}. \end{aligned}$$

The set of sources making a claim on object o is

$$\mathrm{src}_N(o) = \bigcup \{\mathrm{src}_N(c) \mid c \in C, \mathrm{obj}(c) = o\}.$$

The claims associated with o are

$$\mathrm{cl}_N(o) = \{c \in C \mid \mathrm{obj}(c) = o\}.$$

The set of claims in conflict with a given claim $c = (o, v)$, i.e. claims for o with a value other than v, is denoted by

$$\mathrm{conflict}_N(c) = \{(o, v') \mid v' \in D_o \setminus \{v\}\}.$$

The "antisources" of c are then defined to be the sources for claims conflicting with c:

$$\mathrm{antisrc}_N(c) = \bigcup \{\mathrm{src}_N(d) \mid d \in \mathrm{conflict}_N(c)\}.$$

Note that property (2) in the definition of a network ensures $\mathrm{src}_N(c) \cap \mathrm{antisrc}_N(c) = \emptyset$.

Output. With the input defined, we now come to the output of the truth discovery problem. The primary goal is to produce an assessment of the trustworthiness of the sources, and the *true values* for the objects. Approaches differ regarding values: some truth discovery methods output only a single value for each object [64, 27, 97], whereas others give an assessment of the believability (or confidence, probability etc…) of *each claim* (o, v) [100, 73, 40, 104, 102, 103]. We opt for the latter, more general, approach.

On the specific form of these assessments, we face a tension between the social choice and truth discovery perspectives. In social choice theory, one generally looks at *rankings*: e.g. the ranking of candidates in an election result according to a voting rule. Consequently, axioms are generally *ordinal properties*, which constrain how candidates (for example) compare *relative to each other*. In contrast, truth discovery methods universally use *numeric values*. This is more convenient for defining and using truth discovery methods in practise, and induces a ranking by simply comparing the numeric scores. The magnitude of the differences between scores also gives information about *confidence* in distinguishing sources and claims.

However, numeric scores are often not comparable between different methods (for example, some methods output probabilities, whereas others are interpreted as weights which may take negative values) and in general may not carry any semantic meaning at all. This means that meaningful axioms for truth discovery should not refer to specific numeric scores, but only the ranking they introduce.

We will ultimately take a hybrid approach: our methods and example will be defined in terms of numeric scores, but the axioms will only refer to ordinal properties. This approach is summarised succinctly by Altman and Tennenholtz [2], who

write of ranking systems: "We feel that the numeric approach is more suitable for defining and executing ranking systems, while the global ordinal approach is more suitable for axiomatic classification."

An *operator* maps each network to score and claim scores.

Definition 2.1.2. *A* truth discovery operator T *maps each network* N *to a function* $T_N : S_N \cup C_N \to \mathbb{R}$.

Intuitively, the higher the score $T_N(s)$ for a source $s \in S$, the *more trustworthy* s is, according to T on the basis of N. Similarly, the higher $T_N(c)$ for a claim $c \in C$, the *more believable* c is deemed to be. We define the source and claim rankings associated with T and N by

$$s \sqsubseteq_N^T s' \iff T_N(s) \leq T_N(s'),$$
$$c \preceq_N^T c' \iff T_N(c) \leq T_N(c').$$

Then $s \sqsubseteq_N^T s'$ if s' is at least as trustworthy as s, and similar for $\preceq_N^T$. Note that $\sqsubseteq_N^T$ and $\preceq_N^T$ are total preorders. We denote the strict parts by $\sqsubset_N^T$ and $\prec_N^T$ respectively, and the symmetric parts by $\simeq_N^T$ and $\approx_N^T$. We omit the sub- and super-scripts when N and T are clear from context.

Given that our axioms will only refer to the rankings produced by operators, two operators yielding exactly the same rankings – possibly with different scores – appear the same with respect to axiomatic analysis. We say operators T and T' are *ranking equivalent*, denoted $T \sim T'$, if for all networks N we have $\sqsubseteq_N^T = \sqsubseteq_N^{T'}$ and $\preceq_N^T = \preceq_N^{T'}$.

In Section 2.2 we will introduce operators defined as the limit of an iterative procedure. To allow for possible non-convergence we also consider *partial operators*, which assign a mapping $T_N : S \cup C \to \mathbb{R}$ for only a subset of networks.

2.2 Example Operators

In this section we capture several example operators from the literature in our framework: a baseline *voting* method and its generalisation to *weighted* voting, *Sums* [73], *TruthFinder* [100] and *CRH* [64]. As is the case with many methods in the literature, the latter three methods operate iteratively: starting with an initial estimate, scores are repeatedly updated according to some procedure until convergence. Typically the update procedure is recursive, with source scores being updated on the basis of the current claims scores, and vice versa. To simplify the definition and analysis of such methods, we will introduce the class of *recursive operators*.

2.2.1 Voting

It is common in the literature to evaluate truth discovery methods against a non-trust-aware method, such as a simple voting procedure.[1] Here we consider each source to "vote" for their claims, and claims are ranked according to the number of votes received, i.e. by $|\mathrm{src}_N(c)|$. While this ignores the trust aspect of truth discovery entirely, this method will be useful for us as an axiomatic baseline. For example, axioms which aim to address the trust aspect should not hold for voting,

and an axiom referring to the ranking of claims may be too strong if it does hold for voting.

Definition 2.2.1. *T^{vote} is the operator defined by*

$$T_N^{\text{vote}}(s) = 1,$$
$$T_N^{\text{vote}}(c) = |\operatorname{src}_N(c)|.$$

Applying T^{vote} to the network in Fig. 2.1, we have that all sources rank equally ($s \simeq t \simeq u \simeq v$) and $c \approx d \approx e \prec f$.

The problem with T^{vote} is that all reports are equally weighted. If we have a mechanism by which sources can be weighted by trustworthiness, the idea behind voting may still have some merit. We define *weighted voting* as follows.

Definition 2.2.2. *A* weighting *w maps each network N to a function $w_N : S \to \mathbb{R}$. The associated* weighted voting *operator T^w is defined by*

$$T_N^w(s) = w_N(s),$$
$$T_N^w(c) = \sum_{s \in \operatorname{src}_N(c)} w_N(s).$$

Note that T^{vote} arises via the weighting $w_N \equiv 1$. Note that a weighting is essentially just half of a truth discovery operator, where we only output scores for sources. This is completed to an operator T^w by letting the score for a claim be the sum of the weights of its sources. Note also that we allow the possibility of "untrustworthy" sources with $w_N(s) < 0$. Reports from such sources *decrease* the credibility of a claim.

Example 2.2.1. *Set*

$$w_N^{\text{agg}}(s) = \sum_{c \in \operatorname{cl}_N(s)} \frac{|\operatorname{src}_N(c)|}{|\operatorname{cl}_N(s)|}.$$

Then the weight assigned to a source s is the average number of sources agreeing with the claims of s. We call the corresponding operator Weighted Agreement*. Taking N from Fig. 2.1, we have $w_N^{\text{agg}}(s) = 1$, $w_N^{\text{agg}}(t) = 2$, $w_N^{\text{agg}}(u) = 3$, $w_N^{\text{agg}}(v) = 3$. Consequently,*

$$T_N^{w^{\text{agg}}}(c) = w_N^{\text{agg}}(s) = 1,$$
$$T_N^{w^{\text{agg}}}(d) = w_N^{\text{agg}}(t) = 2,$$
$$T_N^{w^{\text{agg}}}(e) = w_N^{\text{agg}}(s) = 1,$$
$$T_N^{w^{\text{agg}}}(f) = w_N^{\text{agg}}(t) + w_N^{\text{agg}}(u) + w_N^{\text{agg}}(v) = 8,$$

yielding the rankings $s \sqsubset t \sqsubset u \simeq v$ and $c \approx e \prec d \prec f$. Note that claim d fares better here than with T^{vote} due to its association with source t, who is more trustworthy than s.

[1] This is often called *majority voting* in the truth discovery literature (e.g. [63, 94, 64]), but using the terminology of social choice theory it is better described as *plurality voting*.

As we will see in [**TODO:** section reference], some operators do not correspond exactly to a weighting w, but give rise to the same rankings. Let us say an operator T is *weightable* if there exists a weighting w such that $T \sim T^w$. Given that weighted voting expresses a clear relationship between source and claim scores, this notion will greatly simplify axiomatic analysis in Section 2.5. [**TODO:** Check afterwards.]

2.2.2 Recursive Operators

To capture the mutual dependence between trust in sources and belief in claims, truth discovery methods generally involve recursive computation [73, 100, 96, 32, 103, 64, 40, 104]. Claim scores are updated on the basis of currently estimated source scores, before claim scores are updated on the basis of the new sources scores. If this process converges, the limiting scores should be a fixed-point of the update procedure, reflecting the desired mutual dependence. To formalise this idea, we define recursive operators.

Definition 2.2.3. *A* recursive scheme *is a tuple* $(\mathcal{D}, T^0, U)$*, where*

- $\mathcal{D}$ *is a set of operators.*
- $T^0 \in \mathcal{D}$ *is the* initial operator*.*
- $U : \mathcal{D} \to \mathcal{D}$ *is the* update function*.*

A recursive scheme converges *to an operator* T^* *if for all networks* N *and all* $z \in S \cup C$*,* $\lim_{n\to\infty} U^n(T_0)_N(z) = T^*_N(z)$*. In this case* T^* *is said to be the* limit *of the scheme.*

The main component of interest here is the update function U, which describes how the scores of one iteration are transformed to obtain scores for the next. The domain of operators $\mathcal{D}$ is used for technical reasons; for example, some operators need to exclude the trivial operator in which scores are identically zero in order for U to be well-defined.

Note that the limit operator T^* is unique, when it exists. We can consider any scheme to converge to a *partial* operator T^*, defined on the networks N such that $\lim_{n\to\infty} U^n(T_0)_N(z)$ exists for all $z \in S \cup C$. Convergence and fixed-point properties – i.e. whether $U(T^*) = T^*$ – will be discussed in Section 2.4. For now, we introduce examples of recursive operators from the literature.

Sums. Sums [73] is a simple and well-known operator adapted from the *Hubs and Authorities* [59] algorithm for ranking web pages. The premise is to extend the linear sum of weighted voting to both claim and source scores: we update the score of each source as the sum of the scores of its claims, and update the score of each claim as the sum of the scores of its sources. To prevent scores from growing without bound, they are normalised at each iteration by dividing by the maximum score (for sources and claims separately).

Definition 2.2.4. Sums *is the recursive scheme* $(\mathcal{D}, T^0, U)$*, where* $\mathcal{D}$ *is the set of all operators with scores in* $[0,1]$*,* $T^0_N \equiv 1/2$*, and* $U(T) = T'$*, with*

$$T'_N(s) = \alpha \sum_{c \in \mathrm{cl}_N(s)} T_N(c),$$
$$T'_N(c) = \beta \sum_{s \in \mathrm{src}_N(c)} T'_N(s).$$

where $\alpha = 1/\max_{t \in S} \left| \sum_{c \in \mathrm{cl}_N(t)} T_N(c) \right|$ *and* $\beta = 1/\max_{d \in C} \left| \sum_{s \in \mathrm{src}_N(d)} T'_N(s) \right|$ *are normalisation factors (which we set to 0 if the denominator is 0). Write* T^{sums} *for the associated limit operator.*

Taking the network N from Fig. 2.1, one can show that $T^{\mathrm{sums}}_N(s) = 0$, $T^{\mathrm{sums}}_N(t) = 1$ and $T^{\mathrm{sums}}_N(u) = T^{\mathrm{sums}}_N(v) = \sqrt{2}/2 \approx 0.7071$, giving a source ranking $s \sqsubset u \simeq v \sqsubset t$. For claims, we have $T^{\mathrm{sums}}_N(c) = T^{\mathrm{sums}}_N(e) = 0$, $T^{\mathrm{sums}}_N(d) = \sqrt{2} - 1 \approx 0.4142$ and $T_N(f) = 1$, giving a claim ranking $c \approx e \prec d \prec f$. Note that the claim ranking is identical to that of Example 2.2.1. For sources, we see that t moves strictly upwards in the ranking compared to Example 2.2.1. Intuitively, this is because source t claims a superset of the claims of u and v, so receives more weight from its claims at each iteration.

TruthFinder. TruthFinder [100] is a pseudo-probabilistic method, and was defined in the first work to introduce (and coin the phrase) truth discovery. It is formulated in a setting more general than ours: the authors suppose claims may *support* each other, as well as conflict, and that support of conflict may occur to varying degrees. Formally, each pair of claims c, c' has an "implication" value $\mathrm{imp}(c \to c') \in [-1, 1]$, where a negative value implies confidence in c should decrease confidence in c', and a positive value implies confidence in c should *increase* confidence in c'. In contrast, our framework assumes claims for the same object are mutually exclusive, so that all implications are negative. To express TruthFinder in our framework, we take $\mathrm{imp}(c \to c')$ to be $-\lambda$ if c and c' have the same object and 0 otherwise, for some fixed parameter $0 \leq \lambda \leq 1$.

Definition 2.2.5. *Given parameters* $\rho, \gamma \in (0, 1)$ *and* $\lambda \in [0, 1]$*,* TruthFinder *is the recursive scheme* $(\mathcal{D}, T^0, U)$*, where* $\mathcal{D}$ *is the set of operators with* $0 < T_N(s) < 1$ *for all* N *and* $s \in S$ *with* $\mathrm{cl}_N(s) \neq \emptyset$*,* $T^0 \equiv 0.9$*, and* $U(T) = T'$*, with*

$$T'_N(c) = \left[1 + \frac{\prod_{s \in \mathrm{src}_N(c)} (1 - T_N(s))^{\gamma}}{\prod_{t \in \mathrm{antisrc}_N(c)} (1 - T_N(t))^{\gamma \rho \lambda}} \right]^{-1}, \tag{2.1}$$
$$T'_N(s) = \sum_{c \in \mathrm{cl}_N(s)} \frac{T'_N(c)}{|\mathrm{cl}_N(s)|}. \tag{2.2}$$

We write T^{tf} *for the associated limit operator.*

We refer the reader to the original TruthFinder paper [100] for the interpretation of ρ and γ. As described above, λ controls the amount to which conflicting claims play a role in the evaluation of a given claim. Of special interest is the case $\lambda = 0$,

in which the denominator in (2.1) is 1. Note that in (2.1) we have unfolded the definitions of [100] in order to obtain a single expression of $T'_N(c)$ in terms of the $T_N(s)$, at the expense of interpretability.

Let us return again to the network in Fig. 2.1. We take parameters $\rho = 0.5$ and $\gamma = 0.3$ (as per the experimental setup of Yin, Han, and Yu [100]) and $\lambda = 0.5$. Assuming that TruthFinder does indeed converge on this network – as it appears to do empirically – we have $T^{\mathrm{tf}}_N(s) \approx 0.5067$, $T^{\mathrm{tf}}_n(t) \approx 0.6590$ and $T^{\mathrm{tf}}_N(u) = T^{\mathrm{tf}}_N(v) = 0.7510$, which gives the ranking $s \sqsubset t \sqsubset u \simeq v$ on the sources. We have $T^{\mathrm{tf}}_N(c) \approx 0.5328$, $T^{\mathrm{tf}}_N(d) \approx 0.5670$, $T^{\mathrm{tf}}_N(e) \approx 0.4807$ and $T^{\mathrm{tf}}_N(f) \approx 0.7510$, which gives the ranking $e \prec c \prec d \prec f$ on the claims. Note that the source ranking coincides with that of Example 2.2.1, and the claim ranking refines that of Example 2.2.1 and Sums by ranking e *strictly* worse than c. Intuitively, this occurs because e has more sources reporting the conflicting claim (namely, f) than c does. If we instead take $\lambda = 0$, so that sources for conflicting claims are not considered, then the ranking reverts to $c \approx e \prec d \prec f$ (and the source ranking remains the same).

CRH. Standing for "Conflict Resolution on Heterogeneous Data", CRH is an optimisation-based framework for truth discovery [64]. It is again set in a more general setting, in which a metric d_o is available to measure the distance between values in D_o, for each object o. The optimisation problem jointly chooses weights for each source and a value for each object, such that the weighted sum of d_o-distances from each source's claim on o is minimised.

To express CRH in our framework we use the "probabilistic" encoding of categorical variables as described in [64, §2.4.1], where each categorical value is represented as a one-hot vector, and the source weight regularisation from [64, Eq. (4)]. We make a minor modification, however, by adding a small quantity ε to α_s and $T'_N(s)$ defined below; this ensures the logarithm in $T'_N(s)$ and the division in $T'_N(c)$ is well-defined and simplifies analysis of CRH later on.

Definition 2.2.6. *Given $\varepsilon > 0$,* CRH-ε *is the recursive scheme $(\mathcal{D}, T^0, U)$, where $\mathcal{D}$ is the set of operators with $0 \leq T_N(c) \leq 1$ for all N and $c \in C$,*

$$T^0_N(s) = 0, \qquad T^0_N(c) = \frac{|\operatorname{src}_N(c)|}{|S|}.$$

and $U(T) = T'$, where

$$T'_N(s) = \varepsilon - \log\left(\frac{\alpha_s}{\sum_{t \in S} \alpha_t}\right),$$
$$T'_N(c) = \frac{\sum_{s \in \operatorname{src}_N(c)} T'_N(s)}{\sum_{t \in S} T'_N(t)},$$

with

$$\alpha_s = \varepsilon + \sum_{c \in \operatorname{cl}_N(s)} \sum_{d \in \operatorname{cl}_N(\operatorname{obj}(c))} (T_N(d) - \mathbb{1}[d = c])^2.$$

The limit operator is denoted by $T^{\text{crh-}\varepsilon}$.[2]

[2] In the degenerate case $S = \emptyset$, we set $T_N \equiv 0$.

Table 2.1: Output rankings of the example operators on the network from Fig. 2.1.

Voting	$s \simeq t \simeq u \simeq v$	$c \approx d \approx e \prec f$
Weighted Agreement	$s \sqsubset t \sqsubset u \simeq v$	$c \approx e \prec d \prec f$
Sums	$s \sqsubset u \simeq v \sqsubset t$	$c \approx e \prec d \prec f$
TruthFinder	$s \sqsubset t \sqsubset u \simeq v$	$e \prec c \prec d \prec f$
TruthFinder ($\lambda = 0$)	$s \sqsubset t \sqsubset u \simeq v$	$c \approx e \prec d \prec f$
CRH-ε	$s \sqsubset t \sqsubset u \simeq v$	$c \approx e \prec d \prec f$

Note that in the case where each source provides a report on *all* objects – which is the setting in which CRH was originally introduced – we have $\sum_{c \in \mathrm{cl}_N(o)} T'_N(c) = 1$. Consequently, T'_N gives rise to a probability distribution over claims for each object o. The term of the sum in α_s corresponding to c is the squared Euclidean distance between this distribution and the distribution put forward by source s, which places all the probability mass in their report c.

In the network from Fig. 2.1 with $\varepsilon = 10^{-5}$, we have $T_N^{\text{crh-}\varepsilon}(s) \approx 0.2577$, $T_N^{\text{crh-}\varepsilon}(t) \approx 1.4827$ and $T_N^{\text{crh-}\varepsilon}(u) = T_N^{\text{crh-}\varepsilon}(v) \approx 9.3567$, giving the source ranking $s \sqsubset t \sqsubset u \simeq v$. Note that this is the same ranking on sources as T^{tf} gives. For claims, we have $T_N^{\text{crh-}\varepsilon}(c) = T_N^{\text{crh-}\varepsilon}(e) \approx 0.0126$, $T_N^{\text{crh-}\varepsilon}(d) \approx 0.0725$ and $T_N^{\text{crh-}\varepsilon}(f) \approx 0.9874$, giving the ranking $c \approx e \prec d \prec f$; this is the same as T^{sums}.

Table 2.1 summaries the source and claim rankings for each example operator on the network N from Fig. 2.1.

2.3 The Axioms

Having laid out the formal framework, we now introduce axioms for truth discovery. Such axioms are formal properties an operator may satisfy, which encode intuitively desirable behaviour. Many of our axioms are adaptations of axioms for various problem in social choice theory (e.g. from voting [105] and ranking systems [2]), in which the axiomatic method has seen great success. We also consider standard social choice axioms which are *not* desirable for truth discovery, to highlight the differences with classical problems such as voting. We will later revisit the example operators of the previous section to see to what extent our axioms hold in practise.

2.3.1 Coherence

The guiding principle of truth discovery is that claims backed by trustworthy sources should be believed, and sources making believable claims are trustworthy. All truth discovery methods aim to implement this principle to some extent, and the examples of Section 2.2 illustrate several different approaches.

We aim to formulate this principle axiomatically as a *coherency* property relating the source ranking $\sqsubseteq$ and the claim ranking $\preceq$: sources making higher $\preceq$-ranked claims should rank highly in $\sqsubseteq$, and vice versa. To do so we adapt the idea behind the *Transitivity* axiom of Altman and Tennenholtz [2] for ranking systems.

Now, a difficulty arises when considering how to compare the claims of two sources. For a simple example, suppose sources have either *low*, *medium* or *high* trustworthiness. How should we rank a claim c with one *medium* sources versus a

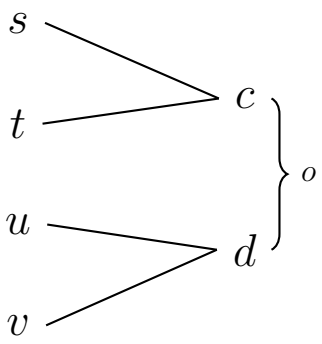


Figure 2.2: A network illustrating ***Claim-coherence****.*

claim d with a *low* and a *high* source? In some situations we may want to prioritise the number of claims, so that d is preferred. In others we may want to avoid trusting *low* sources as much as possible, so that c is preferred. The third option of ranking c and d equally believable is also reasonable.

To avoid these ambiguous cases, we focus on scenarios where there is an "obvious" ordering between two sets of claims (or sources). For example, consider the network depicted in Fig. 2.2. Suppose an operator gives a source ranking $s \sqsubset u \sqsubset t \sqsubset v$. Note that claims c and d have the same number of sources. Moreover, we can pair up these sources one-to-one such that the source for c is less trustworthy than the corresponding source for d: we have $s \sqsubset u$ and $t \sqsubset v$. On aggregate, we may reasonably say that $\mathrm{src}_N(c)$ is less trustworthy (with respect to $\sqsubseteq$) than $\mathrm{src}_N(d)$. We should therefore have $c \prec d$; any operator violating this has failed to realise the dependence between source trustworthiness and claim believability. Similarly, this reasoning can be applied to the set of claims from two sources.

This will form the basis of our first set of axioms. First, we formalise the above idea of a one-to-one correspondence respecting a ranking.

Definition 2.3.1. *If $\leq$ is a relation on a set X and $A, B \subseteq X$, then A* precedes *B* pairwise *with respect to $\leq$ if*

$$\exists f : A \to B \text{ bijective s.t. } \forall x \in A : \; x \leq f(x). \tag{2.3}$$

Say A strictly precedes *B if A precedes B but B does not precede A.*

If f satisfies the condition in (2.3), we say f *witnesses* the fact that A precedes B, and write $f : A \overset{\leq}{\Rightarrow} B$. Note that if $\leq$ is a preorder on X, the "precedes pairwise" relation is a preorder on 2^X. Indeed, it is reflexive (by considering the identity map $A \to A$, for each $A \subseteq X$) and transitive (if $f : A \overset{\leq}{\Rightarrow} B$ and $g : B \overset{\leq}{\Rightarrow} C$, then $g \circ f : A \overset{\leq}{\Rightarrow} C$). The strict pairwise order associated has a natural interpretation, as we now prove: there must exist some x in (2.3) for which the comparison is strict.

Proposition 2.3.1. *Suppose X is finite and $\leq$ is a total preorder on X. Then A strictly precedes B pairwise with respect to $\leq$ if and only if there is $f : A \overset{\leq}{\Rightarrow} B$ such that there is some $x_0 \in A$ with $x_0 < f(x_0)$.*

We need a preliminary lemma.

Lemma 2.3.1. *Suppose $\leq$ is a total preorder on a finite set X and $f : X \to X$ is an injective mapping such that $x \leq f(x)$ for all $x \in X$. Then $x \approx f(x)$ for all x.*

Proof. Take $x \in X$. Consider the sequence of iterates $(f^n(x))_{n\geq 1}$. Since this is an infinite sequence taking values in a finite set, there must be some point at which the sequence repeats, i.e. there are $n, k \geq 1$ such that $f^n(x) = f^{n+k}(x)$. Then $f(f^{n-1}(x)) = f(f^{n+k-1}(x))$, so injectivity gives $f^{n-1}(x) = f^{n+k-1}(x)$. Repeating this argument, we find $x = f^0(x) = f^k(x)$. By hypothesis, $f(x) \leq f^k(x)$, i.e. $f(x) \leq x$. Since $x \leq f(x)$ also, this gives $x \approx f(x)$ as required. □

Proof of Proposition 2.3.1. "if": Clearly A precedes B. Suppose for contradiction that this is not strict. Then there is some $g : B \overset{\leq}{\to} A$. Note that $g \circ f$ is a bijection $A \to A$, and for all $x \in X$ we have $x \leq f(x) \leq g(f(x))$. By Lemma 2.3.1, $x \approx g(f(x))$. In particular, we have $f(x_0) \leq g(f(x_0)) \approx x_0$, but this contradicts $x_0 < f(x_0)$.

"only if": Suppose A strictly precedes B. Then there is some $f : A \overset{\leq}{\to} B$. Note that f^{-1} is a bijection $B \to A$. Since B does not precede A, there must be some $y_0 \in B$ such that $y_0 \not\leq f^{-1}(y_0)$. By totality of $\leq$, we get $f^{-1}(y_0) < y_0$. Taking $x_0 = f^{-1}(y_0)$, we are done. □

We are now ready to state our first two axioms.

> **Claim-coherence**. If $\mathrm{src}_N(c)$ strictly precedes $\mathrm{src}_N(c')$ pairwise with respect to $\sqsubseteq^T_N$, then $c \prec^T_N c'$.
>
> **Source-coherence**. If $\mathrm{cl}_N(s)$ strictly precedes $\mathrm{cl}_N(s')$ pairwise with respect to $\preceq^T_N$, then $s \sqsubset^T_N s'$.

In words, **Claim-coherence** says that whenever we can pair up the sources for c and c' so that each source for c is less trustworthy than the corresponding source for c' (and *strictly* less, for at least one pair of sources), then c is strictly less believable than c'. Likewise, **Source-coherence** says that if the claims of s and s' can be paired up with the claims for s less believable than the claims for s', then s is strictly less trustworthy than s'.

Example 2.3.1. *Consider the network N from Fig. 2.1 again, and consider Sums. Recall that T^{sums} gives the source ranking $s \sqsubset u \simeq v \sqsubset t$, and claim ranking $c \approx e \prec d \prec f$.*

*Note that $\mathrm{src}_N(c) = \{s\}$ and $\mathrm{src}_N(d) = \{t\}$. Since $s \sqsubset t$, we have that $\{s\}$ strictly precedes $\{t\}$ with respect to $\sqsubseteq$. **Claim-coherence** therefore requires that $c \prec d$. Indeed, this does hold.*

*For **Source-coherence**, note that $\mathrm{cl}_N(s) = \{c, e\}$ and $\mathrm{cl}_N(t) = \{d, f\}$. Since $c \prec d$ and $e \prec f$, we see that $\mathrm{cl}_N(s)$ strictly precedes $\mathrm{cl}_N(t)$ with respect to $\preceq$. Accordingly, **Source-coherence** requires $s \sqsubset t$, which does hold.*

So, T^{sums} satisfies both coherence properties for this specific network. We will analyse T^{sums} and the other examples more generally in Section 2.5.

The reader may wonder why we only consider the *strict* pairwise relation in **Claim-coherence** (and **Source-coherence**). An alternative axiom might require that $c \preceq c'$ whenever $\mathrm{src}_N(s)$ precedes $\mathrm{src}_N(s')$ with respect to $\sqsubseteq$ (not necessarily

strictly). However, this property implies that $c \approx c'$ whenever $\mathrm{src}_N(c) = \mathrm{src}_N(c')$. We have already seen an example operator where this does not hold: TruthFinder ranks $e \prec c$ in the network N from Fig. 2.1, but $\mathrm{src}_N(c) = \mathrm{src}_N(e) = \{s\}$. Intuitively, c and e are "tied" when it come to the quality of their own sources, but there are fewer sources *disagreeing* with c (the "antisources") than e. Stating our coherence properties in the strict form permits an operator to consider antisources in cases where there is no clear comparison on the basis of sources alone.

Having said this, an operator with **Claim-coherence** is limited in the extent to which it can take antisources into account. We formulate an antisource version of coherence in Section 2.3.5, and show that it is incompatible with **Claim-coherence** when taken with some other basic axioms.

[**TODO:** Limitation: we can only compare sources/claims with the same number of claims/sources. Signpost if we end up improving this later by considering extra trustworthy sources/claims.]

2.3.2 Symmetry

A standard class of axioms in social choice theory express *symmetry properties*. In voting, for example, symmetry with respect to voters says that a voting rule should not care about the "names" of the voters: if voters i and j swap their ballots, the election result remains the same (this is called *anonymity* in the literature). Similarly, symmetry with respect to candidates says that if we re-label candidates, the outcome remains the same up to re-labelling (this is called *neutrality*). In general, symmetry requires that the output of some process depends only on *structural* features of the input, not the specific "names" of the entities involved.

For truth discovery, we can consider symmetry with respect to sources, objects and claims. The central concept is an *isomorphism* between networks.

Definition 2.3.2. *An* isomorphism *between networks N and N' is mapping $F : S \cup O \cup C \to S' \cup O' \cup C'$ such that*

1. *$F|_S$, $F|_O$ and $F|_C$ are bijections $S \to S'$, $O \to O'$ and $C \to C'$, respectively.*
2. *For all $s \in S$ and $c \in C$: $(s, c) \in R$ iff $(F(s), F(c)) \in R'$.*
3. *For all $c \in C$, $\mathrm{obj}(F(c)) = F(\mathrm{obj}(c))$.*

That is, F is a one-to-one correspondence between the sources, objects and claims of N and their N' counterparts, which respects the structure of the network. One can easily check that we also have $F(\mathrm{src}_N(c)) = \mathrm{src}_{N'}(F(c))$ and $F(\mathrm{cl}_N(s)) = \mathrm{cl}_{N'}(F(s))$. The symmetry axiom says an operator should not distinguish isomorphic networks.

> **Symmetry**. If F is an isomorphism between N and N', then $s \sqsubseteq^T_N s'$ iff $F(s) \sqsubseteq^T_{N'} F(s')$ and $c \preceq^T_N c'$ iff $F(c) \preceq^T_{N'} F(c')$.

We illustrate **Symmetry** with an example.

Example 2.3.2. *Consider the network N from Fig. 2.1 and N' from Fig. 2.3, where we take the sources, objects and domains to be the same in both networks. Then N and N' are isomorphic via the mapping F expressed in cycle notation*

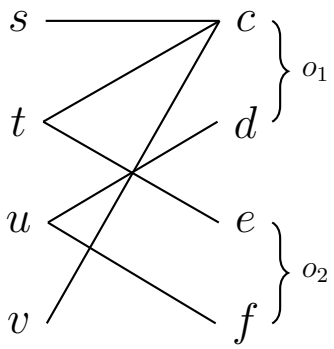


Figure 2.3: A network isomorphic to the one shown in Fig. 2.1.

as $(suv)(cf)(de)(o_1o_2)$*. For example,* s *plays the same role in* N *as* u *in* N'*,* c *plays the same role in* N *as* f *in* N'*, the role of objects* o_1 *and* o_2 *are swapped, etc.* ***Symmetry*** *requires that the source and claim rankings in* N' *are already determined by the rankings of* N*. For example, if the source ranking in* N *is* $s \sqsubset_N u \simeq_N v \sqsubset_N t$*, we must have* $u \sqsubset_{N'} v \simeq_{N'} s \sqsubset_{N'} t$*.*

An *automorphism* is an isomorphism F from a network N to itself. For example, F which swaps u and v in N from Fig. 2.1 is an automorphism, since u and v play exactly the same role in N. **Symmetry** implies that $u \simeq v$, and in fact this holds more generally.

Proposition 2.3.2. *If* F *is an automorphism on* N *and* T *satisfies* ***Symmetry****, then* $s \simeq_N^T F(s)$ *and* $c \approx_N^T F(c)$*, for all* $s \in S$ *and* $c \in C$*.*

Proof. We show $s \simeq_N^T F(s)$ for all sources s; the result for claims is similar. Take $s \in S$. Since S is finite and F restricts to a bijection $S \to S$, an argument identical to the one in the proof of Lemma 2.3.1 shows there is some $k \geq 1$ such that $s = F^k(s)$.

First suppose $s \sqsubseteq_N^T F(s)$. By **Symmetry** we may apply F to both sides; doing so repeatedly yields $F^n(s) \sqsubseteq_N^T F^{n+1}(s)$ for all $n \geq 1$. By transitivity of $\sqsubseteq_N^T$, we get $F(s) \sqsubseteq_N^T F^n(s)$. Taking $n = k$ gives $F(s) \sqsubseteq_N^T F^k(s) = s$, so $s \simeq_N^T F(s)$.

Now suppose $F(s) \sqsubseteq_N^T s$. By an identical argument, $F^n(s) \sqsubseteq_N^T F(s)$ for all $n \geq 1$; taking $n = k$ gives $s \sqsubseteq_N^T F(s)$, so $s \simeq_N^T F(s)$ again.

Since $\sqsubseteq_N^T$ is total these cases are exhaustive, and we are done. □

Proposition 2.3.2 is useful for showing certain sources and claims must rank equally. For example, take the network N from Fig. 2.2. Intuitively this network displays internal symmetry within the sources for each claim and between the claims themselves. Indeed, the functions $F = (st)(uv)$ and $G = (su)(tv)(cd)$ are automorphisms. By Proposition 2.3.2, any operator T satisfying **Symmetry** must output flat rankings $s \simeq t \simeq u \simeq v$ and $c \approx d$.

2.3.3 Monotonicity

Given that voting is not a viable truth discovery method, the believability of a claim c should not increase monotonically with $|\mathrm{src}_N(c)|$. Moreover, it should not increase with the *set* of sources $\mathrm{src}_N(c)$, ordered by set inclusion: $\mathrm{src}_N(c) \subseteq \mathrm{src}_N(d)$

should not in general imply $c \preceq d$. Indeed, consider an adversarial source t deliberately making false claims, and suppose $\mathrm{src}_N(c) = \{s\}$ and $\mathrm{src}_N(d) = \{s, t\}$. Then $\mathrm{src}_N(c) \subseteq \mathrm{src}_N(d)$, but the extra support from t should actually *decrease* the believability of d – since t only provides false claims – not increase it.

Nevertheless, there is a sense in which – all else being equal – a claim with more sources is more believable. The above examples show that some subtlety is needed in formulating this as a general principle, and that trust should be taken into account in doing so.

In this section we consider monotonicity properties of two kinds: monotonicity *within* a network, and monotonicity *between* networks as more reports are added. We start with the latter by adapting the idea of *positive responsiveness* from social choice theory.

Responsiveness. In the context of voting, positive responsiveness requires that if a voter switches their vote from candidate B to a winning candidate A, then A becomes the unique winner [105]. A naive version of positive responsiveness for truth discovery says that if we change a network N by adding a new report (s, c) – possibly removing reports from s conflicting with c – then c should move strictly up in the claim ranking. Clearly this neglects to consider the trustworthiness of s, and is thus an undesirable property (e.g. consider s adversarial as described above). Our first monotonicity axiom weakens this naive property by only considering "fresh" sources s not providing any reports in the original network N. Intuitively, we have no reason to believe such sources are untrustworthy, and they should therefore have a positive effect when making a claim. In what follows, when $\mathrm{cl}_N(s) = \emptyset$ we write $N + (s, c)$ for the network $(S, O, D, R \cup \{(s, c)\})$.

> **Fresh-pos-resp**. Suppose $\mathrm{cl}_N(s) = \emptyset$. Then for all $c \in C$ and $d \in C \setminus \{c\}$, $d \preceq^T_N c$ implies $d \prec^T_{N+(s,c)} c$.

That is, if c was already at least as believable as d, then a fresh report makes c *strictly* more believable in the new network.[3] What about the effects of a fresh report for c on source trustworthiness? According to the mutual dependence between the source and claim rankings – captured in a static network via the coherence properties – sources already claiming c should become more trusted, whereas those claiming a conflicting claim d should become less trusted.

> **Source-pos-resp**. Suppose $s \in \mathrm{antisrc}_N(c)$, $t \in \mathrm{src}_N(c)$, and $\mathrm{cl}_N(u) = \emptyset$. Then $s \sqsubseteq^T_N t$ implies $s \sqsubset^T_{N+(u,c)} t$.

Note that **Source-pos-resp** does not say anything about the ranking of the fresh source u. We consider another example.

Example 2.3.3. *Fig. 2.4 illustrates* ***Fresh-pos-resp*** *and* ***Source-pos-resp****. Let N_0 denote the network including only the solid edges, $N_1 = N_0 + (u, f)$, and $N_2 = N_1 + (v, f)$. Note that N_2 is our running example network from Fig. 2.1. Assuming*

[3] Note that N and $N + (s, c)$ share the same set of objects O and domains D, so the set of possible claims in both networks are the same. Consequently we are justified in treating c and d as claims in both networks.

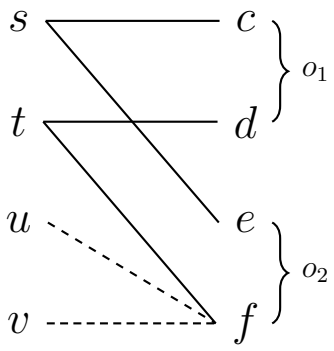


*Figure 2.4: Networks N_0 (solid edges only), $N_1 = N_0 + (u, f)$ and $N_2 = N_1 + (v, f)$ illustrating **Fresh-pos-resp** and **Source-pos-resp**.*

***Symmetry**, everything is tied in N_0: we have $s \simeq_{N_0} t$ and $c \approx_{N_0} d \approx_{N_0} \approx_{N_0} e \approx_{N_0} f$. Since N_1 is the result of adding the report (u, f) and u makes no claims in N_0, **Fresh-pos-resp** gives $e \prec_{N_1} f$. Since $s \in \mathrm{src}_{N_0}(e) \subseteq \mathrm{antisrc}_{N_0}(f)$ and $t \in \mathrm{src}_{N_0}(f)$, **Source-pos-resp** gives $s \sqsubset_{N_1} t$. Going from N_1 to N_2 we can repeat exactly the same arguments to find $e \prec_{N_2} f$ and $s \sqsubset_{N_2} t$.*

*Bringing **Claim-coherence** in too, $s \sqsubset_{N_2} t$ gives $c \prec_{N_2} d$. Thus, **Claim-coherence**, **Symmetry**, **Fresh-pos-resp** and **Source-pos-resp** are enough to capture our intuitions about this network as described in the introduction [**TODO:** check intro.]*

In the special case where a network contains reports only for a single object, the responsiveness properties and **Symmetry** actually force an operator to rank claims by voting, and to rank sources by the vote count of their claims. Note that each source provides at most one report in this case, by condition (2) in the definition of a network. Consequently there is little structure in such networks, as we cannot look at how sources interact over multiple objects to determine trustworthiness. We therefore argue that voting is reasonable behaviour in this special case.

Proposition 2.3.3. *Suppose there is $o \in O$ such that $\mathrm{src}_N(o') = \emptyset$ for all $o \neq o'$. Then*

1. *If T satisfies **Symmetry** and **Fresh-pos-resp**, then for all $c, d \in \mathrm{cl}_N(o)$:*

$$c \preceq_N^T d \iff |\mathrm{src}_N(c)| \leq |\mathrm{src}_N(d)|.$$

2. *If T satisfies **Symmetry** and **Source-pos-resp**, then for all $s, t \in S$ with $\mathrm{cl}_N(s), \mathrm{cl}_N(t) \neq \emptyset$,*

$$s \sqsubseteq_N^T t \iff |\mathrm{src}_N(c_s)| \leq |\mathrm{src}_N(c_t)|,$$

where c_s and c_t are the unique claims reported by s and t respectively.

While Proposition 2.3.3 only addresses a somewhat trivial case, it will turn out to be useful in characterising voting behaviour more generally in Sections 2.3.4 and 2.3.6. It can be seen as one of the many generalisations of *May's Theorem* [66], which characterises the majority voting rule in two-candidate elections. To prove it, we need a preliminary result.

Lemma 2.3.2. *Suppose* $|\operatorname{src}_N(c)| = |\operatorname{src}_N(d)|$, $\operatorname{obj}(c) = \operatorname{obj}(d)$, *and for all* $s \in \operatorname{src}_N(c) \cup \operatorname{src}_N(d)$, $|\operatorname{cl}_N(s)| = 1$. *Then for any operator* T *satisfying* ***Symmetry****,* $c \approx_N^T d$.

Proof. Without loss of generality, assume $c \neq d$. Since $\operatorname{obj}(c) = \operatorname{obj}(d)$, we have $c \in \operatorname{conflict}_N(d)$ and thus $\operatorname{src}_N(c) \cap \operatorname{src}_N(d) = \emptyset$. Since $|\operatorname{src}_N(c)| = |\operatorname{src}_N(d)|$ there exists a bijection $\hat{\varphi} : \operatorname{src}_N(c) \to \operatorname{src}_N(d)$. We extend this to a bijection $\varphi : S \to S$ by

$$\varphi(s) = \begin{cases} \hat{\varphi}(s), & s \in \operatorname{src}_N(c) \\ \hat{\varphi}^{-1}(s), & s \in \operatorname{src}_N(d) \\ s, & \text{otherwise.} \end{cases}$$

Now let $F : S \cup C \cup O \to S \cup C \cup O$ be defined by $F|_S = \varphi$, $F|_C = (cd)$ and $f|_O = \mathrm{id}$. That is, F permutes sources according to φ, swaps claims c and d, and leaves objects as they are. Since $F(c) = d$, to show $c \approx_N^T d$ it is sufficient by Proposition 2.3.2 to show that F is an automorphism on N.

It is easily seen that the restrictions of F to S, C and O respectively, are bijective. Moreover, we have $\operatorname{obj}(F(e)) = F(\operatorname{obj}(e))$ for all claims e since $F(o) = o$ and $\operatorname{obj}(c) = \operatorname{obj}(d)$. It remains to show that $(s, e) \in R$ iff $(F(s), F(e)) \in R$.

For the left-to-right direction, suppose $(s, e) \in R$. First suppose $s \in \operatorname{src}_N(c)$. Then $F(s) = \hat{\varphi}(s) \in \operatorname{src}_N(d)$, so $(F(s), d) \in R$. By assumption we have $|\operatorname{cl}_N(s)| = 1$, so in fact c is the unique claim reported by s. Thus $e = c$. Consequently

$$(F(s), F(e)) = (F(s), d) \in R$$

as required. The case for $s \in \operatorname{src}_N(d)$ follows by a near-identical argument. Finally, if $s \notin \operatorname{src}_N(c) \cup \operatorname{src}_N(d)$ then $F(s) = s$ and $e \notin \{c, d\}$, so $F(e) = e$. Thus $(F(s), F(e)) = (s, e) \in R$.

For the right-to-left direction, suppose $(F(s), F(e)) \in R$. Applying the argument above we have $(F^2(s), F^2(e)) \in R$ also. But note that $F = F^{-1}$, so $F^2 = \mathrm{id}$. Hence $(s, e) \in R$, as required. This completes the proof. □

Proof of Proposition 2.3.3. We prove (1) only, since (2) can be shown using essentially the same argument with **Source-pos-resp** taking the place of **Fresh-pos-resp**.

Suppose T satisfies **Symmetry** and **Fresh-pos-resp**, and take N as stated in Proposition 2.3.3. It is sufficient to show that, for all $c, d \in \operatorname{cl}_N(o)$,

$$|\operatorname{src}_N(c)| \leq |\operatorname{src}_N(d)| \implies c \preceq_N^T d \tag{2.4}$$

$$|\operatorname{src}_N(c)| < |\operatorname{src}_N(d)| \implies c \prec_N^T d. \tag{2.5}$$

First we show (2.4). Suppose $|\operatorname{src}_N(c)| \leq |\operatorname{src}_N(d)|$. Assume without loss of generality that $c \neq d$. Write $k = |\operatorname{src}_N(d)| - |\operatorname{src}_N(c)| \geq 0$. Let $X = \{s_1, \ldots, s_k\}$ be an arbitrary subset of $\operatorname{src}_N(d)$ of size k. Let N_0 denote the network in which all claims from sources in X are removed. Note that since N does not contain reports for objects other than o, by the consistency property (2) in Definition 2.1.1 we have that sources in X *only* report d. We construct networks $N_1, \ldots, N_k$ in which these claims are added back in: for $0 \leq i \leq k - 1$, set

$$N_{i+1} = N_i + (s_{i+1}, d).$$

Then N_k is just the original network N. Note that $\mathrm{cl}_{N_i}(s_j) = \emptyset$ for $j > i$. Next we show by induction that for all $0 \leq i \leq k$,

$$c \preceq^T_{N_i} d, \text{ and if } i > 0 \text{ then } c \prec^T_{N_i} d. \tag{2.6}$$

For the base case $i = 0$, note that since only reports for d were removed in constructing N_0, we have $\mathrm{src}_{N_0}(c) = \mathrm{src}_N(c)$. Consequently,

$$|\mathrm{src}_{N_0}(d)| = |\mathrm{src}_N(d) \setminus X| = |\mathrm{src}_N(d)| - k = |\mathrm{src}_N(c)| = |\mathrm{src}_{N_0}(c)|.$$

Note also that $\mathrm{obj}(c) = \mathrm{obj}(d)$ – since by assumption $c, d \in \mathrm{cl}_N(o)$ – and for $s \in \mathrm{src}_{N_0}(c) \cup \mathrm{src}_{N_0}(d)$ we have $|\mathrm{cl}_{N_0}(s)| = 1$ since N_0 also only contains reports for o. The hypothesis of Lemma 2.3.2 are satisfied, so we have $c \approx^T_{N_0} d$. In particular, $c \preceq^T_{N_0} d$ as required.

Now for the inductive step, suppose (2.6) holds for i. Since $\mathrm{cl}_{N_i}(s_{i+1}) = \emptyset$, **Fresh-pos-resp** and the inductive hypothesis give $c \prec^T_{N_{i+1}} d$, as required.

Finally, (2.4) follows by taking $i = k$ in (2.6), recalling that $N = N_k$. Moreover, (2.5) follows by exactly the same argument, noting that when $|\mathrm{src}_N(c)| < |\mathrm{src}_N(d)|$ we have $k > 0$, so $c \prec^T_{N_k} d$ by (2.6) again. □

Trust-based monotonicity. Suppose $\mathrm{src}_N(d) = \mathrm{src}_N(c) \cup \{s\}$. The relative ranking of c and d depends on the marginal effect of s: if s is "trustworthy" then d gains credibility from the extra support of s, whereas s is "untrustworthy" this extra support has the opposite effect. Our next axiom requires that such marginal effects are compatible with the source trustworthiness ranking. First, some terminology is required.

Definition 2.3.3. *Given a network N, a source $s \in S$ is* marginally trustworthy *with respect to an operator T if there exist claims $c, d \in C$ such that $s \notin \mathrm{src}_N(c)$, $\mathrm{src}_N(d) = \mathrm{src}_N(c) \cup \{s\}$ and $c \preceq^T_N d$. Similarly, s is* marginally untrustworthy *if there are $c, d \in C$ such that $s \notin \mathrm{src}_N(c)$, $\mathrm{src}_N(d) = \mathrm{src}_N(c) \cup \{s\}$ and $d \preceq^T_N c$.*

These properties express something about the trustworthiness of sources via the *claim* ranking $\preceq^T_N$, akin to how **Source-coherence** looks at trustworthiness via the claims reported by a source. Note that it is possible for a source to be both marginally trustworthy and untrustworthy. Naturally, marginally untrustworthy sources should rank lower than marginally trustworthy ones.

> **Marginal-trustworthiness**. If s is marginally untrustworthy and t is marginally trustworthy, then $s \sqsubset^T_N t$.

Equipped with a notion of marginal trustworthiness, we can also state a trust-aware monotonicity axiom for claims.

> **Trust-based-monotonicity**. Suppose $\mathrm{src}_N(d) = \mathrm{src}_N(c) \cup Z$, where $\mathrm{src}_N(c) \cap Z = \emptyset$. Then
>
> 1. If each $s \in Z$ is marginally trustworthy, $c \preceq^T_N d$.
> 2. If each $s \in Z$ is marginally untrustworthy, $d \preceq^T_N c$.

Informally, **Trust-based-monotonicity** says that if each $s \in Z$ has a positive (or at least, not negative) impact on some claim in N, as measured by $\preceq^T_N$, then the sources in Z acting collectively should also have a positive impact. Also note that in the case $Z = \{s\}$, **Trust-based-monotonicity** implies that the marginal impact of s is consistent across the network.

[**TODO:** Example of these postulates? Are they interesting?]

2.3.4 Independence

Another common class of axioms in social choice theory are *independence* axioms, which require that some aspect of the output is independent of "irrelevant" parts of the input. The original example is Arrow's *Independence of Irrelevant Alternatives* (IIA) in voting theory [4], which says, roughly speaking, that the ranking of candidates A and B should depend only on the individual rankings of A and B, not on any "irrelevant" alternative C. It has been adapted to several settings in which the axiomatic method has been applied. Perhaps closest to our setting is judgment aggregation, where independence requires the collective acceptance of a report φ does not depend on how the individuals accept or reject some other report ψ [34].

A version of IIA can be easily stated in our framework: the ranking of claims c and d should depend only on the sources reporting c and d, not on the sources for other claims. However, this axiom is clearly *undesirable* for truth discovery. Indeed, consider again the network N from Fig. 2.1. As we have argued informally, claim c is intuitively weaker than d because how of their respective sources interact with other claims in the network. Nevertheless, we state this axiom as a point of comparison with classical social choice problems such as voting.

> **Classical-independence**. Suppose $C_N = C_{N'}$. Then $\mathrm{src}_N(c) = \mathrm{src}_{N'}(c)$ and $\mathrm{src}_N(d) = \mathrm{src}_{N'}(d)$ implies $c \preceq^T_N d$ iff $c \preceq^T_{N'} d$.

That is, if c and d have the same sources in N and N', they have the same relative ranking in both networks. The undesirability of **Classical-independence** can be formalised axiomatically: together with our earlier axioms, it implies voting-like behaviour within the claims for each object.[4] Note that for the special case of binary networks, similar results have been shown in the literature on binary aggregation with abstentions [18].

Proposition 2.3.4. *Suppose T satisfies **Symmetry**, **Fresh-pos-resp** and **Classical-independence**. Then for all $o \in O$ and $c, d \in \mathrm{cl}_N(o)$,*

$$c \preceq^T_N d \iff |\mathrm{src}_N(c)| \leq |\mathrm{src}_N(d)|.$$

Proof. Take $c, d \in \mathrm{cl}_N(o)$. Let the network N' have the same sources, objects and domains as N, but with reports $R' = R \cap (S \times \{c, d\})$. That is, N' discards all reports for claims other than c and d. Then we have $\mathrm{src}_{N'}(c) = \mathrm{src}_N(c)$, $\mathrm{src}_{N'}(d) = \mathrm{src}_N(d)$, and $\mathrm{src}_{N'}(e) = \emptyset$ for all $e \notin \{c, d\}$. By **Classical-independence**, $c \preceq^T_N d$ iff $c \preceq^T_{N'} d$.

Now, note that since $c, d \in \mathrm{cl}_N(o)$, for $o' \neq o$ and $e \in \mathrm{cl}_N(o')$ we have $e \notin \{c, d\}$, so $\mathrm{src}_N(e) = \emptyset$. Hence $\mathrm{src}_N(o') = \emptyset$ for such o'. Since T satisfies **Symmetry** and

[4] We give a further axiom which implies voting behaviour for claims of *different* objects – and leads to a complete characterisation of voting – in Section 2.3.6.

Fresh-pos-resp, we may apply Proposition 2.3.3 (1) to find $c \preceq^T_{N'} d$ iff $|\mathrm{src}_{N'}(c)| \leq |\mathrm{src}_{N'}(d)|$. But $|\mathrm{src}_{N'}(c)| = |\mathrm{src}_N(c)|$, and likewise for d. Consequently

$$c \preceq^T_N d \iff c \preceq^T_{N'} d \iff |\mathrm{src}_{N'}(c)| \leq |\mathrm{src}_{N'}(d)| \iff |\mathrm{src}_N(c)| \leq |\mathrm{src}_N(d)|$$

as desired. □

While this result appears similar to Proposition 2.3.3, the crucial difference is that we no longer restrict to the case sources only report on a single object, where voting is justified. This is the (overly strong) role **Classical-independence** plays: it allows the complexity of a multi-object network to be reduced to a single-object network, where the ranking trivialises.

Recalling from Example 2.3.3 that **Claim-coherence**, **Symmetry**, **Fresh-pos-resp** and **Source-pos-resp** are enough to ensure $c \prec d$ in our running example network from Fig. 2.1 (whereas per-object voting gives $c \approx d$), we obtain an impossibility result with **Classical-independence**. In fact we obtain *two* impossibility results, since **Source-pos-resp** can also be replaced with **Source-coherence**.

Theorem 2.3.1. *Suppose and operator satisfies* ***Symmetry***, ***Claim-coherence*** *and* ***Fresh-pos-resp****. Then the following axioms cannot hold simultaneously.*

1. ***Source-pos-resp*** *and* ***Classical-independence****.*
2. ***Source-coherence*** *and* ***Classical-independence****.*

[**TODO:** Figure out if these impossibilities are minimal.]

Proof.

1. The impossibility of these axioms holding together follows from Example 2.3.3 and Proposition 2.3.4, as described above.
2. Let N be as shown in Fig. 2.1. Suppose some operator T satisfies the stated axioms. From Proposition 2.3.4 we get $c \approx^T_N d$ and $e \prec^T_N f$. Considering sources s and t, **Source-coherence** gives $s \sqsubset^T_N t$. But now **Claim-coherence** gives $c \prec^T_N d$: contradiction.

□

By only looking at a claim's sources, **Classical-independence** ignores the indirect interaction with other sources and claims in the network. Our next axiom accounts for such interactions by considering networks with *disjoint sub-networks*, such as the one shown in Fig. 2.5. Intuitively, while the sources and claims within a sub-network may interact in complex ways, the fact that the sub-networks have no sources or objects in common means there is no interaction *between* them. Accordingly, the ranking for one should not depend on the other. We formalise this by considering unions of *disjoint networks*.[5]

[5]Note that it is possible to define the disjoint union of an arbitrary collection of (not necessarily disjoint) networks in a manner similar to the disjoint union of a collection of sets $\bigsqcup_{i \in I} X_i$, but we do not need this generality here.

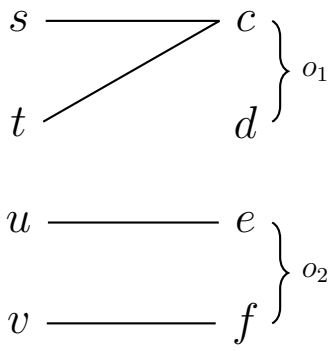


Figure 2.5: A network illustrating ***Disjoint-independence****.*

Definition 2.3.4. *Networks N and N' are* disjoint *if $S \cap S' = \emptyset$ and $O \cap O' = \emptyset$. For N, N' disjoint, their* union *is the network $N \sqcup N' = (S \cup S', O \cup O', \hat{D}, R \cup R')$, where $\hat{D}_o = D_o$ for $o \in O$, and $\hat{D}_o = D'_o$ for $o \in O'$.*

Note that if N and N' are disjoint, it follows that $C \cap C' = \emptyset$ also. The following axiom says that the ranking of sources and claims is unaffected by the addition of a disjoint network.

> **Disjoint-independence**. If N and N' are disjoint, $s, t \in S$, and $c, d \in C$, then $s \sqsubseteq^T_N t$ iff $s \sqsubseteq^T_{N \sqcup N'} t$ and $c \preceq^T_N d$ iff $c \preceq^T_{N \sqcup N'} d$.

[**TODO:** If bothered, explain graph-theoretic interpretation in terms of connected components.]

2.3.5 Conflicting claims

Our axioms so far have not made use of the conflict relation between claims. Intuitively, distinct claims c, c' for the same object o cannot both be true, so belief in c should come at the expense of belief in c'. Similarly, if the antisources of c – that is, the sources who report claims conflicting with c – are seen as less trustworthy than the antisources of c', then the attack on c is less damaging than that of c', so c should be more believable than c'. Note that these are again coherence principles, which constrain how the claim ranking $\preceq$ coheres with both the source ranking $\sqsubseteq$ and the conflict relation. We formulate them as axioms.

> **Conflict-coherence**. If $\text{conflict}_N(c)$ strictly precedes $\text{conflict}_N(c')$ pairwise with respect to $\preceq^T_N$, then $c' \prec^T_N c$.
>
> **Anti-coherence**. If $\text{antisrc}_N(c)$ strictly precedes $\text{antisrc}_N(c')$ pairwise with respect to $\sqsubseteq^T_N$, then $c' \prec^T_N c$.

While both **Conflict-coherence** and **Anti-coherence** appear reasonable in isolation, there is an inherent tension between them and our earlier coherence axioms. Together with symmetry and responsiveness axioms, we have an impossibility result.

Theorem 2.3.2. *Suppose an operator satisfies* ***Symmetry*** *and* ***Claim-coherence****. Then the following axioms cannot hold simultaneously.*

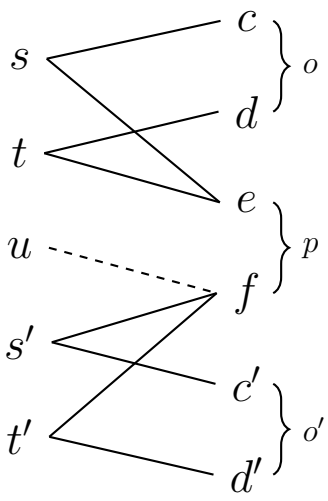


Figure 2.6: Network used to illustrate the impossibility results of Theorem 2.3.2.

1. ***Fresh-pos-resp**, **Source-coherence** and **Conflict-coherence**,*
2. ***Source-pos-resp** and **Conflict-coherence**.*
3. ***Source-pos-resp** and **Anti-coherence**.*

Proof. Suppose T satisfies **Symmetry** and **Claim-coherence**. Throughout the proof, let N_0 denote the network shown in Fig. 2.6 excluding the dashed edge, and let $N_1 = N + (u, f)$ denote the network including the dashed edge. We first note some consequences of the axioms in both networks. In N_0, the mapping $(s\ s')(t\ t')(c\ c')(d\ d')(o\ o')(e\ f)$ is an automorphism, so we have $s \simeq^T_{N_0} s'$ and $e \approx^T_{N_0} f$. Note that $\mathrm{src}_{N_0}(u) = \emptyset$, $s \in \mathrm{antisrc}_{N_0}(f)$ and $s' \in \mathrm{src}_{N_0}(f)$. If T additionally satisfies **Fresh-pos-resp**, we get $e \prec^T_{N_1} f$. If T instead satisfies **Source-pos-resp**, we get $s \sqsubset^T_{N_1} s'$. Considering N_1 alone, the mapping $(s\ t)(s'\ t')(c\ d)(c'\ d')$ is an automorphism, so **Symmetry** gives $c \approx^T_{N_1} d$ and $c' \approx^T_{N_1} d'$.

1. Suppose T also satisfies **Fresh-pos-resp**, **Source-coherence** and **Conflict-coherence**. First we claim $c \approx^T_{N_1} c'$. Indeed, suppose not. If $c' \prec^T_{N_1} c$, we may note that $\mathrm{conflict}_{N_1}(d) = \{c\}$ and $\mathrm{conflict}_{N_1}(d') = \{c'\}$, and apply **Conflict-coherence** to get $d \prec^T_{N_1} d'$. But by **Symmetry** as above, we have $c \approx^T_{N_1} d$ and $c' \approx^T_{N_1} d'$. Consequently $c \approx^T_{N_1} d \prec^T_{N_1} d' \approx^T_{N_1} c'$, i.e. $c \prec^T_{N_1} c'$. Clearly this contradicts $c' \prec^T_{N_1} c$. If $c \prec^T_{N_1} c'$ we obtain a contradiction by an identical argument. Hence $c \approx^T_{N_1} c'$.

   Now, by **Fresh-pos-resp** and **Symmetry** as noted above, we have $e \prec^T_{N_1} f$. **Source-coherence** for s and s' therefore gives $s \sqsubset^T_{N_1} s'$. But considering c and c', **Claim-coherence** gives $c \prec^T_{N_1} c'$. This contradicts $c \approx^T_{N_1} c'$, and we are done.

2. Suppose T additionally satisfies **Source-pos-resp** and **Conflict-coherence**. By the same argument as above, **Conflict-coherence** and **Symmetry** to-

gether dictate that $c \approx^T_{N_1} c'$. But by **Symmetry** and **Source-pos-resp**, we have $s \sqsubset^T_{N_1} s'$. **Claim-coherence** then implies $c \prec^T_{N_1} c'$: contradiction.

3. Suppose T additionally satisfies **Source-pos-resp** and **Anti-coherence**. Again, $s \sqsubset^T_{N_1} s'$. **Claim-coherence** implies $c \prec^T_{N_1} c'$. Since $\text{antisrc}_{N_1}(d) = \{s\}$ and $\text{antisrc}_{N_1}(d') = \{s'\}$, **Anti-coherence** gives $d' \prec^T_{N_1} d$. But recall that, by **Symmetry**, $c \approx^T_{N_1} d$ and $c' \approx^T_{N_1} d'$. Hence $c \prec^T_{N_1} c' \approx^T_{N_1} d' \prec^T_{N_1} d \approx^T_{N_1} c$, i.e. $c \prec^T_{N_1} c$: contradiction.

□

Note that all four coherence axioms can be satisfied at the same time, e.g. by the trivial operator which outputs constant scores $T_N(s) = T_N(c) = 0$. Of course, this operator violates both **Fresh-pos-resp** and **Source-pos-resp**.

2.3.6 Axiomatic Characterisation of Voting

Recall from Proposition 2.3.4 that **Symmetry**, **Fresh-pos-resp** and **Classical-independence** force an operator to rank claims for the object simply by their number of sources, as in voting from Section 2.2.1. In this section we give two further axioms which force this ranking even for claims across different objects, and thus characterise T^{vote} completely. Like **Classical-independence**, these axioms are *not* desirable properties, and are introduced only to capture the behaviour of voting. The first axiom simply says that the source ranking is flat.

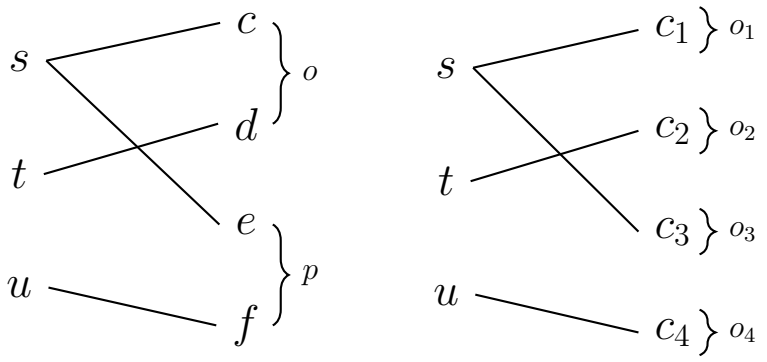
Flat-sources. For all $s, s' \in S$, $s \simeq^T_N s'$.

The second axiom says that objects play no role: it is only the relation between sources and claims which affects the rankings. That is, we can ignore the conflict relation between claims. To define the axiom we introduce a notion of "reducing" the objects of a network.

Definition 2.3.5. *A network N' is an* object reduction *of N via $f : C_N \to C_{N'}$ if*

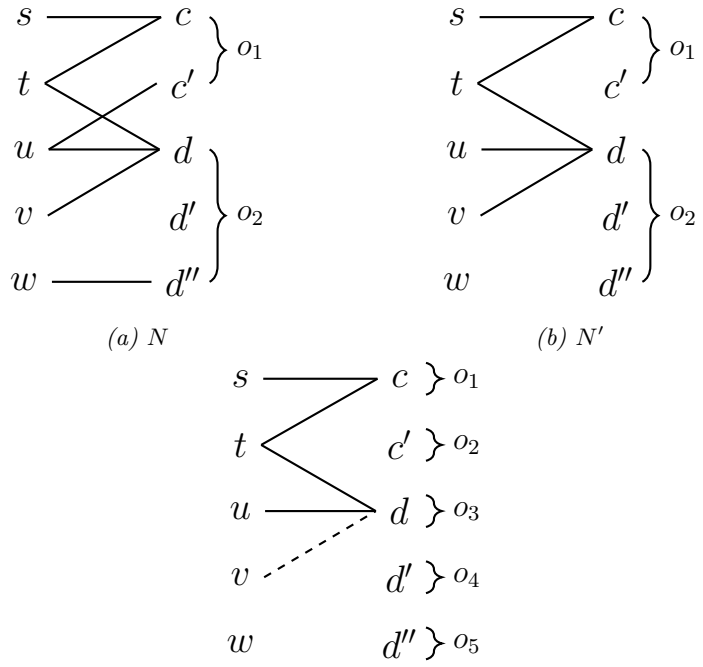
1. *$S' = S$.*
2. *f is a bijection $C_N \to C_{N'}$ such that $(s, c) \in R$ iff $(s, f(c)) \in R'$.*
3. *For al $o \in O'$, $|D'_o| = 1$.*

Note that every network N has an object reduction since the set of possible objects $\mathbb{O}$ is infinite; we may take O' to be any subset of $\mathbb{O}$ of size $|C_N|$, take $D'_o = \{v\}$ for some fixed $v \in \mathbb{V}$, and set R' accordingly. Fig. 2.7 shows an example of an object reduction. Note that the network N' has only a single claim for each object, and the structure of the reports – i.e. the edges shown in Fig. 2.7 – is the same in N and N'. Going from N to N' loses information about which claims conflict with one another, and our axioms in Section 2.3.5 explicitly require that this information *does* affects the rankings. Voting does not use this information, however, which leads to the following axiom.



(a) N *(b) N'*

Figure 2.7: Illustration of an object reduction of a network.



(a) N *(b) N'*

(c) N'' (all edges) and N_0 (excluding dashed edge)

*Figure 2.8: Illustration of the proof of Theorem 2.3.3. In N', reports for claims other than c and d are removed. N'' is an object reduction of N'. The dashed edge shows the reports added when **Fresh-pos-resp** is applied.*

Object-irrelevance. If N' is an object reduction of N via f, then $c \preceq_N^T d$ iff $f(c) \preceq_N^T f(d)$.

Note that **Object-irrelevance** is similar in form to **Symmetry**, but rather than requiring rankings are invariant under isomorphisms – which preserve the relevant structure of a network – it requires rankings are invariant under object reductions.

We can now characterise voting, up to ranking equivalence.

Theorem 2.3.3. *An operator T satisfies **Symmetry**, **Fresh-pos-resp**, **Classical-independence**, **Flat-sources** and **Object-irrelevance** if and only if $T \sim T^{\text{vote}}$.*

Proof (sketch). The "if" direction is straightforward. [**TODO:** Worth sketching?]. For the "only if" direction, take an operator T with the stated axioms. **Flat-sources** immediately implies $\sqsubseteq_N^T = \sqsubseteq_N^{T^{\text{vote}}}$ for all networks N. For the claim rankings, we take a similar approach to the proof of Proposition 2.3.4 and only sketch the argument here. An illustration of the proof is shown in Fig. 2.8.

Take any network N and claims c, d. We first remove all reports for other claims to produce N'; this preserves rankings by **Classical-independence**. Taking N'' to be any object reduction of N', we ensure c and d are the only claims for their respective objects,[6] and rankings are again preserved by **Object-irrelevance**. As before, it suffices to show that $|\operatorname{src}_N(c)| \leq |\operatorname{src}_N(d)|$ implies $c \preceq_N^T d$ and $|\operatorname{src}_N(c)| < |\operatorname{src}_N(d)|$ implies $c \prec_N^T d$, since c and d are arbitrary.

Write $k = |\operatorname{src}_N(d)| - |\operatorname{src}_N(c)| \geq 0$. Choosing k sources from $\operatorname{src}_N(d) \setminus \operatorname{src}_N(c)$, let N_0 be the network obtained from N'' in which reports for d from these sources are removed. Note that such sources *only* report d, since reports for other claims were removed in the construction of N'. Then $|\operatorname{src}_{N_0}(c)| = |\operatorname{src}_{N_0}(d)|$. The fact that $|D''_{\text{obj}(c)}| = |D''_{\text{obj}(d)}| = 1$ ensures we are able to choose an automorphism on N_0 which swaps c and d (and swaps $\operatorname{src}_{N_0}(c) \setminus \operatorname{src}_{N_0}(d)$ with $\operatorname{src}_{N_0}(d) \setminus \operatorname{src}_{N_0}(c)$). By **Symmetry**, $c \approx_{N_0}^T d$.

If $k = 0$ then $N_0 = N''$, and we are done. Otherwise, by repeated applications of **Fresh-pos-resp** we may add the removed reports back in to N_0 to get $c \prec_{N''}^T d$. Since claim rankings are the same in N'' as in N, this completes the proof. □

[**TODO:** Can we get a characterisation of weighted voting? Or a subclass of weighted voting? An easier but still interesting goal might be "binary weighted voting", where $w_N(s) \in \{0, 1\}$.]

2.4 Fixed-points for Recursive Operators

2.5 Satisfaction of the Axioms

In the previous section we introduced several axioms for truth discovery. We now turn back to the example operators from Section 2.2, to assess which axioms hold for each operator. Table 2.2 summarises the results.

[**TODO:** Mention Voting axioms. Proofs are similar to Weighted Agreement.]

Weighted Voting. First we consider weighted voting. The following axioms hold for *any* choice of weighting w.

Lemma 2.5.1. *Let w be a weighting. Then T^w satisfies **Claim-coherence**, **Marginal-trustworthiness** and **Trust-based-monotonicity**.*

Proof. **Claim-coherence** follows easily using the definition of weighted voting and Proposition 2.3.1.

[6] Strictly speaking, we should define an object reduction f between N' and N'', and refer to $f(c)$ and $f(d)$ in N'' instead of c and d. For simplicity we identify c with $f(c)$ and d with $f(d)$ in this proof sketch.

Table 2.2: Axiom satisfaction for the example operators.

	Voting	WeightedAgg	Sums	CRH-ε	TruthFinder	TruthFinder ($\lambda = 0$)
Claim-coherence	✓	✓	✓			
Source-coherence	X	X	✓			
Symmetry	✓	✓	✓			
Fresh-pos-resp	✓	✓	X			
Source-pos-resp	X	✓	X			
Marginal-trustworthiness	✓	✓	✓			
Trust-based-monotonicity	✓	✓	✓			
Classical-independence	✓	X	X			
Disjoint-independence	✓	✓	X			
Conflict-coherence	X	X	X			
Anti-coherence	X	X	X			

One can easily show that if s is marginally trustworthy with respect to T^w then $w_N(s) \geq 0$, and if s is marginally untrustworthy with respect to T^w then $w_N(s) \geq 0$, and **Marginal-trustworthiness** follows.

Finally, for **Trust-based-monotonicity** suppose $\mathrm{src}_N(d) = \mathrm{src}_N(c) \cup Z$, where $\mathrm{src}_N(c) \cap Z = \emptyset$. Then $T^w_N(d) = T^w_N(c) + \sum_{s \in Z} w_N(s)$. If each $s \in Z$ is marginally trustworthy then each $w_N(s)$ is non-negative, and so too is the sum. Hence $T^w_N(d) \geq T^w(c)$, so $c \preceq^{T^w}_N d$. If each $s \in Z$ is marginally untrustworthy then each $w_N(s)$ is non-positive, and similarly we get $d \preceq^{T^w}_N c$ as required. □

Corollary 2.5.1. *Any weightable operator satisfies* ***Claim-coherence***, ***Marginal-trustworthiness*** *and* ***Trust-based-monotonicity****.*

Proof. This follows directly from Lemma 2.5.1 since each axiom only refers to ordinal properties of operators. □

For the particular choice of w for Weighted Agreement from Example 2.2.1, we have the following.

Theorem 2.5.1. *Weighted Agreement satisfies* ***Claim-coherence***, ***Symmetry***, ***Fresh-pos-resp***, ***Source-pos-resp***, ***Marginal-trustworthiness***, ***Trust-based-monotonicity*** *and* ***Disjoint-independence****. It does not satisfy* ***Source-coherence***, ***Classical-independence***, ***Conflict-coherence*** *or* ***Anti-coherence****.*

Proof. For brevity, let w denote w^{agg} and T denote $T^{w^{\mathrm{agg}}}$. **Claim-coherence**, **Marginal-trustworthiness** and **Trust-based-monotonicity** follow from Lemma 2.5.1.

For **Symmetry**, suppose F is an isomorphism betweens networks N and N'. From the definition of an isomorphism we have $(s, c) \in R$ iff $(F(s), F(c)) \in R'$. Consequently $\mathrm{src}_N(c) = \{F^{-1}(s') \mid s' \in \mathrm{src}_{N'}(F(c))\}$ and $\mathrm{cl}_N(s) = \{F^{-1}(c') \mid c' \in \mathrm{cl}_{N'}(F(s))\}$. From this one can show $w_N(s) = w_{N'}(F(s))$, which then implies $T_N(s) = T_{N'}(F(s))$ and $T_N(c) = T_{N'}(F(c))$. **Symmetry** now follows.

For **Fresh-pos-resp** and **Source-pos-resp**, we use the following auxiliary result.

Claim 2.5.1. *Suppose* $\mathrm{cl}_N(u) = \emptyset$ *and let* c *be a claim. Then for all* $s \neq u$ *with* $\mathrm{cl}_N(s) \neq \emptyset$*,*

$$w_{N+(u,c)}(s) = w_N(s) + \frac{\mathbb{1}[c \in \mathrm{cl}_N(s)]}{|\mathrm{cl}_N(s)|}.$$

Proof. First, note that for any claim d,

$$|\operatorname{src}_{N+(u,c)}(d)| = |\operatorname{src}_N(d)| + \mathbb{1}[c = d],$$

and since $s \neq u$ we have $\operatorname{cl}_{N+(u,c)}(s) = \operatorname{cl}_N(s)$. Consequently

$$\begin{aligned} w_{N+(u,c)}(s) &= \sum_{d \in \operatorname{cl}_{N+(u,c)}(s)} \frac{|\operatorname{src}_{N+(u,c)}(d)|}{|\operatorname{cl}_{N+(u,c)}(s)|} \\ &= \sum_{d \in \operatorname{cl}_N(s)} \frac{|\operatorname{src}_N(d)| + \mathbb{1}[c = d]}{|\operatorname{cl}_N(s)|} \\ &= \underbrace{\sum_{d \in \operatorname{cl}_N(s)} \frac{|\operatorname{src}_N(d)|}{|\operatorname{cl}_N(s)|}}_{= w_N(s)} + \sum_{d \in \operatorname{cl}_N(s)} \underbrace{\frac{\mathbb{1}[c = d]}{|\operatorname{cl}_N(s)|}}_{=0 \text{ unless } c=d} \\ &= w_N(s) + \frac{\mathbb{1}[c \in \operatorname{cl}_N(s)]}{|\operatorname{cl}_N(s)|} \end{aligned}$$

□

Now, for **Fresh-pos-resp**, suppose $\operatorname{cl}_N(u) = \emptyset$, $c \neq d$ and $d \preceq_N^T c$. We need to show $d \prec_{N+(u,c)}^T c$. Indeed, using Claim 2.5.1 we have

$$\begin{aligned} T_{N+(u,c)}(c) - T_{N+(u,c)}(d) &= w_{N+(u,c)}(u) + \sum_{s \in \operatorname{src}_N(c)} w_{N+(u,c)}(s) - \sum_{s \in \operatorname{src}_N(d)} w_{N+(u,c)}(s) \\ &= |\operatorname{src}_N(c)| + 1 + \sum_{s \in \operatorname{src}_N(c)} \left(w_N(s) + \frac{1}{|\operatorname{cl}_N(s)|} \right) - \sum_{s \in \operatorname{src}_N(d)} \left(w_N(s) + \frac{\mathbb{1}[c \in \operatorname{cl}_N(s)]}{|\operatorname{cl}_N(s)|} \right) \\ &= |\operatorname{src}_N(c)| + 1 + T_N(c) + \sum_{s \in \operatorname{src}_N(c)} \frac{1}{|\operatorname{cl}_N(s)|} - T_N(d) - \sum_{s \in \operatorname{src}_N(c) \cap \operatorname{src}_N(d)} \frac{1}{|\operatorname{cl}_N(s)|} \\ &= |\operatorname{src}_N(c)| + 1 + \underbrace{T_N(c) - T_N(d)}_{\geq 0} + \sum_{s \in \operatorname{src}_N(c) \setminus \operatorname{src}_N(d)} \frac{1}{|\operatorname{cl}_N(s)|} \\ &\geq 1 \\ &> 0. \end{aligned}$$

This shows $T_{N+(u,c)}c > T_{N+(u,c)}(d)$, and thus $d \prec_{N+(u,c)}^T c$ as required.

For **Source-pos-resp**, suppose $s \in \operatorname{antisrc}_N(c)$, $t \in \operatorname{src}_N(c)$, $\operatorname{cl}_N(u) = \emptyset$ and $s \sqsubseteq_N^T t$. Then

$$\begin{aligned} T_{N+(u,c)}(t) - T_{N+(u,c)}(s) &= w_{N+(u,c)}(t) - w_{N+(u,c)}(s) \\ &= \underbrace{w_N(t) - w_N(s)}_{\geq 0} + \frac{\mathbb{1}[c \in \operatorname{cl}_N(t)]}{|\operatorname{cl}_N(t)|} - \underbrace{\frac{\mathbb{1}[c \in \operatorname{cl}_N(s)]}{|\operatorname{cl}_N(s)|}}_{=0} \\ &\geq \frac{1}{|\operatorname{cl}_N(t)|} \\ &> 0 \end{aligned}$$

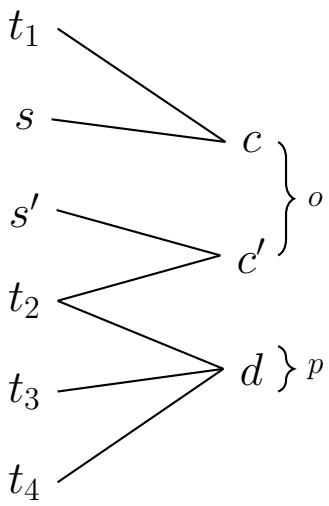


Figure 2.9: Counterexample for ***Source-coherence*** *for Weighted Agreement.*

where we use the fact that $s \in \text{antisrc}_N(c)$ means $c \notin \text{cl}_N(s)$. Hence $s \sqsubset^T_{N+(u,c)} t$.

Finally, **Disjoint-independence** follows easily by noting that for disjoint networks N, N' and $s \in S_N$, $c \in C_N$, we have $\text{cl}_{N \sqcup N'}(s) = \text{cl}_N(s)$ and $\text{src}_{N \sqcup N'}(c) = \text{src}_N(c)$.

To see that **Source-coherence** does not hold, let N be the network shown in Fig. 2.9. One can easily check that $c \prec^T_N c'$ yet $s \simeq^T_N s'$.

Classical-independence cannot hold by the impossibility result Theorem 2.3.1 (1), since **Symmetry**, **Claim-coherence**, **Fresh-pos-resp** and **Source-pos-resp** have already been shown to hold. Similarly, the failure of **Conflict-coherence** and **Anti-coherence** follow from Theorem 2.3.2. □

Sums. To simplify axiomatic analysis of Sums, we first show that T^{sums} is a fixed point of the update function U for Sums. In what follows, let $(\mathcal{D}, T^0, U)$ denote the recursive scheme corresponding to Sums from Definition 2.2.4. Recall that T^{sums} is defined as the limit of this recursive scheme. For simplicity we assume T^{sums} converges on all input networks.[7] We also write $T^n = U^n(T^0)$ for the n-th step of the iteration of Sums.

The following lemma helps to deal with the normalisation factors used in the update function for Sums.

Lemma 2.5.2. *Let* $(x^i_n)_{n\in\mathbb{N}}$ *be convergence sequences in* $\mathbb{R}$*, for* $1 \leq i \leq k$*. Then*

$$\lim_{n\to\infty} \max_i |x^i_n| = \max_i |\lim_{n\to\infty} x^i_n|.$$

Proof. Let $\varepsilon > 0$. Write $y^i = \lim_{n\to\infty} x^i_n$. For each i, hence $|x^i_n| \to |y^i|$ – since the absolute value function $\|\cdot\|$ is continuous – and so there is $n_i \in \mathbb{N}$ such that

[7] While Pasternack and Roth [73] do not consider convergence, Sums is an adaptation of the *Hubs and Authorities* algorithm, for which Kleinberg [59] proves convergence: phrased in our terminology, he shows that the vector of source scores converge to a unit eigenvector of the matrix MM^{T} corresponding to the largest eigenvector (in absolute value), where M is the $|S| \times |C|$ matrix defined by $M_{sc} = \mathbb{1}[s \in \text{src}_N(c)]$. Similarly, claim scores converge to a unit eigenvector of $M^{\text{T}}M$. [**TODO:** possibly signpost that we also take linear algebra approach for unbounded sums?]

$||x_n^i| - |y^i|| < \varepsilon$ for all $n \geq n_i$. Take $m = \max_i n_i$. Let $n \geq m$. For any i, we have

$$|y^i| - \varepsilon < |x_n^i| < |y^i| + \varepsilon.$$

Thus

$$|x_n^i| < |y^i| + \varepsilon \leq \max_j |y^j| + \varepsilon.$$

Since the maximum is achieved for some i, we get

$$\max_i |x_n^i| < \max_j |y^j| + \varepsilon. \tag{2.7}$$

Now, take j such that $\max_i |y^i| = |y^j|$. Then

$$\max_i |x_n^i| \geq |x_n^j| > |y^j| - \varepsilon = \max_i |y^i| - \varepsilon. \tag{2.8}$$

Combining (2.7) and (2.8), we get

$$|\max_i |x_n^i| - \max_i |y^i|| < \varepsilon$$

as required. □

Lemma 2.5.3. *$T^{\text{sums}} \in \mathcal{D}$, and $U(T^{\text{sums}}) = T^{\text{sums}}$.*

Proof. Note that $T_N^n(z) \in [0,1]$ for all n and $z \in S \cup C$. Consequently $T_N^*(z) = \lim_{n\to\infty} T_N^n(z) \in [0,1]$, since $[0,1]$ is closed. Hence $T^{\text{sums}} \in \mathcal{D}$.

Take any network N. If N contains no reports – i.e. $R = \emptyset$, then $T^n \equiv 0$ for all $n > 1$. Hence $T_N^{\text{sums}} \equiv 0$ and $U(T^{\text{sums}})_N = T_N^{\text{sums}}$. Now suppose N contains at least one report (s_0, c_0). It is easily checked that in this case $T_N^n(s_0), T_N^n(c_0) > 0$ for all n. Consequently the maximums in the definition of α and β in Definition 2.2.4 are non-zero. For any $s \in S$, we therefore have

$$\begin{aligned} T_N^{\text{sums}}(s) &= \lim_{n\to\infty} T_N^n(s) \\ &= \lim_{n\to\infty} T_N^{n+1}(s) \\ &= \lim_{n\to\infty} \frac{\sum_{c\in \text{cl}_N(s)} T_N^n(c)}{\max_{t\in S} \left|\sum_{c\in \text{cl}_N(t)} T_N^n(c)\right|} \end{aligned} \tag{2.9}$$

We need to show that the denominator in (2.9) converges to a non-zero limit. By the normalisation step for claim scores, for each $n > 1$ there is a claim c_n with $|T_N^n(c_n)| = 1$. Since there are only finitely many claims, this implies we cannot have $T_N^{\text{sums}}(c) = 0$ for all c, so there is some c_1 with $T_N^{\text{sums}}(c_1) > 0$. Furthermore, $\text{src}_n(c_1) \neq \emptyset$ (otherwise one can easily show $T_N^{\text{sums}}(c_1) = 0$). Likewise, there is some s_1 such that $T_N^{\text{sums}}(s_1) > 0$. Now using the fact that $T_N^n(c) \to T_N^{\text{sums}}(c)$ for each c and taking the limit of the sum, Lemma 2.5.2 gives

$$\lim_{n\to\infty} \max_{t\in S} \left|\sum_{c\in \text{cl}_N(t)} T_N^n(c)\right| = \max_{t\in S} \left|\sum_{c\in \text{cl}_N(t)} T_N^{\text{sums}}(c)\right| \geq T_N^{\text{sums}}(c_1) > 0.$$

Splitting the limit across the quotient in (2.9), we find

$$\begin{aligned} T_N^{\text{sums}}(s) &= \frac{\lim_{n\to\infty} \sum_{c\in \text{cl}_N(s)} T_N^n(c)}{\lim_{n\to\infty} \max_{t\in S} \left|\sum_{c\in \text{cl}_N(t)} T_N^n(c)\right|} \\ &= \frac{\sum_{c\in \text{cl}_N(s)} T_N^{\text{sums}}(c)}{\max_{t\in S} \left|\sum_{c\in \text{cl}_N(t)} T_N^{\text{sums}}(c)\right|} \\ &= U(T^{\text{sums}})_N(s) \end{aligned}$$

as required. One can show $T_N^{\text{sums}}(c) = U(T^{\text{sums}})_N(c)$ for any claim c by a near-identical argument, and thus $U(T^{\text{sums}})_N = T_N^{\text{sums}}$. Since N was arbitrary this shows $U(T^{\text{sums}}) = T^{\text{sums}}$, and the proof is complete. □

Corollary 2.5.2. *T^{sums} is weightable.*

Proof. We define a weighting w as follows. If N contains no reports, set $w_N \equiv 0$. Otherwise, set

$$w_N(s) = \frac{T_N^{\text{sums}}(s)}{\max_{c\in C} \left|\sum_{t\in \text{src}_N(c)} T_N^{\text{sums}}(t)\right|}. \tag{2.10}$$

We need to show $T^{\text{sums}} \sim T^w$, i.e. that T^{sums} and T^w give the same rankings on all networks N. If N contains no reports then both T_N^{sums} and T^w are zero, and therefore output the same rankings. Suppose N contains at least one report. Since we just divide by a constant in (2.10), $s \sqsubseteq_N^{T^{\text{sums}}} s'$ iff $s \sqsubseteq_N^{T^w} s'$ for all sources s and s'. Using the fact that $T^{\text{sums}} = U(T^{\text{sums}})$ from Lemma 2.5.3, it is easily seen that $T_N^{\text{sums}}(c) = \sum_{s\in \text{src}_N(c)} w_N(s) = T_N^w(c)$. Hence T_N^{sums} and T_N^w give exactly the same scores for claims, and in particular the rankings also coincide. □

We come to the axioms satisfied by Sums. While it satisfies both **Claim-coherence** and **Source-coherence**, it is notable that Sums fails both monotonicity properties and **Disjoint-independence**. In some sense these problems are caused by the normalisation step, where source and claim scores are divided by their respective maximums. We present a modified version of Sums without these deficiencies in [**TODO:** section reference].

Theorem 2.5.2. *Sums satisfies **Claim-coherence**, **Source-coherence**, **Symmetry**, **Marginal-trustworthiness**, **Trust-based-monotonicity**. It does not satisfy **Fresh-pos-resp**, **Source-pos-resp**, **Classical-independence**, **Disjoint-independence**, **Conflict-coherence** or **Anti-coherence**.*

Proof. **Claim-coherence**, **Marginal-trustworthiness** and **Trust-based-monotonicity** follow directly from Corollaries 2.5.1 and 2.5.2. For **Source-coherence**, let N be a network and suppose $\text{cl}_N(s)$ strictly precedes $\text{cl}_N(s')$ with respect to $\preceq_N^{T^{\text{sums}}}$. Then by Proposition 2.3.1, there is a bijection $f : \text{cl}_N(s) \to \text{cl}_N(s')$ such that $T_N^{\text{sums}}(c) \leq T_N^{\text{sums}}(f(c))$ for all $c \in \text{cl}_N(s)$, and there is some c_0 with $T_N^{\text{sums}}(c_0) < T_N^{\text{sums}}(f(c_0))$. It follows that N must contain at least one report, since otherwise no strict inequalities hold. For any source t, Lemma 2.5.3 implies

$$T_N^{\text{sums}}(t) = \alpha \sum_{c\in \text{cl}_N(t)} T_N^{\text{sums}}(c),$$

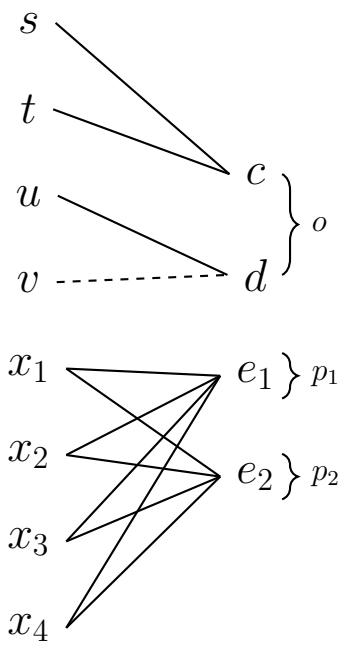


Figure 2.10: Networks used as counterexamples for Sums axiom failures.

where $\alpha = 1/\max_{t' \in S} |\sum_{c \in \mathrm{cl}_N(t')} T_N^{\mathrm{sums}}(c)| > 0$ is a constant. Using the fact that f maps bijectively from $\mathrm{cl}_N(s)$ to $\mathrm{cl}_N(s')$, we get

$$\begin{aligned} T_N^{\mathrm{sums}}(s) - T_N^{\mathrm{sums}}(s') &= \alpha \left(\sum_{c \in \mathrm{cl}_N(s)} T_N^{\mathrm{sums}}(c) - \sum_{c' \in \mathrm{cl}_N(s')} T_N^{\mathrm{sums}}(c') \right) \\ &= \alpha \left(\sum_{c \in \mathrm{cl}_N(s)} T_N^{\mathrm{sums}}(c) - \sum_{c \in \mathrm{cl}_N(s)} T_N^{\mathrm{sums}}(f(c)) \right) \\ &= \alpha \sum_{c \in \mathrm{cl}_N(s)} (T_N^{\mathrm{sums}}(c) - T_N^{\mathrm{sums}}(f(c))). \end{aligned}$$

By assumption $T_N^{\mathrm{sums}}(c) - T_N^{\mathrm{sums}}(f(c)) \leq 0$ for each c, and the inequality is strict for $c = c_0$. Hence $T_N^{\mathrm{sums}}(s) < T_N^{\mathrm{sums}}(s')$, and $s \sqsubset_N^{T^{\mathrm{sums}}} s'$ as required.

Finally, **Symmetry** can be shown in similar way to Weighted Agreement, since Sums is defined only in terms of src_N and cl_N.

For the negative axioms, we refer to networks shown in Fig. 2.10. For **Fresh-pos-resp** and **Source-pos-resp**, let N_0 denote the network without the dashed report (v, d), so that $N_0 + (v, d)$ is the full network. It can be shown that the rankings are the same under Sums in both networks, with $s \simeq t \simeq u \simeq v \sqsubset x_1 \simeq x_2 \simeq x_3 \simeq x_4$ and $c \approx d \prec e_1 \approx e_2$. This violates **Fresh-pos-resp**, since $c \preceq_{N_0}^{T^{\mathrm{sums}}} d$ but $c \not\preceq_{N_0+(v,d)}^{T^{\mathrm{sums}}} d$. It also violates **Source-pos-resp**, since $s \in \mathrm{antisrc}_{N_0}(d)$, $u \in \mathrm{src}_n(d)$ and $s \sqsubseteq_{N_0}^{T^{\mathrm{sums}}} u$, but $s \not\sqsubseteq_{N_0+(v,d)}^{T^{\mathrm{sums}}} u$.

For **Classical-independence** and **Disjoint-independence**, let N_1 and N_2 denote the upper and lower components of the network in Fig. 2.10, excluding the dashed report (v, d). Then $N_0 = N_1 \sqcup N_2$. Hence $c \approx_{N_1 \sqcup N_2}^{T^{\mathrm{sums}}} d$. However, it is

straightforward to check that in the network N_1 alone we have $d \prec_{N_1}^{T^{\text{sums}}} c$; this violates **Disjoint-independence**. Taking N_0' to be the network obtained from N_0 by removing all reports from $x_1, \ldots, x_4$, we have $d \prec_{N_0'}^{T^{\text{sums}}} c$ and $c \approx_{N_0}^{T^{\text{sums}}} d$. Since c and d have the same sources in both networks, this violates **Classical-independence**.

Finally, for **Conflict-coherence** and **Anti-coherence** we can reuse the network N from Fig. 2.6 (including the dashed report). Applying Sums to this network, we have $T_N^{\text{sums}}(s) = T_N^{\text{sums}}(t) = 0$, $T_N^{\text{sums}}(u) = \sqrt{3} - 1 \approx 0.7321$, $T_N^{\text{sums}}(s') = T_N^{\text{sums}}(t') = 1$ and $T_N^{\text{sums}}(c) = T_N^{\text{sums}}(d) = T_N^{\text{sums}}(e) = 0$, $T_N^{\text{sums}}(f) = 1$, $T_N^{\text{sums}}(c') = T_N^{\text{sums}}(d') = \frac{1}{2}(\sqrt{3} - 1) \approx 0.3660$, yielding rankings $s \simeq t \sqsubset u \sqsubset s' \simeq t'$ and $c \approx d \approx e \prec c' \approx d' \prec f$. This ranking violates **Conflict-coherence** since $\text{conflict}_N(c) = \{d\}$ strictly precedes $\text{conflict}_N(c') = \{d'\}$ but $c' \not\prec_N^{T^{\text{sums}}} c$. It also violates **Anti-coherence**, since $\text{antisrc}_N(c) = \{t\}$ strictly precedes $\text{antisrc}_N(c') = \{t'\}$ but $c' \not\prec_N^{T^{\text{sums}}} c$. □

CRH. As with Sums, we can greatly simplify axiomatic analysis of CRH by first showing that the limit operator $T^{\text{crh-}\varepsilon}$ is a fixed point of the update function U. Take $\varepsilon > 0$, and let $(\mathcal{D}, T^0, U)$ denote the recursive scheme of CRH-ε from Definition 2.2.6. As before, we write $T^n = U^n(T^0)$ for the n-th step of the iteration, and assume for simplicity that CRH-ε converges on all networks.

Lemma 2.5.4. *$T^{\text{crh-}\varepsilon} \in \mathcal{D}$, and $U(T^{\text{crh-}\varepsilon}) = T^{\text{crh-}\varepsilon}$.*

Proof. First we show $T^{\text{crh-}\varepsilon} \in \mathcal{D}$; that is, we have $0 \leq T_N^{\text{crh-}\varepsilon}(c) \leq 1$ for all networks N and claims c. First, note that for any operator T and source $s \in S$, we have $U(T)_N(s) > 0$. Consequently $U(T)_N(c) \geq 0$, and

$$U(T)_N(c) = \frac{\sum_{s \in \text{src}_N(c)} U(T)_N(s)}{\sum_{t \in S} U(T)_N(t)} \leq \frac{\sum_{s \in S} U(T)_N(s)}{\sum_{t \in S} U(T)_N(t)} = 1.$$

Since $T^n = U(T^{n-1})$ for $n > 1$, we have $T^n(c) \in [0, 1]$ for all $n > 1$. Consequently $T_N^{\text{crh-}\varepsilon}(c) = \lim_{n \to \infty} T_N^n(c) \in [0, 1]$ also. Thus $T^{\text{crh-}\varepsilon} \in \mathcal{D}$.

Now, take any network N. We aim to show $T_N^{\text{crh-}\varepsilon}(z) = U(T^{\text{crh-}\varepsilon})_N(z)$ for all $z \in S \cup C$. First take $s \in S$. For any $t \in S$, write

$$\alpha_t^n = \varepsilon + \sum_{c \in \text{cl}_N(t)} \sum_{d \in \text{cl}_N(\text{obj}(c))} (T_N^n(d) - \mathbb{1}[d = c])^2.$$

Then $\lim_{n \to \infty} \alpha_t^n = \varepsilon + \sum_{c \in \text{cl}_N(t)} \sum_{d \in \text{cl}_N(\text{obj}(c))} (T_N^{\text{crh-}\varepsilon}(d) - \mathbb{1}[d = c])^2$. We have

$$\begin{aligned} T_N^{\text{crh-}\varepsilon}(s) &= \lim_{n \to \infty} T_N^{n+1}(s) \\ &= \lim_{n \to \infty} \left(\varepsilon - \log \left(\frac{\alpha_s^n}{\sum_{t \in S} \alpha_t^n} \right) \right). \end{aligned} \tag{2.11}$$

Now, since $T_N^n(d) \in [0, 1]$ for all $n \in \mathbb{N}$, and clearly $\mathbb{1}[d = c] \in [0, 1]$, we have

$$\varepsilon \leq \alpha_t^n = \varepsilon + \sum_{c \in \text{cl}_N(t)} \sum_{d \in \text{cl}_N(\text{obj}(c))} \underbrace{(T_N^n(d) - \mathbb{1}[d = c])^2}_{\leq 1} \leq \varepsilon + |C|^2.$$

Hence

$$\frac{\alpha_s^n}{\sum_{t \in S} \alpha_t^n} \geq \frac{\varepsilon}{\sum_{t \in S} (\varepsilon + |C|^2)} = \frac{\varepsilon}{|S|(\varepsilon + |C|^2)} > 0,$$

assuming $S \neq \emptyset$. Since this lower bound is independent of n, this implies $\lim_{n\to\infty} \frac{\alpha_s^n}{\sum_{t\in S} \alpha_t^n} > 0$. By continuity of the logarithm and (2.11), we get

$$T_N^{\text{crh-}\varepsilon}(s) = \varepsilon - \log\left(\frac{\lim_{n\to\infty} \alpha_s^n}{\sum_{t\in S} \lim_{n\to\infty} \alpha_t^n}\right) = U(T^{\text{crh-}\varepsilon})_N(s)$$

as required.

Now take any $c \in C$. From above we have $\lim_{n\to\infty} T_N^n(t) = T_N^{\text{crh-}\varepsilon}(t) \geq \varepsilon > 0$ for each $t \in S$. By simple manipulation of limits we find

$$\begin{aligned} T_N^{\text{crh-}\varepsilon}(c) &= \lim_{n\to\infty} T_N^n(c) \\ &= \lim_{n\to\infty} \frac{\sum_{s\in \text{src}_N(c)} T_N^n(s)}{\sum_{t\in S} T_N^n(t)} \\ &= \frac{\sum_{s\in \text{src}_N(c)} \lim_{n\to\infty} T_N^n(s)}{\sum_{t\in S} \lim_{n\to\infty} T_N^n(t)} \\ &= \frac{\sum_{s\in \text{src}_N(c)} T_N^{\text{crh-}\varepsilon}(s)}{\sum_{t\in S} T_N^{\text{crh-}\varepsilon}(t)} \\ &= U(T^{\text{crh-}\varepsilon})_N(c). \end{aligned}$$

This completes the proof. □

Corollary 2.5.3. *$T^{\text{crh-}\varepsilon}$ is weightable.*

Proof. From Lemma 2.5.4 we have

$$T_N^{\text{crh-}\varepsilon}(c) = \frac{\sum_{s\in \text{src}_N(c)} T_N^{\text{crh-}\varepsilon}(s)}{\sum_{t\in S} T_N^{\text{crh-}\varepsilon}(t)}.$$

Defining a weighting w by $w_N(s) = \frac{T_N^{\text{crh-}\varepsilon}(s)}{\sum_{t\in S} T_N^{\text{crh-}\varepsilon}(t)}$, it is easily seen that $T^{\text{crh-}\varepsilon} \sim T^w$. □

[**TODO:** intro to crh axioms]

Theorem 2.5.3. *Take $\varepsilon > 0$.*

2.5.1 Modifying Sums

Failure of **Disjoint-independence** is bad. Show that Sums converges ordinally, which resolves the issue

2.6 Related Work

2.7 Conclusion

3 Bipartite Tournaments

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 [16] (where players represent alternatives and x beats y if a majority of voters prefer x over y), paired comparisons analysis [45] (where players may represent products and the beating relation the preferences of a user), search engines [86], sports tournaments [14] 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 [55], 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 [25]).

A related example is *truth discovery* [63, 80]: 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 [41]. Other application domains include the evaluation of generative models in machine learning [70] (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

[45, 85, 22], 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 [55] 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.

Contributions. 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.

This chapter is an extension of Singleton and Booth [81], with new results presented in Section 3.5.4.

Chapter outline. In Section 3.1 we define the framework for bipartite tournaments and introduce chain graphs. Section 3.2 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 3.3. Section 3.4 defines a concrete scheme for producing chain-editing-based rankings. Section 3.5 introduces new ranking methods by relaxing the chain editing requirement. Related work is discussed in Section 3.6, and we conclude in Section 3.7.

3.1 Preliminaries

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

3.1.1 Bipartite Tournaments

Following the literature on generalised tournaments [45, 85, 22], we represent a tournament as a matrix, whose entries represent the results of matches between participants. In what follows, $[n]$ denotes the set $\{1, \dots, n\}$ whenever $n \in \mathbb{N}$.

Definition 3.1.1. *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.

Definition 3.1.2. *An* operator T *assigns each tournament K a pair $T(K) = (\preceq_K^T, \sqsubseteq_K^T)$ of total preorders on A and B respectively.*[2]

For $a, a' \in A$, we interpret $a \preceq_K^T a'$ to mean that a' is ranked *at least as strong* as a in the tournament K, according to the operator T (similarly, $b \sqsubseteq_K^T b'$ means b' is ranked at least as strong as b). The strict and symmetric parts of $\preceq_K^T$ are denoted by $\prec_K^T$ and $\approx_K^T$.

As a simple example, consider T_{count}, where $a \preceq_K^{T_{\mathsf{count}}} a'$ iff $|K(a)| \leq |K(a')|$ and $b \sqsubseteq_K^{T_{\mathsf{count}}} b'$ 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 [76], and generalises *Copeland's rule* [16].

3.1.2 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

[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.

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

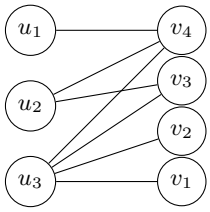


Figure 3.1: An example of a chain graph

a tournament admits a particularly simple solution if this graph happens to be a *chain graph*.

Definition 3.1.3 ([98])**.** *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 3.1 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 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 3.1.4. *A tournament K has the* chain property *if $\leqslant_K^{\mathcal{A}}$ is a total preorder.*

[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.

[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.

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 [16].

Example 3.1.1. *Consider* $K = \left[\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{smallmatrix}\right]$. *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 3.1.*

3.2 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 [55], 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 ranking. In this section we introduce the class of operators producing rankings in this way.

3.2.1 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 3.2.1. *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 [55], amounts to finding a single element of $\mathcal{M}(K)$.[6] We comment further on the computational complexity of chain editing in Section 3.6. The following property characterises chain editing-based operators T.

> **Chain-min**. For every tournament K there is $K' \in \mathcal{M}(K)$ such that $T(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$.

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*.

[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 [31].

Example 3.2.1. *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 3.2.1 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 3.4 we introduce a principled way to single out a *unique* chain tournament and thereby construct a well-defined chain-minimal operator.

3.2.2 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 [15, 45]), voting in social choice theory [33], truth discovery [91], belief merging [36] 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 3.2.2. *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:*

$$\forall a, a' \in A \quad (x_a < x_{a'} \implies \exists b \in B : x_a < y_b \leq x_{a'}) \tag{3.1}$$

$$\forall b, b' \in B \quad (y_b < y_{b'} \implies \exists a \in A : y_b \leq x_a < y_{b'}) \tag{3.2}$$

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 (3.1) 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 ((3.2) is analogous for the Bs). These conditions are intuitive if we assume that skill levels

[7] Here $a_1a_2a_3$ 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.

[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 $b \in B$.

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 [15, 45, 33]. 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 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 3.2.3. *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*

$$P_{\boldsymbol{\alpha}}(X_{ab} = 1 \mid \theta) = \begin{cases} \alpha_+, & x_a < y_b \\ 1 - \alpha_-, & x_a \geq y_b \end{cases} \tag{3.3}$$

$$P_{\boldsymbol{\alpha}}(X_{ab} = 0 \mid \theta) = \begin{cases} 1 - \alpha_+, & x_a < y_b \\ \alpha_-, & x_a \geq y_b \end{cases} \tag{3.4}$$

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 (3.3) and (3.4) correspond to a false positive, true positive, true negative and false negative respectively. We can now define a maximum likelihood operator.

Definition 3.2.4. *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 T 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 $a \preceq_K^T a'$ iff $x_a \leq x_{a'}$ and $b \sqsubseteq_K^T b'$ 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.

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

[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.

Lemma 3.2.1. *Let K be an $m \times n$ tournament, $\boldsymbol{\alpha} \in [0,1]^2$ and $\theta \in \Theta_{m,n}$. Then*

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

Proof. Write $p_{ab,K}$ for $P_{\boldsymbol{\alpha}}(X_{ab} = K_{ab} \mid \theta)$. Expanding the product in Definition 3.2.3, 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 (3.3) and (3.4) 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

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

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 3.2.2. *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 3.2.1,

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

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 3.2.3. *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 (3.1) in the definition of a state (Definition 3.2.2), 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 3.2.4. *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 3.2.3 (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 (3.1) and (3.2) of Definition 3.2.2. For (3.1), 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 (3.1) holds.

For (3.2), 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 (3.2).

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 3.2.3 relies crucially on (3.1) and (3.2) 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 3.2.1. *Let $\boldsymbol{\alpha} = \langle \beta, \beta \rangle$ for some $\beta < \frac{1}{2}$. Then T is a maximum likelihood operator w.r.t $\boldsymbol{\alpha}$ if and only if T 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 3.2.2, 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 3.2.4, $\mathcal{K}_{\Theta_{m,n}}$ 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, T satisfies **Chain-min** iff for every tournament K there is $K' \in \mathcal{M}(K)$ such that $T(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$. Using Lemma 3.2.4 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 $T(K) = (\leqslant_{K_\theta}^{\mathcal{A}}, \leqslant_{K_\theta}^{\mathcal{B}})$. But by Lemma 3.2.3 we have $a \leqslant_{K_\theta}^{\mathcal{A}} a'$ iff $x_a \leq x_{a'}$ and $b \leqslant_{K_\theta}^{\mathcal{B}} b'$ 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. □

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 [55]). Similarly, the case $\alpha_- = 0$ and $\alpha_+ \in (0,1)$ corresponds to *chain deletion*, where edge additions are not allowed.

3.3 Axiomatic analysis

Chain-minimal operators have theoretical backing in a probabilistic sense due to the results of Section 3.2.2, 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 3.4 we describe a scenario in which this is acceptable and define a class of concrete operators for this case, and in Section 3.5 we relax the **Chain-min** requirement in order to gain anonymity.

3.3.1 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 T 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 [76].

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.

> **Anon**. Let $\sigma : A \to A$ and $\pi : B \to B$ be permutations. Then $a \preceq^T_K a'$ iff $\sigma(a) \preceq^T_{\pi(\sigma(K))} \sigma(a')$.

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 3.3.1. *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 $\overline{K}$.

Dual. $b \sqsubseteq_K^T b'$ iff $b \preceq_{\overline{K}}^T b'$.

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 $a \preceq_K^T a'$ iff $a \sqsubseteq_{\overline{K}}^T a'$, 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 [76], which has subsequently been called *Independence of Irrelevant Matches* [45].

IIM. If K_1, K_2 are tournaments of the same size with identical a-th and a'-th rows, then $a \preceq_{K_1}^T a'$ iff $a \preceq_{K_2}^T a'$.

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 [45], 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.

Mon. If $K(a) \subseteq K(a')$ then $a \preceq_K^T a'$.

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

Pos-resp. If $a \preceq_K^T a'$ and $K_{a',b} = 0$ for some $b \in B$, then $a \prec_{K+\mathbf{1}_{a',b}}^T a'$, 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 [76], 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

T_{count}, and it can be shown that T_{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 $a \preceq_K^T a'$, then **Pos-resp** implies $a \prec_{K-\mathbf{1}_{ab}}^T a'$.

3.3.2 Axiom Compatibility with Chain-min

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

Theorem 3.3.1. *There is no operator satisfying* ***Chain-min*** *and any of* ***Anon***, ***IIM*** *or* ***Pos-resp****.*

Proof. We take each axiom in turn. Let T 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 T satisfied **Anon**, we would get $1 \preceq_K^T 2$ iff $\sigma(1) \preceq_{\pi(\sigma(K))}^T \sigma(2)$ iff $2 \preceq_K^T 1$, which implies $1 \approx_K^T 2$. On the other hand, we have

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

Since T satisfies **Chain-min** and $1, 2 \in A$ rank equally in $\preceq_K^T$, there must be $K' \in \mathcal{M}(K)$ such that 1 and 2 rank equally in $\leqslant_{K'}^{\mathcal{A}}$, 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 T cannot satisfy **Anon**.

IIM. Suppose T 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 $1 \preceq_{K_1}^T 2$ iff $1 \preceq_{K_2}^T 2$. 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 $T(K_i) = (\leqslant_{K_i'}^{\mathcal{A}}, \leqslant_{K_i'}^{\mathcal{B}})$. In particular, $1 \prec_{K_1}^T 2$ and $2 \prec_{K_2}^T 1$. But this contradicts **IIM**, and we are done.

Pos-resp. Suppose T 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 $T(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$. Note that $K'(1) = K'(2)$, so we have $1 \approx_K^T 2$. In particular, $1 \preceq_K^T 2$. Since $K_{23} = 0$, we may apply **Pos-resp** to get

$1 \prec^T_{K+\mathbf{1}_{23}} 2$. But $K + \mathbf{1}_{23}$ is just K'. Since the chain property already holds for K', we have $\mathcal{M}(K') = \{K'\}$ and consequently

$$T(K + \mathbf{1}_{23}) = T(K') = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'}) = T(K)$$

so in fact $1 \approx^T_{K+\mathbf{1}_{23}} 2$, 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 3.2.2 provides informal explanation for this result. For K above to arise in the noise model of Definition 3.2.3 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 3.3.2. *For each of **Dual** and **Mon**, there exists an operator satisfying **Chain-min** and the stated property.*

Despite the simplicity of **Mon**, Theorem 3.3.2 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 T, and using properties of the set $\mathcal{M}(K)$ to produce another operator T' 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 3.3.1. *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.

- 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 $b, b' \in B$,

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

so $\leqslant_K^{\mathcal{B}} = \leqslant_{\overline{K}}^{\mathcal{A}}$.

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

  Since $\leqslant_{K'}^{\mathcal{B}}$ always has reflexivity and transitivity, we only need to show the totality property. Let $b, b' \in B$ and suppose $b \not\leqslant_{K'}^{\mathcal{B}} b'$. We must show $b' \leqslant_{K'}^{\mathcal{B}} b$, 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 $b' \leqslant_{K'}^{\mathcal{B}} b$ 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 3.3.1. *There exists an operator T satisfying **Chain-min** and **Dual**.*

Proof. Let T be an arbitrary operator satisfying **Chain-min**. Then there is a function $\alpha : \mathcal{K} \to \mathcal{K}$ such that $T(K) = (\leqslant_{\alpha(K)}^{\mathcal{A}}, \leqslant_{\alpha(K)}^{\mathcal{B}})$ 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

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

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

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 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 3.3.1 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 T' by $T'(K) = (\leqslant^{\mathcal{A}}_{\alpha'(K)}, \leqslant^{\mathcal{B}}_{\alpha'(K)})$. Since $\alpha'(K) \in \mathcal{M}(K)$ for all K, T' satisfies **Chain-min**. Moreover, using Lemma 3.3.1 part (1) and the fact that $\overline{\alpha'(K)} = \alpha'(\overline{K})$, for any tournament K and $b, b' \in B$ we have

$$\begin{aligned} b \sqsubseteq^{T'}_{K} b' &\iff b \leqslant^{\mathcal{B}}_{\alpha'(K)} b' \\ &\iff b \leqslant^{\mathcal{A}}_{\overline{\alpha'(K)}} b' \\ &\iff b \leqslant^{\mathcal{A}}_{\alpha'(\overline{K})} b' \\ &\iff b \sqsubseteq^{T'}_{\overline{K}} b' \end{aligned}$$

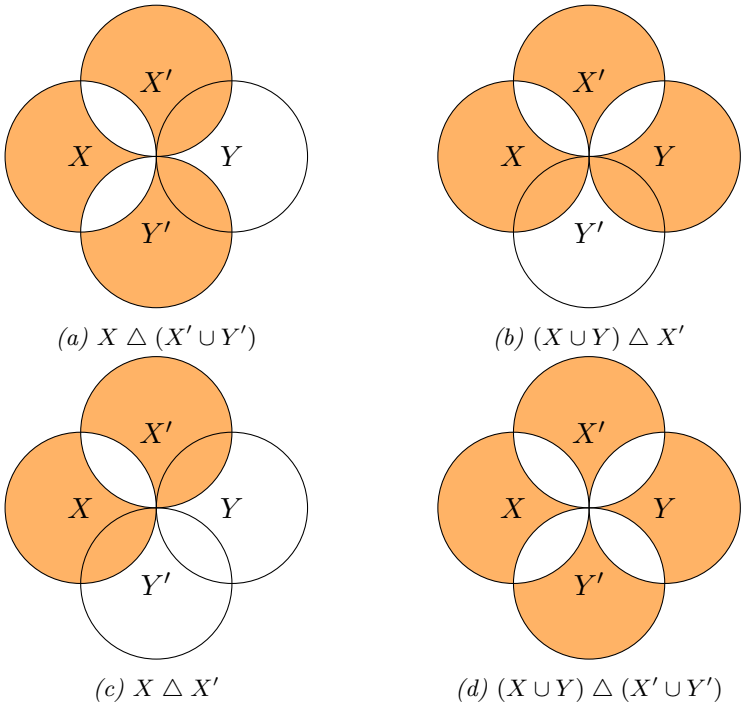
which shows T' also satisfies **Dual**. □

To find an operator satisfying **Chain-min** and **Mon**, we proceed in three stages. First, Lemma 3.3.2 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 3.3.3 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 3.3.4 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 3.3.2. *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.*

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

Lemma 3.3.2. *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 $\mathsf{swap}(K'; a_1, a_2) \in \mathcal{M}(K)$.*



(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 3.2: Depictions of the sets in (3.5)

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

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

[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 $\ll$.

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

$$\begin{aligned} d(K,K') - d(K,K'') &= |K(a_1) \mathbin{\triangle} K'(a_1)| + |K(a_2) \mathbin{\triangle} K'(a_2)| \\ &\quad - |K(a_1) \mathbin{\triangle} K'(a_2)| - |K(a_2) \mathbin{\triangle} K'(a_1)| \\ &= |X \mathbin{\triangle} (X' \cup Y')| + |(X \cup Y) \mathbin{\triangle} X'| \\ &\quad - |X \mathbin{\triangle} X'| - |(X \cup Y) \mathbin{\triangle} (X' \cup Y')| \end{aligned} \tag{3.5}$$

Each of the symmetric differences in (3.5) are depicted in Figure 3.2. 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 (3.5) and consulting Figure 3.2, 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. □

We need some new notation. For a relation R on a set X and $x \in X$, write

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

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

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

Lemma 3.3.3. *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 3.3.2. 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 $a \in U(a_{i+1}, \leqslant^{\mathcal{A}}_{K_{i+1}})$ 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 $|L(a_j, <^{\mathcal{A}}_K)| > |L(a_{i+1}, <^{\mathcal{A}}_K)|$). 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 $a \in U(a_j, <^{\mathcal{A}}_K)$. 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. $\hat{a} \in U(a_j, <^{\mathcal{A}}_K)$. 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 3.3.4. *Let K be a tournament and suppose $K' \in \mathcal{M}(K)$ is such that $U(a, <^{\mathcal{A}}_K) \subseteq U(a, \leqslant^{\mathcal{A}}_{K'})$ for all $a \in A$. Then there is $K'' \in \mathcal{M}(K)$ such that $\leqslant^{\mathcal{A}}_K \subseteq \leqslant^{\mathcal{A}}_{K''}$.*

Proof. Let $A_1, \ldots, A_t \subseteq A$ be the equivalence classes of $\approx^{\mathcal{A}}_K$, the symmetric part of $\leqslant^{\mathcal{A}}_K$. Note that $a \approx^{\mathcal{A}}_K a'$ 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

$$f(A_i) \in \arg\min_{a \in A_i} |B_i \triangle K'(a)| \in A_i$$

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 $a \leqslant_K^{\mathcal{A}} a'$ 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. $f([a]) <_K^{\mathcal{A}} f([a'])$. By the assumption on K' in the statement of the lemma, this means $f([a]) \leqslant_{K'}^{\mathcal{A}} f([a'])$, and so

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

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

The pieces are now in place to prove the following.

Proposition 3.3.2. *There exists an operator T satisfying **Chain-min** and **Mon**.*

Proof. For any tournament K, write

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

By Lemma 3.3.3 and Lemma 3.3.4, $\mathcal{M}_{\mathsf{mon}}(K)$ 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 T by $T(K) = (\leqslant_{\alpha(K)}^{\mathcal{A}}, \leqslant_{\alpha(K)}^{\mathcal{B}})$, we see that T satisfies **Chain-min** and **Mon**, as required. □

Theorem 3.3.2 now follows from Proposition 3.3.1 and Proposition 3.3.2.

3.4 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, $\mathrm{vec}_{\trianglelefteq}(K) = (K_{a_1,b_1},\dots,K_{a_{mn},b_{mn}})$, where $(a_1,b_1),\dots,(a_{mn},b_{mn})$ 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 T by setting $T(K) = (\leqslant^{\mathcal{A}}_{\alpha(K)}, \leqslant^{\mathcal{B}}_{\alpha(K)})$.

Definition 3.4.1. *Let $\trianglelefteq$ be a total order on $\mathbb{N}\times\mathbb{N}$. Define an operator $T_{\trianglelefteq}$ according to the choice function*

$$\alpha_{\trianglelefteq}(K) = \operatorname*{arg\,min}_{K'\in\mathcal{M}(K)} \mathrm{vec}_{\trianglelefteq}(K\oplus K') \tag{3.6}$$

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 3.4.1. *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 3.2.1. Writing $K_1,\dots,K_4$ for the elements of $\mathcal{M}(K)$ in the order that they appear in Example 3.2.1 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 $T_{\trianglelefteq}$ ranks K according to K_4, i.e. $1 \prec^{T_{\trianglelefteq}}_K 2 \prec^{T_{\trianglelefteq}}_K 3$ and $1 \approx^{T_{\trianglelefteq}}_K 3 \sqsubset^{T_{\trianglelefteq}}_K 2 \sqsubset^{T_{\trianglelefteq}}_K 4$.

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

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

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 (3.6). First, a technical result is required.

Lemma 3.4.1. *Let k and l be integers with $1 \leq k \leq l$. 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

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

as required. □

The characterisation in terms of weighted Hamming distances is as follows

Theorem 3.4.1. *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:*

$$\arg\min_{K' \in \mathcal{C}_{m,n}} d_w(K, K') = \{\alpha_{\trianglelefteq}(K)\} \tag{3.7}$$

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

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

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

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 (3.7), 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 3.4.1 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 (3.6). 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 3.4.1 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 3.4.1 and $m = 2$, $n = 3$ are $w = [\,{}^{1.5}_{1.0625}\ {}^{1.25}_{1.03125}\ {}^{1.125}_{1.015625}\,]$.

3.5 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 3.3, and present new axioms which characterise this intuitive instance.

3.5.1 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 $T(K)$ corresponds to *some* chain tournament, not necessarily one closest to K. We call this property *chain-definability*.

> **Chain-def**. For every $m \times n$ tournament K there is $K' \in \mathcal{C}_{m,n}$ such that $T(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$.

Clearly **Chain-min** implies **Chain-def**. 'Chain-definable' operators can also be cast in the MLE framework of Section 3.2.2 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 $\preceq_K^T$ and $\sqsubseteq_K^T$. 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 3.5.1. *T satisfies **Chain-def** if and only if $|\mathsf{ranks}(\preceq_K^T) - \mathsf{ranks}(\sqsubseteq_K^T)| \leq 1$ 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 T satisfies **Chain-def**, and let K be a tournament. We need to show that $|\mathsf{ranks}(\preceq_K^T) - \mathsf{ranks}(\sqsubseteq_K^T)| \leq 1$.

By chain-definability, there is K' with the chain property such that $a \preceq_K^T a'$ iff $K'(a) \subseteq K'(a')$ and $b \sqsubseteq_K^T b'$ iff $(K')^{-1}(b) \supseteq (K')^{-1}(b')$. Write

$$\mathcal{X} = \{[a]_{\preceq_K^T} \mid a \in A, K'(a) \neq \emptyset\}$$

$$\mathcal{Y} = \{[b]_{\sqsubseteq_K^T} \mid b \in B, (K')^{-1}(b) \neq \emptyset\}$$

for the set of ranks in each of the two orders, excluding those who have empty neighbourhoods in K'. Note that $[a]_{\preceq_K^T} = [a']_{\preceq_K^T}$ 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 $a_1 \prec_K^T \cdots \prec_K^T a_s$, 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. $b_i \sqsubset_K^T b_{i+1}$.

We now have $b_1 \sqsubset_K^T \cdots \sqsubset_K^T b_s$; a chain of s strict inequalities in $\sqsubseteq_K^T$. 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 $\mathsf{ranks}(\preceq_K^T) \in \{|\mathcal{X}|, |\mathcal{X}| + 1\}$, 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, $\mathsf{ranks}(\sqsubseteq_K^T) \in \{|\mathcal{Y}|, |\mathcal{Y}| + 1\}$. Since $|\mathcal{X}| = |\mathcal{Y}|$, it is clear that $\mathsf{ranks}(\preceq_K^T)$ and $\mathsf{ranks}(\sqsubseteq_K^T)$ can differ by at most one, as required.

Now we prove the 'if' statement. Let K be a tournament. We have $|\mathsf{ranks}(\preceq_K^T) - \mathsf{ranks}(\sqsubseteq_K^T)| \leq 1$, and must show there is tournament K' with the chain property such that $T(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$.

Let $X_1 < \cdots < X_s$ and $Y_1 < \cdots < Y_t$ be the ranks of $\preceq_K^T$ and $\sqsubseteq_K^T$ 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 $\{X_1, \ldots, X_s\}$ is a partition of A. Note that due to the assumption on the ordering of the X_i, we have $a \preceq_K^T a'$ 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

$$\begin{aligned} a \preceq_K^T a' &\iff p(a) \leq p(a') \\ &\iff N_{p(a)} \subseteq N_{p(a')} \\ &\iff K'(a) \subseteq K'(a') \\ &\iff a \leqslant_{K'}^{\mathcal{A}} a', \end{aligned} \tag{3.8}$$

i.e. $\preceq_K^T = \leqslant_{K'}^{\mathcal{A}}$. Since $\preceq_K^T$ is a total preorder, this shows that K' has the chain property.

It only remains to show that $\sqsubseteq_K^T = \leqslant_{K'}^{\mathcal{B}}$. First note that if $a \in X_i$ and $b \in Y_j$, the fact that $\{Y_1, \ldots, Y_t\}$ 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 (3.8) shows that $\sqsubseteq_K^T = \leqslant_{K'}^{\mathcal{B}}$.

We have shown that $T(K) = (\leqslant_{K'}^{\mathcal{A}}, \leqslant_{K'}^{\mathcal{B}})$ and that K' has the chain property, and the proof is therefore complete. □

3.5.2 Interleaving Operators

According to Theorem 3.5.1, to construct a chain-definable operator it is enough to ensure that the number of ranks of $\preceq_K^T$ and $\sqsubseteq_K^T$ 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 [13].

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 3.5.1. *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 3.5.2. *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 $T = T_{f,g}^{\mathsf{int}}$ *by* $a \preceq_K^T a'$ *iff* $r(a) \geq r(a')$ *and* $b \sqsubseteq_K^T b'$ *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 3.5.1. *Let f and g be selection functions. Fix a tournament K and let A_i, B_i ($i \geq 0$) be as in Definition 3.5.2. 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 3.5.1 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 3.5.2. *An operator T satisfies **Chain-def** if and only if $T = T^{\mathsf{int}}_{f,g}$ for some selection functions f, g.*

Theorem 3.5.2 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 $T = T^{\mathsf{int}}_{f,g}$ be an interleaving operator with selection functions f, g, and fix a tournament K. We will show that $T(K)$ is chain-definable.

As per Proposition 3.5.1, 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 $a \preceq^T_K a'$ iff $r(a) \geq r(a')$, it follows that the non-empty sets $A_0 \setminus A_1, \ldots, A_{j-1} \setminus A_j$ form the ranks of the total preorder $\preceq^T_K$ (that is, the equivalence classes of the symmetric closure $\approx^T_K$). Thus, $\preceq^T_K$ has j ranks. An identical argument shows that $\sqsubseteq^T_K$ has j' ranks.

It follows from Theorem 3.5.1 that $T(K)$ 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 3.5.1, 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 3.5.1 shows that $j = j' + 1$, and we have $|j - j'| = |1| = 1$.

Hence $|j - j'| \leq 1$ in all cases, and $T(K)$ is chain-definable as required.

Now for the 'only if' direction. Suppose T satisfies **Chain-def**. We will define f, g such that $T = T^{\mathsf{int}}_{f,g}$. 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 $\preceq^T_K$ (g will be defined similarly). The interleaving algorithm will then select the ranks of $\preceq^T_K$ and $\sqsubseteq^T_K$ one-by-one; the fact that $T(K)$ 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 3.5.1, $|\mathsf{ranks}(\preceq^T_K) - \mathsf{ranks}(\sqsubseteq^T_K)| \leq 1$. Taking $t = \max\{\mathsf{ranks}(\preceq^T_K), \mathsf{ranks}(\sqsubseteq^T_K)\}$, we can write $X_1, \ldots, X_t \subseteq A$ and $Y_1, \ldots, Y_t \subseteq B$ for the ranks of $\preceq^T_K$ and $\sqsubseteq^T_K$ respectively, possibly with $X_1 = \emptyset$ if $\mathsf{ranks}(\sqsubseteq^T_K) = 1 + \mathsf{ranks}(\preceq^T_K)$ or $Y_1 = \emptyset$ if $\mathsf{ranks}(\preceq^T_K) = 1 + \mathsf{ranks}(\sqsubseteq^T_K)$. Note that $X_i, Y_i \neq \emptyset$ for $i > 1$. Assume these sets are ordered such that $a \preceq^T_K a'$ 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]

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

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

It is not difficult to see that f and g satisfy the conditions of Definition 3.5.1 for selection functions. We claim that for with A_i, B_i denoting the interleaving sets for K and f, g, for all $0 \leq i \leq t$ we have

$$A_i = \bigcup_{j=1}^{t-i} X_j, \quad B_i = \bigcup_{j=1}^{t-i} Y_j \tag{3.9}$$

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

Now suppose (3.9) 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

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

since the X_j form (disjoint) ranks of $\preceq^T_K$ 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

[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 $\prec$).

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, (3.9) holds for all $0 \leq i \leq t$.

It remains to show that $a \preceq_K^T a'$ iff $a \preceq_K^{T_{f,g}^{\mathsf{int}}} a'$ and that $b \sqsubseteq_K^T b'$ iff $b \sqsubseteq_K^{T_{f,g}^{\mathsf{int}}} b'$.

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 $\preceq_K^T$. 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

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

A similar argument shows that $b \sqsubseteq_K^T b'$ iff $b \sqsubseteq_K^{T_{f,g}^{\mathsf{int}}} b'$ for any $b, b' \in B$. Since K was arbitrary, we have shown that $T = T_{f,g}^{\mathsf{int}}$ as required. □

Table 3.1: Iteration of the interleaving algorithm for T_{CI}

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

We now come to an important example.

Example 3.5.1. *Define the* cardinality-based interleaving operator $T_{\mathsf{CI}} = T^{\mathsf{int}}_{f,g}$ *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 3.1 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* $\preceq^T_K$ *(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 3.1.*

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 $a \in f(K, A_i, B_i)$ 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 3.1, 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 T_{CI} from Example 3.5.1 uses simple cardinality-based heuristics, and can be seen as a chain-definable version of T_{count} (which is not chain-definable). It is also the bipartite counterpart to repeated applications of Copeland's rule [13]. 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 T_{CI} can be computed in $O(N^3)$ time.

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

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

3.5.3 Axiom Compatibility

We now revisit the axioms of Section 3.3 in relation to chain-definable operators in general and T_{CI} specifically. Firstly, the weakening of **Chain-min** pays off: **Chain-def** is compatible with all our axioms.

Theorem 3.5.3. *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 3.3.2 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

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

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 T defined by $T(K) = (\leqslant^{\mathcal{A}}_{K'}, \leqslant^{\mathcal{B}}_{K'})$ satisfies **Chain-def**. In particular, $a \preceq^T_K a'$ iff $|K(a)| \leq |K(a')|$.

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

Since $a \preceq^T_K a'$ implies $|K(a)| \leq |K(a')|$, we have $|\hat{K}(a')| = 1 + |K(a')| > |K(a)| = |\hat{K}(a)|$, and therefore $a \prec^T_{\hat{K}} a'$ 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 $\mathsf{ranks}(\preceq^T_K) = 3$. However, **Anon** with **Dual** implies the ranking of B is flat, i.e. $\mathsf{ranks}(\sqsubseteq^T_K) = 1$. This contradicts **Chain-def** by Theorem 3.5.1, yielding the following impossibility result.

Theorem 3.5.4. *There is no operator satisfying **Chain-def**, **Anon**, **Dual** and **Pos-resp**.*

Proof. For contradiction, suppose there is an operator T 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 $1 \approx^T_{K_1} 2$ (e.g. take $\sigma = (1\ 2)$, $\pi = \mathrm{id}_B$). In particular, $1 \preceq^T_{K_1} 2$, so **Pos-resp** implies $1 \prec^T_K 2$. A similar argument with K_2 shows that $3 \approx^T_{K_2} 4$ and $3 \prec^T_K 4$.

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

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 $1 \approx_{\overline{K}}^T 2$ 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. $\mathsf{ranks}(\sqsubseteq_K^T) = 1$. We see that $\mathsf{ranks}(\preceq_K^T)$ and $\mathsf{ranks}(\sqsubseteq_K^T)$ differ by 2, contradicting **Chain-def** according to Theorem 3.5.1. □

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

Lemma 3.5.1. *Let $T = T_{f,g}^{int}$ 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 T 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 T 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 T 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 3.5.2 for K and K' respectively. We claim that for all $i \geq 0$:

$$A'_i = \sigma(A_i), \quad B'_i = \pi(B_i) \tag{3.10}$$

   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 (3.10) 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, (3.10) 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 3.5.2 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 $a \preceq^T_K a'$ iff $\sigma(a) \preceq^T_{K'} \sigma(a')$, i.e. T 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 3.5.2 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 $b \sqsubseteq^T_K b'$ iff $b \preceq^T_{\overline{K}} b'$, 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 $a \preceq^T_K a'$.

   Suppose otherwise, i.e. $a' \prec^T_K a$. 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 T_{CI}, Lemma 3.5.1 yields the following.

Theorem 3.5.5. *T_{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 T_{CI} from Example 3.5.1.

Chain-def. Since T_{CI} is an interleaving operator, **Chain-def** follows from Theorem 3.5.2.

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 3.5.1 part (1) are satisfied.

Let $A' \subseteq A$ and $B' \subseteq B$. We have

$$\begin{aligned} f(K', \sigma(A'), \pi(B')) &= \arg\max_{\hat{a} \in \sigma(A')} |K'(\hat{a}) \cap \pi(B')| \\ &= \sigma(\arg\max_{a \in A'} |K'(\sigma(a)) \cap \pi(B')|) \end{aligned}$$

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

$$\begin{aligned} f(K', \sigma(A'), \pi(B')) &= \sigma(\arg\max_{a \in A'} |K'(\sigma(a)) \cap \pi(B')|) \\ &= \sigma(\arg\max_{a \in A'} |\pi(K(a)) \cap \pi(B')|) \\ &= \sigma(\arg\max_{a \in A'} |\pi(K(a) \cap B')|) \\ &= \sigma(\arg\max_{a \in A'} |K(a) \cap B'|) \\ &= \sigma(f(K, A', B')) \end{aligned}$$

as required. The result for g follows by a near-identical argument. Thus T_{CI} satisfies **Anon** by Lemma 3.5.1 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

$$\begin{aligned} g(K, A', B') &= \arg\min_{b \in B'} |K^{-1}(b) \cap A'| \\ &= \arg\min_{b \in B'} \left(|A'| - |\overline{K}(b) \cap A'|\right) \\ &= \arg\max_{b \in B'} |\overline{K}(b) \cap A'| \\ &= f(\overline{K}, B', A') \end{aligned}$$

and, by Lemma 3.5.1 part (2), T_{CI} satisfies **Dual**.

Mon. Once again, we use Lemma 3.5.1. 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 3.5.1 part (3) is satisfied, and thus T_{CI} satisfies **Mon**.

Pos-resp. We have show that T_{CI} satisfies **Chain-def**, **Anon** and **Dual**; due to impossibility result of Theorem 3.5.4, T_{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 $1 \preceq_{K_1}^{T_{\mathsf{CI}}} 2$ iff $1 \preceq_{K_2}^{T_{\mathsf{CI}}} 2$. However, it is easily verified that $1 \prec_{K_1}^{T_{\mathsf{CI}}} 2$ whereas $2 \prec_{K_2}^{T_{\mathsf{CI}}} 1$. Therefore T_{CI} does not satisfy **IIM**. □

Note that **Anon** *is* satisfied. This makes T_{CI} an important example of a well-motivated, tractable, chain-definable and anonymous operator, meeting the criteria outlined at the start of this section.

3.5.4 Axiomatic Characterisation of T_{CI}

In Theorem 3.5.5 we saw which of the axioms from Section 3.3 hold for T_{CI}. We now characterise T_{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 $\preceq_K^T$ should coincide with that of $\preceq_K^{T_{\mathsf{CI}}}$. Together with **Dual** and **Chain-def**, these will characterise T_{CI}.

Unlike the axioms introduced so far, which were straightforward, general properties for ranking methods, the new axioms are geared specifically towards characterising T_{CI}. Thus, this section takes a *descriptive* perspective as opposed to the *normative* perspective of Section 3.3.

Towards the characterisation result, we first note that T_{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 3.1 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 $T_{f,g}^{\mathsf{int}}$, called *sub-matrix independence*.

> **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 3.5.2. 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 T 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 3.5.4) that this is not possible: **SMI** either holds for T or does not, independently of the choice of selection functions f

and g. Nevertheless, to avoid circularity we will say T satisfies **SMI** if *there exist* f, g with the **SMI** property such that $T = T^{\mathsf{int}}_{f,g}$.

First, T_{Cl} does indeed satisfy **SMI**.

Proposition 3.5.2. *T_{Cl} satisfies **SMI**.*

Proof. Let f and g denote the selection functions for T_{Cl}. 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 T_{Cl} 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 3.5.3. *Let T and T' be interleaving operators satisfying **SMI**. Suppose that for all tournaments K,*

$$\begin{aligned} \max(A, \preceq^T_K) &= \max(A, \preceq^{T'}_K) \\ \max(B, \sqsubseteq^T_K) &= \max(B, \sqsubseteq^{T'}_K). \end{aligned}$$

Then $T = T'$.

Proof. Let f, g and f', g' be selection functions corresponding to T and T' respectively. Take a tournament K. To show $T(K) = T(K')$ 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 3.5.2 for T and T' 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 3.5.1 for f and **SMI**, we have

$$\begin{aligned} f(K, A_i, B_i) &= \{a_p \mid p \in f(K^-, [m'], [n'])\} \\ f'(K, A'_i, B'_i) &= \{a_p \mid p \in f'(K^-, [m'], [n'])\}. \end{aligned} \tag{3.11}$$

But $[m']$ and $[n']$ are the full set of players in K^-, so

$$\begin{aligned} f(K^-, [m'], [n']) &= \max(A_{K^-}, \preceq^T_{K^-}) \\ &= \max(A_{K^-}, \preceq^{T'}_{K^-}) \\ &= f'(K^-, [m'], [n']), \end{aligned}$$

where our assumption on the maximal ranks for T and T' is employed in the second step. Consulting (3.11) 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. □

This result simplifies the task of characterising T_{CI} among the interleaving operators with **SMI**, since we only need to consider the maximal ranks of A and B. In fact, given that T_{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 T_{CI} together with **Dual**.

Argmax. $\max(A, \preceq^T_K) = \arg\max_{a \in A} |K(a)|$.

Note that **Argmax** does *not* require $\preceq^T_K$ to reduce to T_{count}, since we only consider a such that $|K(a)|$ is *maximal*. Now, since **Chain-def** characterises interleaving operators, to obtain a characterisation of T_{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.

Rank-removal. Suppose $\max(A, \preceq^T_K) \neq A$ and $\max(B, \sqsubseteq^T_K) \neq B$. Write $A \setminus \max(A, \preceq^T_K) = \{a_1, \ldots, a_{m'}\}$ and $B \setminus \max(B, \sqsubseteq^T_K) = \{b_1, \ldots, b_{m'}\}$, 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',

$$\begin{aligned} a_p \preceq^T_K a_{p'} &\iff p \preceq^T_{K^-} p' \\ b_q \sqsubseteq^T_K b_{q'} &\iff q \sqsubseteq^T_{K^-} q'. \end{aligned}$$

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 3.5.2. *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*

$$f(K, A_i, B_i) = \max(A_i, \preceq_K^{T_{f,g}^{\mathsf{int}}})$$
$$g(K, A_i, B_i) = \max(B_i, \sqsubseteq_K^{T_{f,g}^{\mathsf{int}}}).$$

Proof. We show the first statement; the second follows by an identical argument. For brevity, write T for $T_{f,g}^{\mathsf{int}}$. 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 $a' \preceq_K^T a$. 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 $a' \preceq_K^T a$.

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 $a \notin \max(A_i, \preceq_K^T)$. 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 $a \prec_K^T a'$. Hence $a \notin \max(A_i, \preceq_K^T)$, as required. □

Proposition 3.5.4. *T satisfies **Chain-def** and **Rank-removal** if and only if T is an interleaving operator satisfying **SMI**.*

Proof. For the 'if" direction, suppose $T = T_{f,g}^{\mathsf{int}}$ is an interleaving operator satisfying **SMI**. Then T satisfies **Chain-def** by Theorem 3.5.2. We show **Rank-removal**. Let $\{a_1, \ldots, a_{m'}\}$, $\{b_1, \ldots, b_{n'}\}$ and K^- be as in the statement of **Rank-removal**.

Let A_i, B_i denote the interleaving sets for T applied to K, and A_i^-, B_i^- those for T applied to K^-. First we claim that for all $p \in [m']$, $q \in [n']$ and $i \geq 0$,

$$\begin{aligned} p \in A_i^- &\iff a_p \in A_{i+1}, \\ q \in B_i^- &\iff b_q \in B_{i+1}. \end{aligned} \tag{3.12}$$

We prove (3.12) by induction on i. Take $i = 0$. By definition, $A_0^- = A_{K^-} = [m']$, so $p \in A_0^-$ always holds. Recall that $\{a_1, \ldots, a_{m'}\} = A \setminus \max(A, \preceq_K^T)$. It is easily seen that $\max(A, \preceq_K^T) = f(K, A_0, B_0)$, so in fact $\{a_1, \ldots, a_{m'}\} = A_1$. Hence $a_p \in A_1$ always holds. The argument for the Bs is identical. This proves (3.12) for $i = 0$.

For the inductive step, suppose (3.12) 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 (3.12) holds for $i+1$. For the second part, using property (3) from Definition 3.5.1 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 (3.12) 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 $A_i^- = \{p_1, \ldots, p_{m''}\}$ and $B_i^- = \{q_1, \ldots, q_{n''}\}$ 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 $T = T^{\mathsf{int}}_{f,g}$ 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 (3.12).

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 (3.12) 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 $\preceq^T_{K^-}$ corresponds exactly to that of $\{a_1, \ldots, a_{m'}\}$ in $\preceq^T_K$, as required for **Rank-removal**. The case for the B ranking is identical, and we are done.

Now for the 'only if" direction, suppose T satisfies **Chain-def** and **Rank-removal**. By Theorem 3.5.2 there are selection functions f and g such that $T = T^{\mathsf{int}}_{f,g}$. 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 3.5.2 we see that A_i is the result of removing the top $i-1$ ranks of players in A, according to the ranking $\preceq_K^T$. Applying **Rank-removal** $i-1$ times, we get

$$a_p \preceq_K^T a_{p'} \iff p \preceq_{K^-}^T p'$$

for all $p, p' \in [m']$. Consequently, using Lemma 3.5.2 again,

$$\begin{aligned} a_p \in f(K, A_i, B_i) &\iff a_p \in \max(A_i, \preceq_K^T) \\ &\iff \forall p' \in [m'] : \ a_{p'} \preceq_K^T a_p \\ &\iff \forall p' \in [m'] : \ p' \preceq_{K^-}^T p \\ &\iff p \in \max([m'], \preceq_{K^-}^T) \\ &\iff p \in f(K^-, [m'], [n']) \end{aligned}$$

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. □

Finally, we can state the axiomatic characterisation of T_{CI}.

Theorem 3.5.6. *T_{CI} is the unique operator satisfying **Dual**, **Chain-def**, **Rank-removal** and **Argmax**.*

Proof. We have already seen in Theorem 3.5.5 that T_{CI} satisfies **Dual** and **Chain-def**. For **Argmax**, note that for any tournament K,

$$\begin{aligned} \max(A, \preceq_K^{T_{\mathsf{CI}}}) &= f(K, A, B) \\ &= \arg\max_{a \in A} |K(a) \cap B| \\ &= \arg\max_{a \in A} |K(a)| \end{aligned}$$

directly from the definition of the selection function f for T_{CI}. Finally, **Rank-removal** follows from Proposition 3.5.4 since T_{CI} is an interleaving operator with **SMI** (by Proposition 3.5.2).

For uniqueness, suppose some operator T also satisfies the stated axioms. By Proposition 3.5.4, T is an interleaving operator satisfying **SMI**. To show $T = T_{\mathsf{CI}}$ it is sufficient by Proposition 3.5.3 to show that T and T_{CI} agree on maximal ranks for all tournaments K. Clearly this is the case for the ranking of A, since **Argmax** completely prescribes $\max(A, \preceq_K^T)$ solely in terms of K. Moreover, by **Dual** we have $\sqsubseteq_K^T = \preceq_{\overline{K}}^T$, so

$$\begin{aligned} \max(B_K, \sqsubseteq_K^T) &= \max(A_{\overline{K}}, \preceq_{\overline{K}}^T) \\ &= \arg\max_{b \in A_{\overline{K}}} |\overline{K}(b)| \\ &= \max(A_{\overline{K}}, \preceq_{\overline{K}}^{T_{\mathsf{CI}}}) \\ &= \max(B, \sqsubseteq_K^{T_{\mathsf{CI}}}) \end{aligned}$$

where we apply **Argmax** for T and T_{CI} to the dual tournament $\overline{K}$. This completes the proof. □

3.6 Related Work

On chain graphs. Chain graphs were originally introduced by Yannakakis [98], 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* [69] (where only edge deletions are allowed) and chain *editing* [31] (where both additions and deletions are allowed). We refer the reader to the work of Jiao, Ravi, and Gatterbauer [55] and Drange et al. [31] for a more detailed account of this literature. Outside of complexity theory, chain graphs have been studied for their spectral properties in [3, 42], and the more general notion of a *nested colouring* was introduced in [21].

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* [16], 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 [16] – 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 [45] is a pair (X, A), where $X = [t]$ for some $t \in \mathbb{N}$ and $A \in \mathbb{R}_{\geqslant 0}^{t \times t}$ is a non-negative $t \times t$ matrix with $A_{ii} = 0$ for all $i \in X$. In this formalism each encounter between a pair of players i and j is represented by *two* numbers: A_{ij} and A_{ji}. This allows one to model both intensities of victories and losses (including draws) via the difference $A_{ij} - A_{ji}$, and the case where a comparison is not available (where $A_{ij} = A_{ji} = 0$).

Any $m \times n$ bipartite tournament K has a natural generalised tournament representation via the $(m+n) \times (m+n)$ *anti-diagonal block matrix* $A = \left[\begin{smallmatrix} 0 & K \\ \bar{K} & 0 \end{smallmatrix}\right]$, where the 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 A is strongly connected. This is not the case in general for A constructed as above, which means not all existing tournament operators (and tournament axioms) are well-defined for bipartite inputs.[21] Consequently,

[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 [33].

bipartite tournaments are a special case of generalised tournaments *in principle*, but not in practise.

3.7 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 T_{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 [55]; 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 [68]. The interleaving operators of Section 3.5.2 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 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.

[21] We note that Slutzki and Volij [85] side-step the reducibility issue by decomposing T into irreducible components and ranking each separately, although their methods may give only *partial* orders.

Interlude

In Chapters 2 and 3 we studied how to rank sources by trustworthiness, broadly speaking, in the context of truth discovery and bipartite tournament ranking. However, we had no formal semantics to define the *meaning* of trustworthiness, and indeed this meaning varies between truth discovery operators and ranking methods. This flexibility was useful for our social-choice-style analysis, where rankings are commonly used in this manner. For instance, truth discovery operators using different interpretations of trustworthiness can still be meaningfully compared by the axiomatic method.

In the remainder of the thesis we take a stricter view on trustworthiness, positioned in relation to *expertise* in a logic-based framework. Informally, a source is trustworthy with respect to a topic X if we *believe they are an expert* on matters relating to X. Under this interpretation, the truth discovery problem can be modelled in a similar way to *belief revision* [1] and *belief merging* [61] by considering how to form *beliefs* on the basis of input reports. Specifically, by considering beliefs both about the state of the world and the expertise of the sources, we have analogues of both the source and claim rankings from truth discovery.

Beyond its relation to trustworthiness, expertise is also a topic of interest in its own right. The logical properties of expertise – and the connections to the truthfulness of information – are explored in detail in the next chapter using the tools of modal logic. This framework is used as the basis for a belief change problem in Chapter 5, which can be thought of as a complementary, logical version of truth discovery. Finally, Chapter 6 shifts the focus away from normative properties of aggregation methods – as expressed by axioms in Chapter 2 and postulates in Chapter 5 – towards *truth-tracking*, i.e. how one can find the truth given reports from non-expert sources.

4 Expertise and Information

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.

Problems associated with expertise have been exacerbated recently by the COVID-19 pandemic, in which false information from non-experts has been shared widely on social media [65, 26]. There have also been high-profile instances of experts going beyond their area of expertise to comment on issues of public health [93], highlighting the importance of *domain-specific* notions of expertise. Identifying experts is also an important task for *liquid democracy* [11], in which voters may delegate their votes to expertise on a given policy issue.

Expertise has been well-studied, with perspectives from behavioural and cognitive science [17, 35], sociology [19], and philosophy [58, 92, 44], among other fields. In this work we study the *logical* content of expertise, and its relation to truthfulness of information.

Specifically, we develop a *modal logic* framework to model *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 p, on which the source *does* have expertise, and which is true. If this holds for all possible ways to weaken $p \wedge q$ (this is the case, for instance, if the source does not have expertise on any statement strictly stronger than p), then $p \wedge q$

[1] Note that we could instead consider the dual case: expertise means being able to *verify* when a proposition is true.

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, a committee may consist of several experts across different domains, so that by working together the group achieves expertise beyond any of its individual members. Indeed, such pooling of expertise becomes necessary in cases where it is infeasible for an individual to be a specialist in all relevant sub-areas. As a concrete example, consider the *Rogers Commission report*[2] into the 1986 Challenger disaster, whose members included politicians, military generals, physicists, astronauts and rocket scientists. Beyond extending the expertise of its constituents, the breadth of expertise among the commission allowed it to collectively assess issues at the *intersection* of its members' specialities.

Towards defining collective expertise we will again turn to (multi-agent) epistemic logic, borrowing from the well-known notions of *distributed* and *common knowledge* [38]. 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.

While the picture of expertise painted so far has been static, it is also natural to consider the *dynamics* of expertise. For example, how does expertise change over time as sources interact with the world and gain new knowledge? What are the effects of *announcements*, particularly when sources are non-experts? We study the logic of such events via dynamic operators in the style of dynamic epistemic logic [28].

Contributions. On the conceptual side, we develop a modal logic framework to reason about the expertise of a source and soundness of information. We also study collective expertise among multiple sources, and consider how expertise may evolve via learning and announcements. Importantly, both singular and collective expertise are shown to connected in a precise sense to standard notions from epistemic logic. This formalises the conceptual link between expertise and *knowledge*. On the technical side we obtain a sound and complete axiomatisation, and axiomatise several sub-classes of models with additional axioms.

[**TODO:** Publication statement. Either a highly extended version of ESSLLI, or slightly extended version of Synthese (dynamics).]

Chapter outline. In Section 4.1 we give a motivating example and define the syntax and semantics. Section 4.2 looks at how expertise may be closed under certain operations (e.g. conjunction, negation). The core connection with epistemic logic is given in Section 4.3. We turn to axiomatics in Section 4.4, and give sound and complete logics for various classes of expertise models. In Section 4.5 we generalise to multiple sources. Section 4.6 introduces the dynamic extension of the logic, and

[2] https://en.wikipedia.org/wiki/Rogers_Commission_Report

we conclude in Section 4.7. Several of the main proofs have also been formalised with the Lean theorem prover.[3]

4.1 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 4.1.1. *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 ($p \wedge \neg d$). 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 $p \wedge \neg d$ 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 $p \wedge \neg d$ 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:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \mathsf{E}\varphi \mid \mathsf{S}\varphi \mid \mathsf{A}\varphi$$

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 [46] for technical convenience; $\mathsf{A}\varphi$ is read as "φ holds in all states". Other logical connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) and constants ($\top$, $\bot$) are introduced as abbreviations.

Semantics. On the semantic side, we use the notion of an *expertise model*.

Definition 4.1.1. *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.

[3] https://github.com/anonymous-logician/expertise. [**TODO:** de-anonymise repo.]

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* [78, 67, 72], 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:[4] $\mathsf{E}\varphi$ holds iff $\|\varphi\|_M$ is closed, and $\mathsf{S}\varphi$ holds at x iff x lies in the *closure* of $\|\varphi\|_M$.[5] 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 4.1.2. *To formalise Example 4.1.1, consider the model $M = (X, P, V)$ shown in Fig. 4.1, where $X = 2^{\{i,p,d\}}$, $P = \{\{ipd, pd, ip, p\}, \{id, d, i, \emptyset\}\}$ (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 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 4.1.1.*

We further illustrate the semantics by listing some valid formulas.

Proposition 4.1.1. *The following formulas are valid:*

[4] 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 4.2.

[5] 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 $\|\varphi\|$.

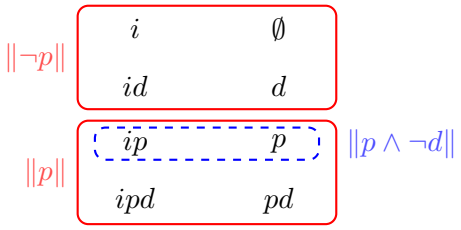


Figure 4.1: Expertise model from Example 4.1.2, which formalises Example 4.1.1.

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.

4.2 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 4.2.1. *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 $F \models \mathsf{E}\varphi \wedge \mathsf{E}\psi \to \mathsf{E}(\varphi \vee \psi)$ 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 4.2.1 (1) and Proposition 4.1.1 (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 4.2.1. *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 $\mathcal{L}$.[6] 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 4.2.2. *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 $x \notin B_1$.
- $M, x \not\models \mathsf{S}(p \wedge \neg q)$: Note that $\|p \wedge \neg q\|_M = V(p) \setminus V(q) = B_2 \setminus B_1$. Therefore we have $B_2 \in P$ and $\|p \wedge \neg q\|_M \subseteq B_2$, but $x \notin B_2$.

[6] 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 4.2.1. Then $M', 1 \models \mathsf{E}(p \wedge q)$ and $M, 1 \not\models \mathsf{E}(p \wedge q)$, but M and M' agree on $\mathcal{L}_{\mathsf{SA}}$ formulas. Hence $\mathsf{E}(p \wedge q)$ is not equivalent to any $\mathcal{L}_{\mathsf{SA}}$ formula.

Now set $\varphi = \neg q$ and $\psi = \neg p$. We have

$$\hat{\mathsf{S}}(\varphi \to \psi) = \neg \mathsf{S} \neg (\varphi \to \psi) \equiv \neg \mathsf{S}(\varphi \wedge \neg \psi) \equiv \neg \mathsf{S}(p \wedge \neg q)$$

$$\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi = \neg \mathsf{S} \neg \varphi \to \neg \mathsf{S} \neg \psi \equiv \neg \mathsf{S} q \to \neg \mathsf{S} p \equiv \mathsf{S} p \to \mathsf{S} q$$

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

$$\|\neg\psi\|_M \subseteq \|\varphi \wedge \neg\psi\|_M \cup \|\neg\varphi\|_M = \|\neg(\varphi \to \psi)\|_M \cup \|\neg\varphi\|_M \subseteq A \cup B.$$

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 $F \models \hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi)$. □

4.3 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 4.3.1. *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 source 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 [9, 71].[7] 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 4.3.2. *For a set X and $P \subseteq 2^X$, let R_P be the binary relation on X given by*

$$xR_Py \iff \forall A \in P : (y \in A \implies x \in A)$$

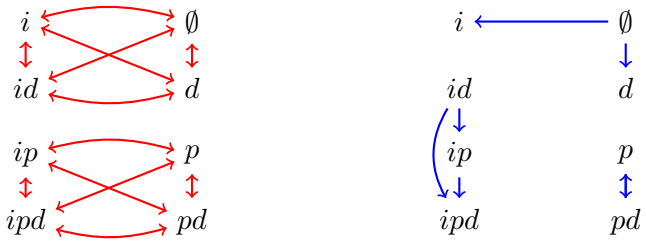


Figure 4.2: Left: the relation R_P corresponding to X and P from Example 4.1.2 (with reflexive edges omitted). Note that R_P is an equivalence relation, with equivalence classes $\|p\|$ and $\|\neg p\|$. Right: an example of a non-symmetric relation R_P, corresponding to $P = \{\emptyset, X, \{id, ip, ipd\}, \{id, ip\}, \{id\}, \{i, \emptyset\}, \{\emptyset, d\}, \{p, pd\}\}$.

In the case where P is the collection of closed sets of a topology on X, R_P is the *specialisation preorder*. Fig. 4.2 shows an example of R_P for X and P from Example 4.1.2. 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 4.3.1. *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 4.3.2. *Let X be a set.*

[7] 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* [71, pp. 24].

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 $A \subseteq X$:*

$$A \in P \iff A \text{ is downwards closed wrt } R_P.$$

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 4.3.1, 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 4.3.1. *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 4.3.2 (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 4.3.2 (2), since P is fully determined by R_P for expertise models closed under unions and intersections. Finally, Lemma 4.3.2 (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 4.3.2. *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 4.3.1 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 4.3.2 (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 4.3.1. 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 4.3.2 (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 4.3.3. *Let* $f : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}} \to \mathbb{M}^*_{\mathsf{S4}}$ *be the bijection from Theorem 4.3.1. Then for all* $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$, $x \in X$ *and* $\varphi \in \mathcal{L}$*:*

$$M, x \models \varphi \iff f(M), x \models t(\varphi) \tag{4.1}$$

Moreover, if $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$ *is the bijection from Theorem 4.3.2, then for all* $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$*:*

$$M, x \models \varphi \iff g(M), x \models t(\varphi) \tag{4.2}$$

Proof. Note that since g is defined as the restriction of f to $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$, (4.2) follows from (4.1). We show (4.1) 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 4.3.2 (2), we have

$$A \in P \iff A \text{ is downwards closed wrt R} \qquad (*)$$

We show (4.1) 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.

4.4 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 4.2.

Table 4.1: 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})
$\mathsf{A}(\varphi \to \psi) \to (\mathsf{S}\varphi \wedge \mathsf{E}\psi \to \psi)$	(W_{E})
$\varphi \to \mathsf{S}\varphi$	(T_{S})
$\mathsf{SS}\varphi \to \mathsf{S}\varphi$	(4_{S})
$\mathsf{A}(\varphi \to \psi) \to (\mathsf{S}\varphi \to \mathsf{S}\psi)$	(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)

The General Case. Let L be the extension of propositional logic generated by the axioms and inference rules shown in Table 4.1. Note that we treat A as a "box" and S as a "diamond" modality. Some of the axioms were already seen in Proposition 4.1.1; 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 4.4.1. 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 4.1.1 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 $\vdash (\psi_0 \wedge \cdots \wedge \psi_n) \to \varphi$. 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 4.4.2. *Let* Γ *be a maximally consistent set and* $\varphi, \psi \in \mathcal{L}$. *Then*

1. $\varphi \in \Gamma$ *iff* $\Gamma \vdash \varphi$
2. *If* $\varphi \to \psi \in \Gamma$ *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 $\varphi \to \varphi$ is an instance of the propositional tautology $p \to p$, we have $\vdash \varphi \to \varphi$. 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 $\vdash (\psi_0 \wedge \cdots \wedge \psi_n) \to \bot$. 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), $\vdash (\psi_1 \wedge \cdots \wedge \psi_n) \to \neg\varphi$. 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 $\vdash ((\varphi \to \psi) \wedge \varphi) \to \psi$. 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 $\Gamma \nvdash \varphi$ 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.

□

Lemma 4.4.3 (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 4.4.4. *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 $\Gamma R \Lambda$. □

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 4.4.5. *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 $\psi_0, \ldots, \psi_n \in \Gamma_0$ 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 4.4.3), 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 4.4.1. *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 4.4.5 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 4.4.5. 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

$$\begin{aligned} \widehat{P}^0_\Sigma &= \{|\varphi|_\Sigma \times \mathbb{R} \mid \mathsf{E}\varphi \in \Sigma\}, \\ \widehat{P}^1_\Sigma &= \{A \subseteq \widehat{X}_\Sigma \mid A \text{ is S-closed and } \forall \varphi \in \mathcal{L} : A \neq |\varphi|_\Sigma \times \mathbb{R}\}. \end{aligned}$$

We have a version of the truth lemma for the model $\widehat{M}_\Sigma = (\widehat{X}_\Sigma, \widehat{P}_\Sigma, \widehat{V}_\Sigma)$.

Lemma 4.4.6. *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 4.4.5 and Corollary 4.4.1. We treat the cases of E and S formulas.

- E: First suppose $\mathsf{E}\varphi \in \Gamma$. By Corollary 4.4.1, $\mathsf{E}\varphi \in \Sigma$. 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 $\|\varphi\|_{\widehat{M}_\Sigma} \in \widehat{P}_\Sigma$. By the induction hypothesis, $\|\varphi\|_{\widehat{M}_\Sigma} = |\varphi|_\Sigma \times \mathbb{R}$. Hence $|\varphi|_\Sigma \times \mathbb{R} \in \widehat{P}_\Sigma$. 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 4.4.5 then gives $\mathsf{A}(\varphi \leftrightarrow \psi) \in \Sigma$. By Corollary 4.4.1, 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 4.4.5, $\mathsf{A}(\varphi \to \psi) \in \Sigma$. By Corollary 4.4.1 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 is $\mathcal{D}$. Since $\mathbb{R}$ is uncountable, there is some $s \in \mathbb{R}$ such that $(\Gamma, s) \notin \mathcal{D}$.[8] 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 ($\mathrm{T}_\mathsf{S}$) and ($\mathrm{Nec}_\mathsf{A}$), we have $\mathsf{A}(\varphi \to \mathsf{S}\varphi) \in \Sigma$, and consequently $|\varphi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$ by Lemma 4.4.5. 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 4.4.5, $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Sigma$.

  Now, take any $(\Lambda, u) \in |\mathsf{S}\psi|_\Sigma \times \mathbb{R}$. Since $\Lambda \in X_\Sigma$, Corollary 4.4.1 gives $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Lambda$. By ($\mathrm{W}_\mathsf{S}$), $\mathsf{S}\psi \to \mathsf{SS}\varphi \in \Lambda$. Since $\Lambda \in |\mathsf{S}\psi|_\Sigma$, we get $\mathsf{SS}\varphi \in \Lambda$.

[8] If not, then $s \mapsto (\Gamma, s)$ is an injective mapping $\mathbb{R} \to \mathcal{D}$, which would imply $\mathbb{R}$ is countable.

But then (4_{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 4.4.1. L *is strongly complete*[9] *with respect to* $\mathbb{M}$.

Proof. We show the contrapositive. Suppose $\Gamma \nvdash \varphi$. 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 4.4.6 (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 sub-classes of $\mathbb{M}$ corresponding to closure conditions.

To start, consider closure under intersections. It was shown in Proposition 4.2.1 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 4.4.2. $\mathsf{L}_{\mathsf{int}}$ *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 4.2.1 (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 4.4.5 and Corollary 4.4.1.

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

$$P_\Sigma = \{A \subseteq X_\Sigma \mid A \text{ is S-closed}\}.$$

[9] That is, for all sets $\Gamma \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, if $\Gamma \models \varphi$ then $\Gamma \vdash \varphi$.

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 4.4.6). 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), $\mathsf{S}\psi \to \mathsf{SS}\varphi \in \Delta$. Now $\mathsf{S}\psi \in \Delta$ implies $\mathsf{SS}\varphi \in \Delta$, 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 $\Gamma \nvdash_{\mathsf{L}_{\mathsf{int}}} \varphi$, 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 $\Gamma \not\models_{\mathbb{M}_{\mathsf{int}}} \varphi$. □

Now we add finite unions to the mix. It was shown in Proposition 4.2.2 that within class $\mathbb{M}_{\mathsf{int}}$, the **K** axiom for the dual operator $\hat{\mathsf{S}}\varphi = \neg\mathsf{S}\neg\varphi$ characterises closure under finite unions. Note that any frame (X, P) closed under intersections and finite unions is a topological space,[10] where P is the set of *closed* sets. Write

[10] 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.

$\mathbb{M}_{\mathsf{top}} = \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{finite-unions}}$ 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:

$$\begin{array}{ll} \hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi) & (\mathrm{K}_{\mathsf{S}}) \\ \mathsf{A}\varphi \to \hat{\mathsf{S}}\varphi & (\mathrm{Inc}) \end{array}$$

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 4.3.3 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 $\mathbb{M}_{\mathsf{top}}$.[11]

Theorem 4.4.3. $\mathsf{L}_{\mathsf{top}}$ *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 4.2.2. 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 4.3.3. 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 4.2. It is well known that $\mathsf{L}_{\mathsf{S4A}}$ is strongly complete with respect to $\mathbb{M}^*_{\mathsf{S4}}$ [10, Theorem 7.2].

Table 4.2: 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:

$$\begin{array}{ll} u(p) & = p \\ u(\varphi \wedge \psi) & = u(\varphi) \wedge u(\psi) \\ u(\neg\varphi) & = \neg u(\varphi) \\ u(\mathsf{K}\varphi) & = \neg\mathsf{S}\neg u(\varphi) \\ u(\mathsf{A}\varphi) & = \mathsf{A}u(\varphi). \end{array}$$

Recall the translation $t : \mathcal{L} \to \mathcal{L}_{\mathsf{KA}}$ from Section 4.3. 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 $\mathsf{L}_{\mathsf{top}}$-provably equivalent to $u(t(\varphi))$.

[11] Note that **KT4** is also complete for topological spaces with respect to the interior semantics [9].

Claim 4.4.1. *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 4.4.2. *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 $\mathsf{L_{S4A}}$ proofs. The base case consists of showing that if φ is an instance of an $\mathsf{L_{S4A}}$ axiom or a substitution instance of a propositional tautology, then $\vdash_{\mathsf{L_{top}}} u(\varphi)$. The case for instances of tautologies is straightforward, since u does not affect the structure of a propositional formula. We take the axioms of $\mathsf{L_{S4A}}$ in turn.

- (K_{K}): We have
$$\begin{aligned} &u(\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)) \\ &\quad = \neg \mathsf{S} \neg (u(\varphi) \to u(\psi)) \to (\neg \mathsf{S} \neg u(\varphi) \to \neg \mathsf{S} \neg u(\psi)) \\ &\quad = \hat{\mathsf{S}}(u(\varphi) \to u(\psi)) \to (\hat{\mathsf{S}} u(\varphi) \to \hat{\mathsf{S}} u(\psi)) \end{aligned}$$
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 L_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 L_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 L_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 L_top.

- (Inc_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 $\vdash_{\mathsf{L}_\mathsf{top}} u(\psi_i)$ for each i then $\vdash_{\mathsf{L}_\mathsf{top}} u(\varphi)$.

- (Nec_A): If $\vdash_{\mathsf{L}_\mathsf{top}} u(\varphi)$, then from ($\mathrm{Nec}_\mathsf{A}$) in L_top we get $\vdash_{\mathsf{L}_\mathsf{top}} \mathsf{A}u(\varphi)$. But $\mathsf{A}u(\varphi) = u(\mathsf{A}\varphi)$, so we are done.

- (MP): Similarly, this clear from (MP) for L_top and the fact that $u(\varphi \to \psi) = u(\varphi) \to u(\psi)$.

□

Claims 4.4.1 and 4.4.2 easily imply the following.

Claim 4.4.3. *Let $\varphi \in \mathcal{L}$. Then $\vdash_{\mathsf{L}_\mathsf{S4A}} t(\varphi)$ implies $\vdash_{\mathsf{L}_\mathsf{top}} \varphi$.*

Proof. Suppose $\vdash_{\mathsf{L}_\mathsf{S4A}} t(\varphi)$. By Claim 4.4.2, $\vdash_{\mathsf{L}_\mathsf{top}} u(t(\varphi))$. By Claim 4.4.1, $\vdash_{\mathsf{L}_\mathsf{top}} \varphi \leftrightarrow u(t(\varphi))$. By (MP), $\vdash_{\mathsf{L}_\mathsf{top}} \varphi$. □

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 4.3.3 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 4.4.3, $\vdash_{\mathsf{L_{top}}} \psi_0 \wedge \cdots \wedge \psi_n \to \varphi$. 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. 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 4.4.4. $\mathsf{L_{int-compl}}$ *is sound and strongly complete with respect to* $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$.

Proof. For soundness, we need to check that (5_{S}) is valid on $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$. 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 4.1.1 (4), we get $M, x \models \neg\mathsf{S}\varphi$.

The completeness proof goes in exactly the same way as Theorem 4.4.3. 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 4.4.3, 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 $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$. □

4.5 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 [38], and establish connections between collective expertise and collective knowledge.

4.5.1 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 [38]:

- $\mathsf{K}_j\varphi$ (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.

- $\mathsf{K}_J^{\mathsf{dist}}\varphi$ (distributed knowledge): for $J \subseteq \mathcal{J}$ non-empty, set

$$M^*, x \models \mathsf{K}_J^{\mathsf{dist}}\varphi \iff \forall y \in X : (x, y) \in \bigcap_{j\in J} R_j \implies M^*, y \models \varphi.$$

  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.

- $\mathsf{K}_J^{\mathsf{sh}}\varphi$ (shared knowledge):[12] for $J \subseteq \mathcal{J}$ non-empty, set

$$M^*, x \models \mathsf{K}_J^{\mathsf{sh}}\varphi \iff \forall j \in J : M^*, x \models \mathsf{K}_j\varphi.$$

  That is, a group J have shared knowledge of φ exactly when each agent in J knows φ. Thus we have $\mathsf{K}_J^{\mathsf{sh}}\varphi \equiv \bigwedge_{j\in J} \mathsf{K}_j\varphi$.

- $\mathsf{K}_J^{\mathsf{com}}\varphi$ (common knowledge): write $\mathsf{K}_J^1\varphi$ for $\mathsf{K}_J^{\mathsf{sh}}\varphi$, and for $n \in \mathbb{N}$ write $\mathsf{K}_J^{n+1}\varphi$ for $\mathsf{K}_J^{\mathsf{sh}}\mathsf{K}_J^n\varphi$. Then

$$M^*, x \models \mathsf{K}_J^{\mathsf{com}}\varphi \iff \forall n \in \mathbb{N} : M^*, x \models \mathsf{K}_J^n\varphi.$$

  Here $\mathsf{K}_J^1\varphi$ says that everyone in J knows φ, $\mathsf{K}_J^2\varphi$ 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 $\mathcal{L}_{\mathsf{KA}}^{\mathcal{J}}$ for the language formed from Prop with knowledge operators K_j, $\mathsf{K}_J^{\mathsf{dist}}$, $\mathsf{K}_J^{\mathsf{sh}}$ and $\mathsf{K}_J^{\mathsf{com}}$, for $j \in \mathcal{J}$ and $J \subseteq \mathcal{J}$ non-empty, and the universal modality A.

4.5.2 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 4.3.3) 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.

[12] In Fagin et al. [38], 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.

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_J^{dist}. A first candidate for P_J^{dist} 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_J^{dist} be the closure of $\bigcup_{j\in J} P_j$ under unions and intersections:

$$P_J^{\mathsf{dist}} = \bigcap \left\{ P' \supseteq \bigcup_{j\in J} P_j \mid P' \text{ is closed under unions and intersections} \right\}.$$

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.

P_J^{dist} also has a topological interpretation. As in Section 4.3, each P_j gives rise to an Alexandrov topology τ_j (where P_j are the closed sets) if it is closed under unions and intersections. By the aforementioned properties, τ_J^{dist} corresponds to the coarsest Alexandrov topology finer than each τ_j. On the other hand, since the join (in the lattice of topologies on X) of finitely many Alexandrov topologies is again Alexandrov [88, Theorems 2.4, 2.5], it follows that τ_J^{dist} is equal to the join $\bigvee_{j\in J} \tau_j$.

Now, recall from Theorem 4.3.3 that our semantics for expertise and soundness is connected to relational semantics via the mapping $P \mapsto R_P$ (Definition 4.3.2). 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 4.5.1. *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 4.3.3. Indeed, shared knowledge may fail positive introspection (axiom **4**: $\mathsf{K}\varphi \to \mathsf{KK}\varphi$), 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 4.5.2. *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 4.5.2 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. [38]. In what follows, write $R^+ = \bigcup_{n\in\mathbb{N}} R^n$ for the transitive closure of R.

Lemma 4.5.1 (Fagin et al. [38], 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 4.5.1, 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 4.5.3. *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 4.3.2 (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 4.3.1 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 4.3.2 (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

$$\begin{aligned} M, x &\models E_j\varphi \iff \|\varphi\|_M \in P_j \\ M, x &\models E_J^g\varphi \iff \|\varphi\|_M \in P_J^g && (g \in \{\mathsf{dist}, \mathsf{com}\}) \\ M, x &\models S_j\varphi \iff \forall A \in P_J : \|\varphi\|_M \subseteq A \implies x \in A \\ M, x &\models S_J^g\varphi \iff \forall A \in P_J^g : \|\varphi\|_M \subseteq A \implies x \in A && (g \in \{\mathsf{dist}, \mathsf{com}\}) \end{aligned}$$

Note that expertise and soundness are interpreted as before, but with respect to different collections P. Consequently, the interactions shown in Proposition 4.1.1 still hold for individual and collective notions of expertise and soundness.

Example 4.5.1. *Extending Examples 4.1.1 and 4.1.2, 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 4.1.2,*

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 4.1.1. 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 4.5.4. *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

$$\begin{array}{lll} t(\mathsf{E}_j\varphi) & = \mathsf{A}(\neg t(\varphi) \to \mathsf{K}_j \neg t(\varphi)) & \\ t(\mathsf{E}^g_J\varphi) & = \mathsf{A}(\neg t(\varphi) \to \mathsf{K}^g_J \neg t(\varphi)) & (g \in \{\mathsf{dist}, \mathsf{com}\}) \\ t(\mathsf{S}_j\varphi) & = \neg\mathsf{K}_j \neg t(\varphi) & \\ t(\mathsf{S}^g_J\varphi) & = \neg\mathsf{K}^g_J \neg t(\varphi) & (g \in \{\mathsf{dist}, \mathsf{com}\}) \end{array}$$

where the other cases are as for t in Section 4.3. 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 4.3.3.

Theorem 4.5.1. *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 4.3.1 and 4.3.2. For the stated property we proceed by induction on $\mathcal{L}^{\mathcal{J}}$ formulas. As in Theorem 4.3.3, the cases for atomic propositions, propositional connectives and A are straightforward. For expertise and soundness, the argument in the proof of Theorem 4.3.3 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 4.5.1. For common expertise this was shown in Lemma 4.5.1 and Proposition 4.5.3. □

Theorem 4.5.1 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 4.4.3 and 4.4.4.

4.6 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.

4.6.1 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 4.2, 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 4.2 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 4.6.1. *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 $M, 1 \not\models [+\varphi]\mathsf{E}\varphi$.[13] 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 [7].

Next we give reduction axioms to express any formula involving $[+\varphi]$ by an equivalent formula in the static language $\mathcal{L}$.

Proposition 4.6.1. *The following formulas are valid on $\mathbb{M}$:*

$$\begin{aligned} [+\varphi]p &\leftrightarrow [+\varphi]p \\ [+\varphi](\psi \wedge \theta) &\leftrightarrow [+\varphi]\psi \wedge [+\varphi]\theta \\ [+\varphi]\neg\psi &\leftrightarrow \neg[+\varphi]\psi \\ [+\varphi]\mathsf{A}\psi &\leftrightarrow \mathsf{A}[+\varphi]\psi \\ [+\varphi]\mathsf{S}\psi &\leftrightarrow \mathsf{S}[+\varphi]\psi \wedge (\mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi) \\ [+\varphi]\mathsf{E}\psi &\leftrightarrow \mathsf{A}((\mathsf{S}[+\varphi]\psi \wedge (\mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi)) \to [+\varphi]\psi) \end{aligned}$$

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 $x \in X$.

[13] 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 $M^{+\varphi}, 1 \not\models \mathsf{E}\varphi$.

"→": 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.

"←": 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. □

Note that only the reduction axiom for $[+\varphi]\mathsf{E}\psi$ requires $M^{+\varphi}$ to be closed under intersections.

4.6.2 Sound Announcements

In logics of public announcement [74, 28], 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 4.6.2. *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 4.6.1. *The report of the economist in Example 4.1.1 can be modelled as* $[?(p \wedge \neg d)]$. *Note that, with M as in Example 4.1.2, $\|\mathsf{S}(p \wedge \neg d)\|_M = \|p\|_M$. The updated model $M^{?(p\wedge\neg d)}$ therefore consists only of the bottom half of M as shown in Fig. 4.1. 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 source 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 4.6.2. *The following formulas are valid on $\mathbb{M}$:*

$$\begin{aligned} [?\varphi]p &\leftrightarrow \mathsf{S}\varphi \to p \\ [?\varphi](\psi \wedge \theta) &\leftrightarrow [?\varphi]\psi \wedge [?\varphi]\theta \\ [?\varphi]\neg\psi &\leftrightarrow \mathsf{S}\varphi \to \neg[?\varphi]\psi \\ [?\varphi]\mathsf{A}\psi &\leftrightarrow \mathsf{S}\varphi \to \mathsf{A}[?\varphi]\psi \\ [?\varphi]\mathsf{S}\psi &\leftrightarrow \mathsf{S}\varphi \to \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi) \end{aligned}$$

and the following is valid on $\mathbb{M}_{\mathsf{int}}$:

$$[?\varphi]\mathsf{E}\psi \leftrightarrow \mathsf{S}\varphi \to \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$$

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 4.6.3. *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 4.1.1 (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).

□

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 4.1.1 (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.

4.7 Conclusion

Summary. This chapter presented a modal logic framework to reason about the expertise of information sources and soundness of information. 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. The framework was then extended to handle multiple sources, permitting the study of several notions of collective expertise. Finally, we considered dynamic operators to model evolving expertise and sound announcements.

Limitations and future work. On a technical level, some open questions remain. For example, can frames closed under arbitrary unions be be expressed in our language, as other closure properties were expressed in Proposition 4.2.1? Similarly, can one axiomatise the class of models closed under arbitrary unions, without also requiring closure under intersections? One could also consider computational properties, such as decidability and the complexity of the satisfiability problem.

There are also conceptual limitations and areas for future study. Firstly, our notion of expertise is absolute: either the source is an expert on φ or they are not. In reality things are more nuanced, and source may have varying levels of expertise. Our assumption that expertise is independent of the actual state of the world could

also be considered too strong, since it forbids any possibility of *context-dependent* expertise. As a somewhat contrived example, the economist in our running examples may have expertise on p in ordinary times, but not if they are suffering from the virus which affects cognitive ability.

Outlook. Equipped with the notions of expertise and soundness from this chapter, the following chapter poses a belief change problem – in the style of AGM revision [1] and belief merging [61] – in which expertise is not assumed to be known upfront, but must be estimated from a sequence of reports. For simplicity we dispense with some of the generality of this chapter, by (i) considering only finite models whose states are the propositional valuations over a fixed, finite set of variables; and (ii) assuming expertise collections are closed under both intersections and complements. By Theorem 4.3.2, such models are in one-to-one correspondence with S5 relational models, so that their corresponding binary relation R_P is an equivalence relation. Equivalently, each collection P closed under intersections and complements corresponds to a partition Π_P over the set of states, whose cells are simply the equivalence classes of R_P. Since one can express the semantic conditions for expertise and soundness directly in terms of this partition, in what follows we in fact take the partition Π_P as primitive instead of the expertise collection P. Given that the equivalence relation R_P corresponding to Π_P can be understood as an epistemic accessibility relation (by Theorem 4.3.3), we can interpret Π_P as expressing an *indistinguishibility relation* over states: two states lie in the same partition cell if the source lacks expertise to distinguish them.

5 Belief Change with Non-Expert Sources

In the previous chapter we introduced a logical framework to reason about the expertise of sources and soundness of information. We now build on this framework to study a belief change problem in which expertise is not fixed upfront, but is to be estimated on the basis of reports from multiple sources. In this way we develop a logic-based analogue of the truth discovery aggregation problem, which complements the social-choice-style framework of Chapter 2. Specifically, we identify trustworthiness with *belief in expertise*: a source i is deemed trustworthy on φ if the belief change operator includes $\mathsf{E}_i\varphi$ in its belief set. By also including propositional formulas in belief sets, we are able to express beliefs about the actual world, i.e. the aggregation of the reports from sources.

Our point of departure is logic-based belief change in the AGM paradigm [1]. To illustrate the problem – and how it differs from existing forms of belief change – consider the following scenario in a hospital. Suppose we observe the results of a blood test of patient 1, confirming condition X. Assuming the test is reliable, AGM revision 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 [50].

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. This is also analogous to *semi-supervised truth discovery* [101, 75], in which some ground truths are known ahead of time. In some respects we also extend approaches to non-prioritised revision (e.g. selective revision [39], credibility-limited revision [51], and trust-based revision [12]), which assume information about the reliability of sources is known up front. The problem

is also related to *belief merging* [61] which deals with combining belief bases from multiple sources; a detailed comparison will be given in Section 5.6.

Our work is also connected to trust and belief revision. As Yasser and Ismail [99] 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.

To unify the trust and belief aspects, we work in (a fragment of) the language of expertise and soundness from Chapter 4, including formulas of the form $\mathsf{E}_i\varphi$, read as "source i has expertise on φ", and $\mathsf{S}_i\varphi$, read as "φ is sound for source i to report". The output of our belief change problem is then a collection of belief and knowledge sets in the 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.

We do, however, make some simplifying assumptions compared to the modal framework in Chapter 4. Firstly, we only consider only a finite set of propositional variables, and identify states of the world with propositional valuations. Secondly, we assume expertise is closed under both intersections and complements, so that – by Theorem 4.3.2 – expertise of a source is fully captured by an equivalence relation; or equivalently, a *partition* of the propositional valuations. In other words, each source has a *indistinguishibility relation* over valuations, whereby any two valuations in the same partition cell are indistinguishable.

The semantics of expertise and soundness can be expressed directly in terms of partitions; we have that a source is an expert on a proposition φ exactly when they can distinguish every φ valuation from every $\neg\varphi$ valuation, and φ is sound for i if the "actual" state of the world is indistinguishable from a φ valuation.

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 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 [53]). 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 as per the dynamics of Section 4.6).

Contributions. The main contribution of this chapter is the formulation of a belief change problem in the setting of the logic of expertise developed in the previous chapter. This allows us to explore how belief and expertise-based trust should interact and evolve as reports are received from the various sources. 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 5.1 we set out the formal framework. Section 5.2 introduces the problem and lists some core postulates. We give two constructions and specific example operators in Section 5.3. Section 5.4 introduces some further postulates concerning belief change on the basis of one new report. An analogue of selective revision [39] is presented Section 5.5. Section 5.6 discusses related work, and we conclude in Section 5.7.

5.1 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. In this chapter we work with the fragment of the language from Chapter 4, in which expertise and soundness formulas are restricted to propositional formulas only[1] and the universal modality A is excluded. Concretely, we assume a fixed *finite* set Prop of propositional variables, and let $\mathcal{L}_0$ denote the set of propositional formulas generated from Prop using the usual propositional connectives. Formulas in $\mathcal{L}_0$ are used to describe properties of the world in each case $c \in \mathcal{C}$. 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:

$$\Phi ::= p \mid \Phi \wedge \Phi \mid \neg\Phi \mid \mathsf{E}_i\varphi \mid \mathsf{S}_i\varphi$$

where $i \in \mathcal{S}$, $p \in \mathsf{Prop}$ 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.

[1] While this assumption is made for simplicity's sake, we do not lose much by excluding iterated applications of E and S, at least for expertise models closed under intersections and complements. Indeed, we have that $\mathsf{E}\varphi$ either holds globally in a model or holds nowhere, so $\mathsf{EE}\varphi$ always holds. One can show that $\mathsf{ES}\psi$ also always holds in such models, by taking $\varphi = \mathsf{S}\varphi$ in Proposition 4.2.1 (1) and recalling that $\mathsf{SS}\psi \rightarrow \mathsf{S}\psi$ is valid. Similarly, one can show that $\mathsf{SS}\varphi \leftrightarrow \mathsf{S}\varphi$ and $\mathsf{SE}\varphi \leftrightarrow \mathsf{E}\varphi$ in such models.

As before, the intuitive reading of $\mathsf{E}_i\varphi$ is source i has expertise on φ. 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. Since both operators are restricted to propositional formulas, we will not consider iterated formulas such as $\mathsf{E}_i\mathsf{S}_j\varphi$.

Semantics. The semantics in this chapter are, in essence, a special case of Chapter 4. Instead of considering arbitrary sets of possible states, as in Definition 4.1.1, we fix states as propositional valuations over Prop. Expertise is also assumed to be closed under both intersections and complements for all sources. At the same time, we generalise slightly by considering the distinguished source $*$ and multiple cases $c \in \mathcal{C}$. For convenience we also offer a different presentation, using *partitions* instead of expertise collections to represent expertise.

Formally, let $\mathcal{V}$ denote the set of propositional valuations over Prop. For each $\varphi \in \mathcal{L}_0$, the set of valuations making φ true is denoted by $\|\varphi\|$. 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 $\Pi[\|\varphi\|]$. 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:

$$\begin{array}{lll} W, c \models p & \iff & v_c \in \|p\| \\ W, c \models \mathsf{E}_i\varphi & \iff & \Pi_i[\varphi] = \|\varphi\| \\ W, c \models \mathsf{S}_i\varphi & \iff & v_c \in \Pi_i[\varphi] \end{array}$$

where $i \in \mathcal{S}$, $\varphi \in \mathcal{L}_0$, and the clauses for conjunction and negation are the expected ones. Since $\|\varphi\| \subseteq \Pi_i[\varphi]$ 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 [12]). 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*,[2] so can be considered a closure operator of the set of propositional valuations. Then $W, c \models \mathsf{E}_i\varphi$ iff $\|\varphi\|$ is closed in V, and $W, c \models \mathsf{S}_i\varphi$ iff v_c lies in

the closure of $\|\varphi\|$, i.e. φ is true after closing $\|\varphi\|$ 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.

The semantics so defined are indeed the same as those of Chapter 4, as the following result shows.

Proposition 5.1.1. *Let W be a world. Then there is a multi-source expertise model $M = (X, \{P_i\}_{i\in\mathcal{S}}, V)$ and $\{x_c\}_{c\in\mathcal{C}} \subseteq X$ such that for all $\Phi \in \mathcal{L}$ and $c \in \mathcal{C}$,*

$$W, c \models \Phi \iff M, x_c \models \Phi. \tag{5.1}$$

Moreover, (i) X is finite; (ii) each P_i is closed under intersections and complements; and (iii) using the notation from Definition 4.3.2, $uR_{P_i}v$ iff $\Pi_i[u] = \Pi_i[v]$, i.e. R_{P_i} is the equivalence relation associated with the partition Π_i.

Proof. Take $X = \mathcal{V}$ and set $V(p) = \|p\|$. For each $i \in \mathcal{S}$, set $P_i = \{A \subseteq \mathcal{V} \mid \Pi_i[A] = A\}$. For each $c \in \mathcal{C}$, simply let $x_c = v_c$. Then one can easily show that for all $U \subseteq \mathcal{V}$ and $i \in \mathcal{S}$,

$$\Pi_i[U] = \bigcap\{A \in P_i \mid U \subseteq A\}. \tag{5.2}$$

A simple induction on $\mathcal{L}$ formulas then shows (5.1).

Since Prop is finite there are only finitely many propositional valuations, and thus $X = \mathcal{V}$ is finite. It is easily checked that each P_i is closed under intersections and complements using properties of partitions. Finally, we have by (5.2) and the definition of R_{P_i} that

$$\begin{aligned} uR_{P_i}v &\iff \forall A \in P_i : (v \in A \implies u \in A) \\ &\iff u \in \bigcap\{A \in P_i \mid v \in A\} \\ &\iff u \in \bigcap\{A \in P_i \mid \{v\} \subseteq A\} \\ &\iff u \in \Pi_i[\{v\}] \\ &\iff \Pi_i[u] = \Pi_i[v] \end{aligned}$$

as required. □

In other words, a world corresponds to a particular kind of expertise model together with a state x_c for each case $c \in \mathcal{C}$. Having shown this equivalence, we henceforth deal exclusively with worlds and models instead of expertise models and collections. We come to an example.

Example 5.1.1. *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 $\mathsf{Prop} = \{x, y\}$, 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*

[2] That is, (i) $\Pi[\emptyset] = \emptyset$, (ii) $U \subseteq \Pi[U]$, (iii) $\Pi[\Pi[U]] = \Pi[U]$, and (iv) $\Pi[U_1 \cup U_2] = \Pi[U_1] \cup \Pi[U_2]$.

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 $W = \langle \{v_c\}_{c\in\{c_1,c_2\}}, \{\Pi_i\}_{i\in\{*,a,b\}}\rangle$*:*

$$\begin{array}{ll} v_{c_1} = xy; & v_{c_2} = \bar{x}y; \\ \Pi_a = xy, \bar{x}y \mid x\bar{y}, \bar{x}\bar{y}; & \Pi_b = xy, x\bar{y} \mid \bar{x}y, \bar{x}\bar{y}. \end{array}$$

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* $\Pi_b[\neg x] = \{\bar{x}y, \bar{x}\bar{y}\} = \|\neg x\| \not\ni xy = v_{c_1}$*).*

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.[3]

Proposition 5.1.2. *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* $\mathsf{E}_i\varphi \wedge \mathsf{E}_i\psi \rightarrow \mathsf{E}_i(\varphi \wedge \psi)$
3. $\mathsf{E}_i p_1 \wedge \cdots \wedge \mathsf{E}_i p_k \rightarrow \mathsf{E}_i\varphi$*, where* $p_1, \ldots, p_k$ *are the propositional variables appearing in* φ
4. $\mathsf{E}_i\varphi \wedge \mathsf{S}_i\varphi \rightarrow \varphi$*, and* $\mathsf{S}_i\varphi \wedge \neg\varphi \rightarrow \neg\mathsf{E}_i\varphi$
5. $\mathsf{S}_i\varphi \wedge \mathsf{S}_i\neg\varphi \rightarrow \neg\mathsf{E}_i\varphi$
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 $\|\varphi\| = \|\psi\|$; since the semantics for $\mathsf{S}_i\varphi$ and $\mathsf{E}_i\varphi$ only refer to $\|\varphi\|$ (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 $\|\varphi\| = \Pi_i[\varphi]$. We show $W, c \models \mathsf{E}_i\neg\varphi$. Indeed, take $v \in \Pi_i[\neg\varphi]$. Then there is $v' \in \|\neg\varphi\|$ such that $v \in \Pi_i[v']$. Thus $v' \in \Pi_i[v]$ also. Supposing for contradiction that $v \in \|\varphi\|$, we get
$$v' \in \Pi_i[v] \subseteq \Pi_i[\varphi] = \|\varphi\|.$$

[3]Note that some of these validities follow from Proposition 5.1.1 and the validities in Chapter 4.

But then $v' \in \|\neg\varphi\| \cap \|\varphi\| = \emptyset$; contradiction. Hence $v \notin \|\varphi\|$, i.e. $v \in \|\neg\varphi\|$. This shows that $\Pi_i[\neg\varphi] \subseteq \|\neg\varphi\|$, 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

$$\Pi_i[\varphi \wedge \psi] \subseteq \Pi_i[\varphi] = \|\varphi\|$$

and, similarly, $\Pi_i[\varphi \wedge \psi] \subseteq \|\psi\|$. Hence

$$\Pi_i[\varphi \wedge \psi] \subseteq \|\varphi\| \cap \|\psi\| = \|\varphi \wedge \psi\|,$$

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 $v_c \in \Pi_i[\varphi] = \|\varphi\|$, 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 $\Pi_*[\varphi] = \|\varphi\|$, so $W, c \models \mathsf{E}_*\varphi$. Similarly, $W, c \models \mathsf{S}_i\varphi$ iff $v_c \in \Pi_*[\varphi] = \|\varphi\|$ iff $W, c \models \varphi$.

□

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 5.1.2. *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 5.1.2 part (3),* $W, c_1 \models \mathsf{E}_i(p \wedge q)$. *Since* W *satisfies* Γ_{c_1} *in case* c_1, *Proposition 5.1.2 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 $\mathrm{Cn}_0([\Gamma_c]) = [\Gamma_c]$, i.e. the propositional parts of G are (classically) closed.

In propositional logic, $\|\cdot\|$ 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 [24], 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 5.1.3. $S \subseteq \mathcal{W}$ *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 $\|\varphi_U\| = U$. 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

$$R_{W',i,U} = \begin{cases} \mathsf{E}_i\varphi_U, & W', c_0 \models \mathsf{E}_i\varphi_U \\ \neg\mathsf{E}_i\varphi_U, & \text{otherwise} \end{cases}$$

[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 [24]) 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.

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) $v_c^W = v_c^{W_c}$; 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. □

5.2 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* $K_c^\sigma \subseteq \mathcal{L}$.

5.2.1 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 $[B_c^\sigma]$ are closed under (propositional) consequence, and that $\mathsf{E}_i\varphi \in B_c^\sigma$ iff $\mathsf{E}_i\varphi \in B_d^\sigma$. **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 5.1.2 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 [77]. 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 5.2.1. *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.

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.

[6] This argument is from [23].

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. $K^{\sigma\cdot\rho} = \mathrm{Cn}(K^\sigma \sqcup K^\rho)$

This postulate 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 5.2.2. *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 5.2.1,

$$K^{\sigma\cdot\rho} = \mathrm{Cn}(G^{\sigma\cdot\rho}_{\mathsf{snd}} \sqcup K^\emptyset)$$

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 5.2.1 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 5.2.3. *Suppose an operator satisfies the basic postulates. Then*

1. *If* $\varphi \in K_c^\sigma$ *and* $\neg\psi \in \mathrm{Cn}_0(\varphi)$ *then* $\neg \mathsf{E}_i\psi \in K_c^{\sigma\cdot\langle i,c,\psi\rangle}$.
2. *If* $\langle i, c, \varphi\rangle \in \sigma$ *and* $\mathsf{E}_i\varphi \in B_c^\sigma$ *then* $\varphi \in B_c^\sigma$.

Proof.

1. Suppose $\varphi \in K_c^\sigma$ and $\neg\psi \in \mathrm{Cn}_0(\varphi)$. Write $\rho = \sigma \cdot \langle i, c, \psi\rangle$. By **Soundness**, $\mathsf{S}_i\psi \in K_c^\rho$. By **K-conjunction**, $\varphi \in K_c^\sigma \subseteq (K^\sigma \sqcup K^{\langle i,c,\psi\rangle})_c \subseteq \mathrm{Cn}_c(K^\sigma \sqcup K^{\langle i,c,\psi\rangle}) = K_c^\rho$. Since $\neg\psi \in \mathrm{Cn}_0(\varphi)$ and $\varphi \in K_c^\rho$, **Closure** gives $\neg\psi \in K_c^\rho$. Recalling from Proposition 5.1.2 part (4) that $\mathsf{S}_i\psi \wedge \neg\psi \to \neg\mathsf{E}_i\psi$, **Closure** gives $\neg\mathsf{E}_i\psi \in K_c^\rho$, as desired.
2. Suppose $\langle i, c, \varphi\rangle \in \sigma$ and $\mathsf{E}_i\varphi \in B_c^\sigma$. By **Soundness** and **Containment**, $\mathsf{S}_i\varphi \in B_c^\sigma$. From Proposition 5.1.2 part (4) again we have $\mathsf{E}_i\varphi \wedge \mathsf{S}_i\varphi \to \varphi$. By **Closure**, $\varphi \in B_c^\sigma$.

□

(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 5.1.2 part (6) and **Closure**). We illustrate the basic postulates by formalising the introductory hospital example.

Example 5.2.1. *Set* $\mathcal{S}, \mathcal{C}$ *and* Prop *as in Example 5.1.1, 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 5.1.2 part (6) and* ***Closure****, the report from* $*$ *gives* $x \in K_{c_1}^\sigma$*, i.e. reliable reports are known.* ***Soundness*** *also gives* $\mathsf{S}_a x \wedge \mathsf{S}_b\neg x \in K_{c_2}^\sigma$*. Combined with Proposition 5.1.2 parts (2), (4) and* ***Closure****, this yields* $\neg(\mathsf{E}_a x \wedge \mathsf{E}_b x) \in K_c^\sigma$ *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* $x \wedge \mathsf{S}_a\neg x \in K_{c_1}^\sigma$*, and thus* $\neg\mathsf{E}_a x \in K_c^\sigma$ *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.* $\mathsf{E}_a x \vee \mathsf{E}_b x \in B_c^\sigma$*). Combined with* $\neg\mathsf{E}_a x \in K_c^\sigma$*,* ***Closure*** *and* ***Containment****, we get* $\mathsf{E}_b x \in B_{c_2}^\sigma$*. Symmetry of expertise together with Proposition 5.2.3 part (2) then gives* $\neg x \in B_{c_2}^\sigma$*, i.e. we trust Dr. B in the example and believe patient 2 does not suffer from condition X.*

5.2.2 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 5.2.1. *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 5.2.1. *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 $\Phi \in K^\sigma_c$ 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 $\Phi \in B^\sigma_c$, 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 5.2.1. $\mathcal{X}_\sigma \supseteq \mathcal{Y}_\sigma$ follows from **Containment** and the definition of a model of a collection. For the second property, note that $\Phi \in K^\sigma_c$ 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 5.2.2. *Define the model-based operator* weak-mb *by*

$$\mathcal{X}_\sigma = \mathcal{Y}_\sigma = \{W \mid W, c \models \mathsf{S}_i\varphi \text{ for all } \langle i, c, \varphi \rangle \in \sigma\}.$$

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 5.2.1.[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 5.2.2. *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 5.1.2, 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* $B_c^\sigma = K_c^\sigma$.

5.3 Constructions

For model-based operators in Definition 5.2.1, 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.

5.3.1 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 5.3.1. *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 5.2.1 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 5.3.2. *Define the conditioning operator* var-based-cond *by setting* $\mathcal{X}_\sigma$ *in the same way as* weak-mb *in Definition 5.2.2, and* $W \leq W'$ *iff* $r(W) \leq r(W')$, *where*

$$r(W) = -\sum_{i \in \mathcal{S}} \left| \{ p \in \mathsf{Prop} \mid \Pi_i^W[p] = \|p\| \} \right|.$$

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 5.3.1. *Revisiting the sequence* $\sigma = (\langle *, c, p\rangle, \langle i, c, \neg p \wedge q\rangle)$ *from Example 5.2.2 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* $\leq$:

[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] A total preorder is a reflexive, transitive and total relation.

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 5.3.3. *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.

Example 5.3.2. *Applying* part-based-cond *to σ from Examples 5.2.2 and 5.3.1, 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 5.2.1, conditioning operators can be characterised using an approach similar to that of Delgrande, Peppas, and Woltran [24] 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 [62].

> **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

[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.

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 5.3.1 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]

Theorem 5.3.1. *Suppose an operator satisfies the basic postulates of Section 5.2.1.*[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 [24], although some differences arise due to the form of our input as finite sequences of reports. First, we need a preliminary result.

Lemma 5.3.1. *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 $\Phi \in K_c^\sigma$ iff $\mathcal{X}_\sigma \subseteq \mathrm{mod}_c(\Phi)$, we have $K_c^\sigma = \mathrm{Cn}_c(G)$, 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 5.3.1 – from which Theorem 5.3.1 immediately follows.

Proposition 5.3.1. *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 5.2.1 – our operator is model based corresponding to this choice of $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$. Clearly both are

[11] Equivalently, there is a total preorder $\leq$ such that $\mathrm{mod}(B^\sigma) = \min_{\leq} \mathrm{mod}(K^\sigma)$ for all σ.

[12] Strictly speaking, we only need **Closure**, **Containment**, **K-conjunction**, **Equivalence** and **Rearrangement**.

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 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 5.3.1. *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 $W_1, W_2 \in \mathcal{X}_{\rho(W_1,W_2)}$. □

Claim 5.3.2. *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 5.3.1, 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 5.3.1 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. □

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 5.3.3. *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 5.3.2 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 5.3.3, every world in $\mathcal{Y}_\sigma$ is $\leq_0$-minimal in $\mathcal{X}_\sigma$. In fact, the converse is also true.

Claim 5.3.4. *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 5.3.3, $W'RW$ 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 $W_0, \ldots W_n$ 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

$$\begin{aligned} \rho_j &= \rho(W_j, W_{j+1}) \qquad (0 \leq j < n) \\ \rho_n &= \rho(W_0, W_n) \end{aligned}$$

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

$$W_{j+1} \in \mathcal{X}_{\rho_j} \cap \mathcal{Y}_{\rho_{j+1}}$$

[13] If $\rho_j \neq \emptyset$ this follows from Claim 5.3.1. 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**.

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 5.3.2 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. □

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 5.3.5. *For any sequence σ, $\mathcal{Y}_\sigma = \min_\leq \mathcal{X}_\sigma$*

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 5.3.3, 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 5.3.4 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 5.2.1. By elementariness and Lemma 5.3.1, $\mathcal{X}_\sigma = \mathrm{mod}(K^\sigma)$ and $\mathcal{Y}_\sigma = \mathrm{mod}(B^\sigma)$.

The following claim will be useful at various points.

[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.

Claim 5.3.6. *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 5.3.6.

- **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 5.3.6.

- **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 5.3.1 that $\mathcal{X}_\sigma$ and $\mathcal{Y}_\sigma$ are elementary is a technical condition,[15] the characterisation in Proposition 5.1.3 implies a simple sufficient condition for elementariness.

Proposition 5.3.2. *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 5.1.3. 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. □

Proposition 5.3.2 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.

[15] **Inclusion-vacuity** may fail for non-elementary conditioning.

5.3.2 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 5.2.1), conditioning gives $p \notin B_d^\sigma$ and $\neg p \notin B_d^\sigma$. 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 $p \in B_d^\sigma$.[16] This shows that we need the plausibility order to be responsive to the input sequence for adequate belief change.[17]

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 5.3.4. *An operator is* score-based *if there is a mapping $\sigma \mapsto \langle \mathcal{X}_\sigma, \mathcal{Y}_\sigma\rangle$ as in Definition 5.2.1 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 5.3.3. *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

$$k(W) = |\{W' \in \mathcal{W} \mid W' \leq W\}|$$

[16] At the very least, the case $p \in B_d^\sigma$ should not be *excluded*.

[17] In Section 5.4 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.

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 5.3.5. *Define a score-based operator* excess-min *by setting* $r_0(W) = 0$ *and*

$$d(W, \langle i, c, \varphi \rangle) = \begin{cases} |\Pi_i^W[\varphi] \setminus \|\varphi\||, & W, c \models \mathsf{S}_i\varphi \\ \infty, & \textit{otherwise.} \end{cases}$$

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 5.2.1. 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 5.3.1 that excess-min is *not* a conditioning operator.[18]

Example 5.3.3. *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 5.1.2 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.*

[18] We will later give an alternative proof of this fact, via an impossibility result for conditioning operators (Proposition 5.4.3).

5.4 One-Step Revision

The postulates of Section 5.2.1 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.

> **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 [49, 56], it is no surprise that our order-based constructions are consistent with **AGM-$*$**.

Proposition 5.4.1. 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 5.4.1. *For any model-based operator, sequence σ, case c, and valuation v in$\mathcal{V}$,*

$$v \in \|[B_c^\sigma]\| \iff \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v} \neq \emptyset$$

Proof. $\Longrightarrow$: We show the contrapositive. Suppose $\mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v} = \emptyset$. Let ψ be any propositional formula such that $\|\psi\| = \mathcal{V} \setminus \{v\}$. Now for any $W \in \mathcal{Y}_\sigma$, we have $W \notin \mathcal{W}_{c\,:\,v}$, i.e. $v_c^W \neq v$. Hence $v_c^W \in \|\psi\|$, 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 $v \notin \|\psi\|$, we have $v \notin \|[B_c^\sigma]\|$.

$\Longleftarrow$: 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, $v = v_c^W \in \|\varphi\|$. Since $\varphi \in [B_c^\sigma]$ was arbitrary, we have $v \in \|[B_c^\sigma]\|$. □

We have a sufficient condition for **AGM-$*$** for score-based operators.

Lemma 5.4.2. *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 $K_c^\sigma = \mathcal{L}$. But since K_c^σ 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 $[B_c^\sigma]$ by

$$[B_c^\sigma] \star \varphi = \{\psi \in \mathcal{L}_0 \mid \min_\preceq \|\varphi\| \subseteq \|\psi\|\}$$

To show that $\star$ satisfies the AGM postulates (for $[B_c^\sigma]$) it is sufficient to show that the models of $[B_c^\sigma]$ are exactly the $\preceq$-minimal valuations.

Claim 5.4.1. $\|[B_c^\sigma]\| = \min_\preceq \mathcal{V}$.

Proof. "$\subseteq$": let $v \in \|[B_c^\sigma]\|$. By Lemma 5.4.1, 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 $\hat{v} = v_c^{\hat{W}}$, 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 $v \preceq v_c^{W'}$, so

$$r_\sigma(W) = k(v) \leq k(v_c^{W'}) \leq r_\sigma(W')$$

That is, $W \in \mathcal{Y}_\sigma \cap \mathcal{W}_{c\,:\,v}$. By Lemma 5.4.1, $v \in \|[B_c^\sigma]\|$. □

So, $\star$ is indeed an AGM operator for $[B_c^\sigma]$. Now take $\varphi \in \mathcal{L}_0$ such that $\neg\varphi \notin K_c^\sigma$. Write $\rho = \sigma \cdot \langle *, c, \varphi\rangle$. We claim the following.

Claim 5.4.2. $\|[B_c^\rho]\| = \min_\preceq \|\varphi\|$.

Proof. "$\subseteq$": let $v \in \|[B_c^\rho]\|$. By Lemma 5.4.1 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 $v = v_c^W \in \|\varphi\|$.

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 $v' \in \|\varphi\|$. 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 $v_c^{W'} = v' \in \|\varphi\|$, 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 $v \in \min_\preceq \|\varphi\|$. 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 $\hat{v} \in \|\varphi\|$. 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 $v_c^W = v \in \|\varphi\|$, 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, $v_c^{W'} \in \|\varphi\|$. 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 5.4.1, we are done. □

Noting that $\|[B_c^\sigma]\star\varphi = \min_\preceq \|\varphi\|$, it follows from Claim 5.4.2 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 5.3.3 (and the construction of d in its proof), one can apply Lemma 5.4.2 with $M = 0$ for conditioning operators with **K-conjunction** and a certain natural property.

Corollary 5.4.1. *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 5.4.1.

Proof of Proposition 5.4.1. For the conditioning operators var-based-cond and part-based-cond, it is easily verified that the condition in Corollary 5.4.1 holds, and thus **AGM-$*$** does also. For the score-based operator excess-min, we may use Lemma 5.4.2 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 5.2.2 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.

Proposition 5.4.2. 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 5.4.3. *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 5.1.2 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 5.4.4. *Suppose a score-based operator is such that for any $i \in \mathcal{S}$, $c \in \mathcal{C}$, $\varphi \in \mathcal{L}_0$ and $W, W' \in \mathcal{W}$,*

$$\begin{aligned} W, c \models \mathsf{E}_i\varphi \wedge \varphi \text{ and } W', c \models \neg\varphi \\ \implies d(W, \langle i, c, \varphi\rangle) < d(W', \langle i, c, \varphi\rangle) \end{aligned}$$

*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 5.4.2. For the conditioning operators var-based-cond and part-based-cond, **Cond-success** follows from Lemma 5.4.3 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 5.4.4 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 5.4.3. *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 5.3.1, 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 $\|\varphi_k\| = v_k$ ($k \in \{1, 2\}$). 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 5.4.3. $W_1 \simeq W_2$.

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 5.2.1 we have $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$. By elementariness and Lemma 5.3.1, 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 5.4.4. $W_1, W_2 \in \mathcal{Y}_\sigma$.

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 $\|\varphi_1\| = \{v_1\}$, 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] [**TODO:** Proposition 23 from online notes] □

Proposition 5.4.3 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**.

5.5 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* [39], in which propositional beliefs are formed by "selecting" only a part of each input report (e.g., some part consistent with the source's expertise). For example, in Example 5.3.1 we saw that when given $\sigma = (\langle *, c, p\rangle, \langle i, c, \neg p \wedge q\rangle)$, var-based-cond outputs propositional beliefs $[B^\sigma_c] = \mathrm{Cn}_0(p \wedge q)$. Intuitively, the report from $*$ is taken as-is, whereas the report of $\neg p \wedge q$ from i is weakened to just q. The resulting formulas are combined conjunctively to form the propositional belief set. We formalise this idea via *selection schemes*. In what follows, write $\sigma \restriction c = \{\langle i, \varphi\rangle \mid \langle i, c, \varphi\rangle \in \sigma\}$ for the c-reports in σ.

Definition 5.5.1. *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 $c \in \mathcal{C}$,*

$$[B^\sigma_c] = \mathrm{Cn}_0(\{f_\sigma(i, c, \varphi) \mid \langle i, \varphi\rangle \in \sigma \restriction c\}).$$

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 $[B^\sigma_c] = \mathrm{Cn}_0(\emptyset)$, so selectivity already rules out non-tautological prior propositional beliefs. Also note that in the presence of **Closure**, **Containment** and **Soundness**, selectivity implies that $[B^\sigma_c] = [B^\rho_c]$, 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^σ_c. For any sequence σ and case c, write $\Gamma^\sigma_c = \{\varphi \in \mathcal{L}_0 \mid \exists i \in \mathcal{S} : \langle i, \varphi\rangle \in \sigma \restriction c\}$.

Boundedness. If σ is $*$-consistent, $[B^\sigma_c] \subseteq \mathrm{Cn}_0(\Gamma^\sigma_c)$

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 $\left[B^\emptyset_c\right] = \mathrm{Cn}_0(\emptyset)$. We have the following characterisation.

Theorem 5.5.1. *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

$$M_c = \|[B^\sigma_c]\|.$$

By **Boundedness**, we have $M_c \supseteq \|\Gamma^\sigma_c\|$. Now set

$$F_\sigma(i, c, \varphi) = \|\varphi\| \cup M_c.$$

Define a selection function f_σ by letting $f_\sigma(i, c, \varphi)$ be any formula with $\|f_\sigma(i, c, \varphi)\| = F_\sigma(i, c, \varphi)$. 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 $c \in \mathcal{C}$,

$$M_c = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i, c, \varphi).$$

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 $v \in \|\varphi\|$. This shows $v \in \|\Gamma_c^\sigma\|$. But $\|\Gamma_c^\sigma\| \subseteq M_c$ by **Boundedness**, so $v \in M_c$; contradiction.

From this we get

$$\begin{aligned}
\|[B_c^\sigma]\| &= M_c \\
&= \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i, c, \varphi) \\
&= \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} \|f_\sigma(i, c, \varphi\|) \\
&= \|\{f_\sigma(i, c, \varphi) \mid \langle i, \varphi\rangle \in \sigma \restriction c\}\|
\end{aligned}$$

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**. □

The characterisation in Theorem 5.5.1 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'. Recall that $\mathcal{W}_{c\,:\,v} = \{W \in \mathcal{W} \mid v_c^W = v\}$ denotes the set of worlds whose c valuation is v. We have the following.

Proposition 5.5.1. *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 5.5.1 is somewhat technical, it is implied by the very natural *partition-equivalence* property from Section 5.1. 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.

To prove Proposition 5.5.1, we first state some preliminary results.

Lemma 5.5.1. *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. $\|\varphi\| \cap \Pi_i^W[v_c^W] \neq \emptyset$. By refinement, $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. Hence $\|\varphi\| \cap \Pi_i^{W'}[v_c^{W'}] \neq \emptyset$, so $v_c^{W'} \in \Pi_i^{W'}[\varphi]$. That is, $W', c \models \mathsf{S}_i\varphi$. □

Lemma 5.5.2. *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 $W' \in \mathcal{W}$,*

$$W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}}) \iff W \text{ refines } W' \text{ at } c.$$

Proof. For a valuation $v \in \mathcal{V}$, let $\varphi(v)$ be a propositional formula such that $\|\varphi(v)\| = \{v\}$. 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.

$\Longrightarrow$: 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.

$\Longleftarrow$: 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 5.5.1 gives $W', c \models \mathsf{S}_i\varphi(v)$. Hence $W' \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$. □

Proof of Proposition 5.5.1. 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 5.5.1 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**, $\|[B_c^\sigma]\| \supseteq \|\Gamma_c^\sigma\|$.

Take any $v \in \|\Gamma_c^\sigma\|$. 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 5.2.1, elementariness and Lemma 5.3.1, 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 $\langle i, \varphi\rangle \in \sigma \restriction c$, so $\varphi \in \Gamma_c^\sigma$. Hence $v \in \|\Gamma_c^\sigma\| \subseteq \|\varphi\|$. 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 5.5.1, $W', d \models \mathsf{S}_i\varphi$ 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 5.4.1 that $v \in \|[B^\sigma_c]\|$, 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 5.5.2, 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 $\|\varphi\| = \{v\}$ and $\|\psi\| = \{v^W_c\}$. 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

$$\|[B^\rho_c]\| \supseteq \|\Gamma^\rho_c\| = \|\varphi \vee \psi\| = \{v, v^W_c\}.$$

In particular, $v \in \|[B^\rho_c]\|$. By Lemma 5.4.1, 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. □

5.5.1 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 5.5.2. *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

$$\Gamma^{\sigma,H}_c = \{\varphi \in \mathcal{L}_0 \mid \exists i \in \mathcal{S} : \langle i, \varphi \rangle \in \sigma \restriction c \text{ and } \forall d \in H : \langle i, \varphi \rangle \notin \sigma \restriction d\}.$$

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 5.5.1. *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 5.5.2. *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 $M_d = \|[B_d^\sigma]\|$. Note that with c_0 an arbitrary fixed case, and writing $F_\sigma(i, \varphi) = \|f_\sigma(i, c_0, \varphi\|)$, 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

$$M_c \supseteq \|\Gamma_c^{\sigma,H}\| \cap \bigcap_{d \in H} M_d. \tag{5.3}$$

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

$$\begin{aligned} v &\in \|\Gamma_c^{\sigma,H}\| \\ &\subseteq \|\varphi\| \\ &\subseteq \|f_\sigma(i, c, \varphi\|) \\ &= F_\sigma(i, \varphi) \end{aligned}$$

(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 (5.3), 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 $\|[B_c^\sigma]\|$. 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

$$F_\sigma(i, \varphi) = \|\varphi\| \cup \bigcup_{c \in \mathcal{C}(i,\varphi)} M_c.$$

Define f by letting $f_\sigma(i, c, \varphi)$ be any propositional formula with $\|f_\sigma(i, c, \varphi\|) = F_\sigma(i, \varphi)$. 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

$$H = \{d \in \mathcal{C} \mid v \in M_d\}.$$

Applying **H-Boundedness** and taking the models of both sides, we obtain

$$M_c \supseteq \|\Gamma_c^{\sigma,H}\| \cap \bigcap_{d \in H} M_d. \tag{5.4}$$

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 $v \in \|\varphi\|$. 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 $v \in \|\varphi\|$. Since φ was arbitrary, we have $v \in \|\Gamma_c^{\sigma,H}\|$. By (5.4) 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.

5.5.2 Expertise and Selectivity

In the existing literature on selective belief change (e.g. [39, 12]), 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 5.5.3. *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 $\langle i, c, \varphi\rangle \in \sigma$,*

$$\mathsf{E}_i f_\sigma(i, c, \varphi) \in B_c^\sigma.$$

That is, i is trusted on the weakened report $f_\sigma(i, c, \varphi)$ whenever i reports φ in case c in σ. 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

$$\sigma = (\langle i, c, p\rangle, \langle j, c, p\rangle, \langle i, d, p\rangle, \langle j, d, \neg p\rangle)$$

By **Soundness** and **Closure**, we cannot have both $\mathsf{E}_i p$ and $\mathsf{E}_j p$ in B_c^σ. Ideas of symmetry 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 5.2.1. 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**. This example already shows that var-based-cond, part-based-cond and excess-min are not EC-selective.

The core issue here is that, unlike in ealier work on trust-based selective revision, the expertise of sources is part of the operator's output and is thus uncertain. In order for $\mathsf{E}_i\varphi$ to be believed, i needs to be trusted on φ in *every* maximally plausible world. If there are several such worlds with different assessments of expertise – e.g. if 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.

For model-based operators satifying **Soundness**, this phenomenon can be formalised: a report of φ from source i is expanded by the *join*[19] of the partitions Π_i^W, for $W \in \mathcal{Y}_\sigma$. As the above example shows, this join may be strictly coarser than any of the individual partitions Π_i^W. Formally, 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. We first need a preliminary result.

Lemma 5.5.3. *For any model-based operator,* $\mathsf{E}_i\varphi \in B_c^\sigma$ *iff* $\Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \|\varphi\|$.

Proof. "if": Suppose $\Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \|\varphi\|$. Take $W \in \mathcal{Y}_\sigma$. Then since Π_i^W refines $\Pi_i^{\mathcal{Y}_\sigma}$, we have $\Pi_i^W[\varphi] \subseteq \Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \|\varphi\|$. Since $\|\varphi\| \subseteq \Pi_i^W[\varphi]$ always holds, we have $\Pi_i^W[\varphi] = \|\varphi\|$ and thus $W, c \models \mathsf{E}_i\varphi$. Since $W \in \mathcal{Y}_\sigma$ was arbitrary, this shows $\mathsf{E}_i\varphi \in B_c^\sigma$.

"only if": Suppose $\mathsf{E}_i\varphi \in B_c^\sigma$. We need to show $\Pi_i^{\mathcal{Y}_\sigma}[\varphi] \subseteq \|\varphi\|$. Take $v \in \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. Let R_i^W be the equivalence relation corresponding to the partition Π_i^W. Then the relation R corresponding to the join $\Pi_i^{\mathcal{Y}_\sigma}$ is the smallest equivalence relation containing each of the R_i^W, which is given explictly by the transitive closure $R = \left(\bigcup_{W \in \mathcal{Y}_\sigma} R_i^W\right)^+$.

Now, from $v \in \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$ there is some $u \in \|\varphi\|$ such that vRu. By definition of the transitive closure, there are $x_0, \ldots, x_n \in \mathcal{V}$ such that $v = x_0$, $u = x_n$, and for each $0 \leq k < n$, $(x_k, x_{k+1}) \in \bigcup_{W \in \mathcal{Y}_\sigma} R_i^W$. That is, there are $W_0, \ldots, W_{n-1} \in \mathcal{Y}_\sigma$ such that $(x_k, x_{k+1}) \in R_i^{W_k}$. We wil show that each x_k lies in $\|\varphi\|$ by backwards

[19] The join of a set of partitions is its least upper bound with respect to the refinement order.

induction. For $k = n$, we have $x_n = u \in \|\varphi\|$ by assumption. If $x_{k+1} \in \|\varphi\|$, then $(x_k, x_{k+1}) \in R_i^{W_k}$ gives $x_k \in \Pi_i^{W_k}[x_{k+1}] \subseteq \Pi_i^{W_k}[\varphi]$. By assumption, $\mathsf{E}_i\varphi \in B_c^\sigma$. Since $W_k \in \mathcal{Y}_\sigma$, this means $W_k, c \models \mathsf{E}_i\varphi$ and $\Pi_i^{W_k}[\varphi] = \|\varphi\|$. Hence $x_k \in \|\varphi\|$ as desired. This shows $v = x_0 \in \|\varphi\|$, and we are done. □

Proposition 5.5.2. *If a model-based operator is EC-selective and satisfies* ***Soundness****, then*

$$\|[B_c^\sigma]\| = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$$

for all $*$-consistent σ and $c \in \mathcal{C}$.

Proof. Let σ be an $*$-consistent sequence and take $c \in \mathcal{C}$.

$\subseteq$: Let $v \in \|[B_c^\sigma]\|$. By Lemma 5.4.1, there is $W \in \mathcal{Y}_\sigma$ such that $v = v_c^W$. Take $\langle i, \varphi\rangle \in \sigma \restriction c$. By **Soundness** (and **Containment**, which holds for all model-based operators) we have $\mathsf{S}_i\varphi \in B_c^\sigma$, so $W, c \models \mathsf{S}_i\varphi$. Consequently $v = v_c^W \in \Pi_i^W[\varphi] \subseteq \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$, where we use the fact that Π_i^W refines $\Pi_i^{\mathcal{Y}_\sigma}$ in the last step.

$\supseteq$: Let $v \in \bigcap_{\langle i,\varphi\rangle \in \sigma\restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. By EC-selectivity there is some expertise-compatible selection scheme f. Write $F_\sigma(i, c, \varphi) = \|f_\sigma(i, c, \varphi)\|$. Then we have

$$\|[B_c^\sigma]\| = \bigcap_{\langle i,\varphi\rangle \in \sigma \restriction c} F_\sigma(i, c, \varphi) \tag{5.5}$$

and $\mathsf{E}_i f_\sigma(i, c, \varphi) \in B_c^\sigma$ for each $\langle i, \varphi\rangle \in \sigma \restriction c$. By Lemma 5.5.3, $\Pi_i^{\mathcal{Y}_\sigma}[F_\sigma(i, c, \varphi)] = F_\sigma(i, c, \varphi)$. We show $v \in \|[B_c^\sigma]\|$ using (5.5). Take $\langle i, \varphi\rangle \in \sigma \restriction c$. By definition of a selection scheme we have $f_\sigma(i, c, \varphi) \in \mathrm{Cn}_0(\varphi)$, so $\|\varphi\| \subseteq F_\sigma(i, c, \varphi)$. Since $v \in \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$ by assumption, we get

$$v \in \Pi_i^{\mathcal{Y}_\sigma}[\varphi] \subseteq \Pi_i^{\mathcal{Y}_\sigma}[F_\sigma(i, c, \varphi)] = F_\sigma(i, c, \varphi)$$

as required. □

Note that Proposition 5.5.2 immediately implies selectivity with respect to any scheme f such that $\|f_\sigma(i, c, \varphi)\| = \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. Since the right-hand side does not depend on the case c, we get the following corollary.

Corollary 5.5.1. *If a model-based operator is EC-selective and satisfies* ***Soundness****, then it is case-independent-selective.*

Proposition 5.5.2 also shows that propositional beliefs in case c are determined only by the reports in $\sigma \restriction c$ together with the expertise part of B^σ, via the partitions Π_i^W for $W \in \mathcal{Y}_\sigma$. This property can be expressed syntactically as follows, where for a collection G we write $\mathsf{E}(G)$ for the sub-collection of formulas of the form $\mathsf{E}_i\varphi$.

Determination. For any $*$-consistent σ and $c \in \mathcal{C}$, $[B_c^\sigma] = [\mathrm{Cn}_c(K^\sigma \sqcup \mathsf{E}(B^\sigma))]$.

In other words, **Determination** says that propositional beliefs may be fully recovered by taking (the c-consequences of) the knowledge set K^σ together with just the expertise formulas in B^σ. Surprisingly, **Determination** in fact characterises EC-selectivity, under additional mild assumptions. In what follows, recall that G_{snd}^σ denotes the collection with $(G_{\mathsf{snd}}^\sigma)_c = \{\mathsf{S}_i\varphi \mid \langle i, \varphi\rangle \in \sigma \restriction c\}$.

Theorem 5.5.3. *A model-based operator satisfying* ***Consistency*** *and* $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$ *for all* σ *is EC-selective if and only if it satisfies* ***Determination****.*

Proof. Take any model-based operator satisfying **Consistency** and which has $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$ for all σ. Note that the latter property implies **Soundness**.

"if": Suppose **Determination** holds. We claim that for any $*$-consistent σ and $c \in \mathcal{C}$,

$$\|[B^\sigma_c]\| = \bigcap_{\langle i,\varphi\rangle \in \sigma\restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]. \tag{5.6}$$

This implies selectivity upon letting $f_\sigma(i, c, \varphi)$ be any formula with models $\Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. Furthermore it implies EC-selectivity by Lemma 5.5.3.

The left-to-right inclusion of (5.6) follows by an argument identical to that of Proposition 5.5.2 using **Soundness**. It suffices to show the right-to-left inclusion. Take $v \in \bigcap_{\langle i,\varphi\rangle\in\sigma\restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. By **Consistency**, there is some $W_0 \in \mathcal{Y}_\sigma$. Consider W obtained from W_0 by setting its c-valuation to v, and by setting the partition of source i to $\Pi_i^{\mathcal{Y}_\sigma}$:

$$v^W_d = \begin{cases} v^{W_0}, & d \neq c \\ v, & d = c \end{cases},$$

$$\Pi^W_i = \Pi_i^{\mathcal{Y}_\sigma}.$$

Note that since $W_0 \in \mathcal{Y}_\sigma$, $\Pi^{W_0}_i$ refines Π^W_i for each i. We aim to show $W \in \mathrm{mod}(K^\sigma \sqcup \mathsf{E}(B^\sigma))$. Recall that, by assumption, $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$. It therefore suffices to show that $W \in \mathrm{mod}(G^\sigma_{\mathsf{snd}}) \cap \mathrm{mod}(\mathsf{E}(B^\sigma))$. First take $\langle i, d, \varphi\rangle \in \sigma$. By **Soundness** and **Containment** we have $W_0, d \models \mathsf{S}_i\varphi$, i.e. $v^{W_0}_d \in \Pi^{W_0}_i[\varphi]$. If $d \neq c$ then

$$v^W_d = v^{W_0}_d \in \Pi^{W_0}_i[\varphi] \subseteq \Pi^W_i[\varphi],$$

where we use the fact that $\Pi^{W_0}_i$ refines Π^W_i in the last step. Thus $W, d \models \mathsf{S}_i\varphi$ as required. If instead $d = c$, then by our assumption on v,

$$v^W_c = v \in \Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \Pi^W_i[\varphi]$$

so that $W, c \models \mathsf{S}_i\varphi$ as required. This shows $W \in \mathrm{mod}(G^\sigma_{\mathsf{snd}})$. For $W \in \mathrm{mod}(\mathsf{E}(B^\sigma))$, take any $\mathsf{E}_i\varphi \in B^\sigma_c$ (note that by **Closure** $\mathsf{E}(B^\sigma)$ contains the same formulas in each case, so we may choose c without loss of generality). Then by Lemma 5.5.3, $\Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \|\varphi\|$. By construction of W we evidently have $W, c \models \mathsf{E}_i\varphi$.

This shows $W \in \mathrm{mod}(K^\sigma \sqcup \mathsf{E}(B^\sigma))$. Finally, to show $v \in \|[B^\sigma_c]\|$, take any $\psi \in [B^\sigma_c]$. By **Determination**, $\psi \in \mathrm{Cn}_c(K^\sigma \sqcup \mathsf{E}(B^\sigma))$. Thus $W, c \models \psi$. But by construction the c-valuation in W is v, so $v \in \|\psi\|$ and we are done.

"only if": Suppose the operator is EC-selective according to some scheme f. To show **Determination**, take any $*$-consistent σ and $c \in \mathcal{C}$. By **Containment** we have $K^\sigma \sqsubseteq B^\sigma$, and clearly $\mathsf{E}(B^\sigma) \sqsubseteq B^\sigma$. Consequently $\mathsf{K}^\sigma \sqcup \mathsf{E}(B^\sigma) \sqsubseteq B^\sigma$; by monotonicity of Cn and **Closure** we get $\mathrm{Cn}(K^\sigma \sqcup \mathsf{E}(B^\sigma)) \sqsubseteq B^\sigma$. This in turn implies $[B^\sigma_c] \supseteq [\mathrm{Cn}_c(K^\sigma \sqcup \mathsf{E}(B^\sigma))]$.

For the reverse inclusion, it is sufficient by **Closure** to show

$$\|[\mathrm{Cn}_c(K^\sigma \sqcup \mathsf{E}(B^\sigma))]\| \subseteq \|[B^\sigma_c]\|. \tag{5.7}$$

So, take v in the set on the left-hand side. By an argument identical to the proof of Lemma 5.4.1, there is some $W \in \mathrm{mod}(K^\sigma \sqcup \mathsf{E}(B^\sigma))$ such that $v = v_c^W$. Since **Soundness** holds by the assumption that $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$, EC-selectivity and Proposition 5.5.2 give $\|[B_c^\sigma]\| = \bigcap_{\langle i,\varphi\rangle \in \sigma\restriction c} \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$.

Take $\langle i, \varphi\rangle \in \sigma \restriction c$. Let ψ be any propositional formula with $\|\psi\| = \Pi_i^{\mathcal{Y}_\sigma}[\varphi]$. Then $\mathsf{E}_i\psi \in B_c^\sigma$, so $W \in \mathrm{mod}(\mathsf{E}(B^\sigma))$ gives $W, c \models \mathsf{E}_i\psi$. Now, **Soundness** and $W \in \mathrm{mod}(K^\sigma)$ also gives $W, c \models \mathsf{S}_i\varphi$, i.e. $v = v_c^W \in \Pi_i^W[\varphi]$. Since $\|\varphi\| \subseteq \Pi_i^{\mathcal{Y}_\sigma}[\varphi] = \|\psi\|$, we get

$$v \in \Pi_i^W[\varphi] \subseteq \Pi_i^W[\psi] = \|\psi\| = \Pi_i^{\mathcal{Y}_\sigma}[\varphi].$$

This shows (5.7) and completes the proof. □

Note that if the basic postulates are given, the condition $K^\sigma = \mathrm{Cn}(G^\sigma_{\mathsf{snd}})$ in Theorem 5.5.3 is equivalent to $K^\emptyset = \mathrm{Cn}(\emptyset)$ by Proposition 5.2.1. In particular, Theorem 5.5.3 applies to our concrete operators var-based-cond, part-based-cond and excess-min. Since we have already seen these operators are *not* EC-selective, we also have that they each fail **Determination**.

The potential problem with EC-selectivity, as expressed by **Determination**, is that is only permits belief formation on the basis of soundness statements together with firmly believed expertise statements in $\mathsf{E}(B^\sigma)$. A natural weaker notion of expertise-compatible selectivity requires not that i is believed to have expertise on $f_\sigma(i, c, \varphi)$, but merely that such expertise is *consistent* with B^σ.

Definition 5.5.4. *A selection scheme f is* weakly expertise-compatible *with an operator $\sigma \mapsto \langle B^\sigma, K^\sigma\rangle$ if for all $*$-consistent σ and $\langle i, c, \varphi\rangle \in \sigma$,*

$$\neg\mathsf{E}_i f_\sigma(i, c, \varphi) \notin B_c^\sigma.$$

Mirroring earlier terminology, say an operator is *weakly EC-selective* if it is selective according to some weakly expertise-compatible scheme. Weak EC-selectivitity overcomes the issues of EC-selectivity highlighted above on the sequence

$$\sigma = (\langle i, c, p\rangle, \langle j, c, p\rangle, \langle i, d, p\rangle, \langle j, d, \neg p\rangle).$$

For example, each of our example operators var-based-cond, part-based-cond and excess-min are weakly EC-selective for this particular σ according to the selection

$$f_\sigma(i, c, p) = f_\sigma(j, c, p) = p$$
$$f_\sigma(i, d, p) = f_\sigma(j, d, \neg p) = \top.$$

However, this selection is *not* case independent. Questions around the interaction between weak EC-selectivity and case independence, as well as the whether the example operators are weakly EC-selective and/or case-independent-selective in general, are left for future work.

5.6 Related Work

In this section we discuss related work.

Belief Merging. In the framework of Konieczny and Pérez [61], 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 [60] 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 [61]. 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 $\Delta_\mu(\Phi) \not\vdash \psi$ 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 5.2.1. It also facilitates postulates which refer directly to expertise, such as the weakenings of **Success** in Section 5.4.

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 [47], 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 [90]. 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 [79] – 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 $\mathsf{E}\varphi$ holds iff $\mathsf{A}(\varphi \to \mathsf{K}\varphi)$ 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, $\mathsf{S}\varphi$ translates to $\neg\mathsf{K}\neg\varphi$. That is, φ is sound exactly when the source does not *know* φ is false.

In the present framework, if we set $W, c \models \mathsf{K}_i\varphi$ iff $\Pi_i[v_c] \subseteq \|\varphi\|$ and $W, c \models \mathsf{A}\Phi$ 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 $\mathsf{E}_i\varphi \equiv \mathsf{A}(\varphi \to \mathsf{K}_i\varphi)$ and $\mathsf{S}_i\varphi \equiv \neg\mathsf{K}_i(\neg\varphi)$. 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.

5.7 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. [39, 51, 12, 23]). 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**.

Limitations and 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 [99]. Moreover, there are many weaker version of **Success** which have been considered in the literature (e.g. in [39, 51, 12]); 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* [53, 99, 23].

6 Truth-Tracking

In this chapter we study truth-tracking in the logical framework of Singleton and Booth [83] for reasoning about multiple non-expert information sources. Broadly speaking, the goal of truth-tracking is to find the true state of the world given some input which describes it. In our case this involves finding the true state of some propositional domain about which the sources give reports, and finding the extent of the expertise of the sources themselves.

The general problem of truth-tracking has been studied in various forms across many domains. Perhaps the oldest approach goes back to Condorcet [20], whose celebrated *Jury Theorem* states that a majority vote on a yes/no issue will yield the "correct" answer with probability approaching 1 as the number of voters tends to infinity, provided that each voter is more reliable than random choice. This result has since been generalised in many directions Grofman, Owen, and Feld [48]. More widely, *epistemic social choice* [33] studies aggregation methods (e.g. voting rules) from the point of finding the "correct" result with high probability, where individual votes are seen as noisy approximations. Of particular relevance to our work is truth-tracking in *judgement aggregation* in social choice [52, 89], which also takes place in a logical framework. *Belief merging* has close links with judgement aggregation, and generalised jury theorems have been found here too [37].

In crowdsourcing, the problem of *truth discovery* [63] looks at how information from unreliable sources can be aggregated to find the true value of a number of variables, and to find the true reliability level of the sources. This is close to our setting, since incoming information is not always assumed to be reliable, and information about the sources themselves is sought after. Work in this area combines empirical results (e.g. how well methods find the truth on test datasets for which true values are known) and theoretical guarantees, and is typically set in a probabilistic framework.

On the other hand, *formal learning theory* [54] offers a non-probabilistic view on truth-tracking, stemming from the framework of Gold [43] for identification in the limit. In this paradigm a learner receives an infinite sequence of information step-by-step, such that all true information eventually appears in the sequence. The learner outputs a hypothesis at each step, and aims to stabilise on the correct hypothesis after some finite number of steps. This framework has been combined with belief revision theory [57, 6] and dynamic epistemic logic [8].

This is the approach we take, and in particular we adapt the truth-tracking setting of Baltag, Gierasimczuk, and Smets [6]. We apply this to the logical framework of Singleton and Booth [83]. Briefly, this framework extends finite propositional

logic with two new notions: that of a source having *expertise* on a formula, and a formula being *sound* for a source to report. Intuitively, expertise on φ means the source has the epistemic capability to distinguish between any pair of φ and $\neg\varphi$ states: they know whether or not φ holds in any state. A formula is sound for a source if it is true *up to their lack of expertise*. For example, if a source has expertise on φ but not ψ, then $\varphi \wedge \psi$ is sound whenever φ holds, since we can ignore the ψ part (on which the source has no expertise). The resulting logical language therefore addresses both the *ontic* facts of the world, through the propositional part, and the *epistemic* state of the sources, via expertise and soundness.

Generally speaking, formal learning theory supposes that all information received is true, and that all true information is eventually received.[1] This is not a tenable assumption with non-expert sources: some sources may simply lack the expertise to know whether φ is true or false. Instead we make a different (and strong) assumption: all and only *sound* reports are received. Thus, sources report everything consistent with their expertise. This necessitates inconsistent reports from non-experts, even if we assume sources are rational. Consequently, the input to our learning methods should be distinguished from the inputs to belief revision and belief merging methods [1, 61] – also propositional formulas – which represent *beliefs* of the reporting sources. Indeed, we do not model beliefs of the sources at all.

The following example informally illustrates the core concepts of the logical framework and truth-tracking, and will be returned to throughout the chapter.

Example 6.0.1. *Consider a medical scenario in which patient A is checked for conditions p and q. By examining A, a doctor* D *has expertise to determine whether A has at least one of p or q, but cannot tell which one(s) without a blood test. A test is only available for p, however, so that the technician* T *performing the test has expertise on p but not q.*

Supposing A in fact suffers from q but not p, D *considers each of* $p \wedge q$, $\neg p \wedge q$ *and* $p \wedge \neg q$ *possible, whereas* T *considers both* $\neg p \wedge q$ *and* $\neg p \wedge \neg q$ *possible. Assuming both sources report* all *they consider possible, their combined expertise leaves* $\neg p \wedge q$ *as the only possibility. Intuitively, this means we* can *find the true values of p and q in this case.*

Now consider a patient B who suffers from both conditions. D *cannot distinguish A and B, so will provide the same reports, and* T *considers both* $p \wedge q$ *and* $p \wedge \neg q$ *possible. In this case* T *is more knowledgable than* D *– since they consider fewer situations possible – but we cannot narrow down the true value of q. Thus truth-tracking is only possible for p. The second patient still provides useful information, though, since together with the reports on A,* T*'s lack of expertise tells us all the (in)distinctions between states they are able to make. Namely,* T *cannot distinguish between* $p \wedge q$ *and* $p \wedge \neg q$*. Thus we can find the truth about* T*'s expertise.*

Contributions. [**TODO:** Explicitly list contributions]

Chapter outline. In Section 6.1 we outline the logical framework for reasoning about expertise. Section 6.2 introduces the key concepts of truth-tracking and solv-

[1] Baltag, Gierasimczuk, and Smets [6] consider erroneous reports, but only provided that all errors are eventually corrected.

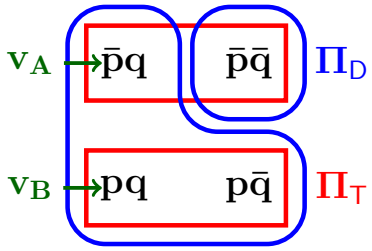


Figure 6.1: Example of a world W, which formalises Example 6.0.1. Here $\mathsf{Prop} = \{p, q\}$, $\mathcal{S} = \{\mathsf{D}, \mathsf{T}\}$ *and* $\mathcal{C} = \{A, B\}$.

able questions. We characterise solvable questions in Section 6.3, and explore what they can reveal about the actual world in Section 6.4. Section 6.5 looks at learning methods themselves, and characterises truth-tracking methods. We conclude in Section 6.6.

6.1 Preliminaries

In this section we recall the logical framework of Singleton and Booth [83] for reasoning with non-expert sources.

Syntax. Let Prop be a finite set of propositional variables, and let $\mathcal{L}_0$ denote the propositional language generated from Prop. We use $\mathcal{L}_0$ to model the domain underlying the truth-tracking problem; it describes the "ontic" facts of the world, irrespective of the expertise of the sources. Formulas in $\mathcal{L}_0$ will be denoted by lower-case Greek letters (φ, ψ, etc).

Let $\mathcal{S}$ be a finite set of sources. The language $\mathcal{L}$ extends $\mathcal{L}_0$ with expertise and soundness formulas for each source $i \in \mathcal{S}$, and is defined by the following grammar:

$$\Phi ::= \varphi \mid \mathsf{E}_i\varphi \mid \mathsf{S}_i\varphi \mid \Phi \wedge \Phi \mid \neg\Phi,$$

for $\varphi \in \mathcal{L}_0$ and $i \in \mathcal{S}$. Formulas in $\mathcal{L}$ will be denoted by upper-case Greek letters (Φ, Ψ etc). Other logical connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) are introduced as abbreviations. We read $\mathsf{E}_i\varphi$ as "i has expertise on φ", and $\mathsf{S}_i\varphi$ as "φ is sound for i". Note that we restrict the expertise and soundness formulas to propositional arguments, and do not considered iterated formulas such as $\mathsf{E}_i\mathsf{S}_j\varphi$.

Semantics. Let $\mathcal{V}$ denote the set of propositional valuations over Prop. We represent the expertise of a source i with a *partition* Π_i of $\mathcal{V}$. Intuitively, this partition represents the distinctions between states the source is able to make: valuations in the same cell in Π_i are indistinguishable to i, whereas i is able to tell apart valuations in different cells. We say i has expertise on φ iff i can distinguish all φ states from $\neg\varphi$ states, and φ is sound for i if the "actual" state is indistinguishable from some φ state. Note that these notions are closely linked to *S5 knowledge* from epistemic logic; see [83, 79] for further discussion.

Let $\mathcal{C}$ be a finite set of *cases*, thought of as independent instantiations of the domain of interest. For example, the cases in Example 6.0.1 are the patients A and B. We consider the expertise of sources to be fixed across all cases.

A *world* is a pair $W = \langle \{v_c\}_{c\in\mathcal{C}}, \{\Pi_i\}_{i\in\mathcal{S}} \rangle$, where

- $v_c \in \mathcal{V}$ is the "actual" valuation for case c;
- $\Pi_i \subseteq 2^{\mathcal{V}}$ is a partition representing the expertise of i.

Let $\mathcal{W}$ denote the set of worlds. Note that $\mathcal{W}$ is finite, since $\mathcal{V}$, $\mathcal{C}$ and $\mathcal{S}$ are. For $\varphi \in \mathcal{L}_0$, write $\|\varphi\| \subseteq \mathcal{V}$ for the models of φ, and write $v \Vdash \varphi$ iff $v \in \|\varphi\|$. The consequences of a set $\Gamma \subseteq \mathcal{L}_0$ is denoted by $\mathrm{Cn}_0(\varphi)$, and we write $\Gamma \Vdash \varphi$ if $\varphi \in \mathrm{Cn}_0(\Gamma)$. For a partition Π, let $\Pi[v]$ denote the unique cell in Π containing v, and write $\Pi[U] = \bigcup_{v \in U} \Pi[v]$ for $U \subseteq \mathcal{V}$. For brevity, we write $\Pi[\varphi]$ instead of $\Pi[\|\varphi\|]$. We evaluate $\mathcal{L}$ formulas with respect to a world W and a case c as follows:

$$\begin{aligned} W, c \models \varphi &\iff v_c \Vdash \varphi \\ W, c \models \mathsf{E}_i\varphi &\iff \Pi_i[\varphi] = \|\varphi\| \\ W, c \models \mathsf{S}_i\varphi &\iff v_c \in \Pi_i[\varphi], \end{aligned}$$

where the clauses for conjunction and negation are as standard. The semantics follows the intuition outlined above: $\mathsf{E}_i\varphi$ holds when Π_i separates the φ states from the $\neg\varphi$ states, and $\mathsf{S}_i\varphi$ holds when v_c is indistinguishable from some φ state. Consequently, $\mathsf{S}_i\varphi$ means φ is true *up to the expertise of* i: if we weaken φ according to i's expertise, the resulting formula (with models $\Pi_i[\varphi]$) is true.

Example 6.1.1. *Take W from Fig. 6.1, which formalises Example 6.0.1. Then $W, c \models \mathsf{E}_{\mathsf{D}}(p \vee q)$ for all $c \in \mathcal{C}$, since $\|p \vee q\|$ is a cell in Π_{D}. We also have $W, A \models \neg p \wedge \mathsf{S}_{\mathsf{D}}p$, i.e. patient A does not suffer from condition p, but it is consistent with D's expertise that they do.*

We write $W, c \models \Gamma$, for a set of formulas $\Gamma \subseteq \mathcal{L}$, if $W, c \models \Phi$ for all $\Phi \in \Gamma$. For a set $S \subseteq \mathcal{W}$, we write $S, c \models \Phi$ iff $W, c \models \Phi$ for all $W \in S$.

Reports. A *report* is triple $\langle i, c, \varphi \rangle$, where $i \in \mathcal{S}$, $c \in \mathcal{C}$ and $\varphi \in \mathcal{L}_0$ with $\varphi \not\equiv \bot$. In this chapter, we interpret such triples as source i reporting that φ is possible in case c. An *input sequence* σ is a finite sequence of reports.

A *method* L maps each input sequence σ to a set of worlds $L(\sigma) \subseteq \mathcal{W}$, representing the worlds L considers most plausible given σ.[2] We say L *believes* $S \subseteq \mathcal{W}$ on the basis of σ if $L(\sigma) \subseteq S$. L is *consistent* if $L(\sigma) \neq \emptyset$ for all input sequences σ.

6.2 Truth-Tracking

We adapt the framework for truth-tracking from [5, 6], which finds its roots in formal learning theory. In this framework, a learning method receives increasing initial segments of an infinite sequence – called a *stream* – which enumerates all (and only) the true propositions observable at the "actual" world. Truth-tracking requires the method to eventually find the actual world (or some property thereof), given *any* stream.

As mentioned in the introduction, in our setting we cannot assume the sources themselves report only true propositions. Instead, our streams will enumerate all

[2] We depart from the original framework here by taking a semantic view of belief change operators, with the output a set of worlds instead of formulas.

the *sound* reports. Thus, a stream may include false reports, but such false reports only arise due to lack of expertise of the corresponding source.[3] Moreover, *all* sound reports will eventually arise. Since $\mathsf{S}_i\varphi$ means φ is possible from the point of view of i's expertise, we can view a stream as each source sharing *all that they consider possible* for each case $c \in \mathcal{C}$. In particular, a non-expert source may report both φ and $\neg\varphi$ for the same case.

Definition 6.2.1. *An infinite sequence of reports ρ is a* stream *for W iff for all i, c, φ:*

$$\langle i, c, \varphi\rangle \in \rho \iff W, c \models \mathsf{S}_i\varphi.$$

We refer to the left-to-right implication as *soundness* of ρ for W, and the right-to-left direction as *completeness*. Note that every world W has some stream: the set $\{\langle i, c, \varphi\rangle \mid W, c \models \mathsf{S}_i\varphi\}$ is countable, so can be indexed by $\mathbb{N}$ to form a stream. For $n \in \mathbb{N}$ we let ρ_n denote the n-th report in ρ, and write $\rho[n]$ for the finite initial segment of ρ of length n.

Example 6.2.1. *Consider W from Fig. 6.1 and case A. From the point of view of* D*'s expertise, the "actual" world could be pq, $\bar{p}q$, $p\bar{q}$. Consequently, in a stream for W,* D *will report p, $\neg p$, q, $\neg q$, $p \vee q$, and so on. A report that* D *will* not *give is $\neg(p \vee q)$, since* D *has expertise to know this is false.*

Note that v_A and v_B are indistinguishable to D*, so the reports of* D *in any stream will be the same for both cases. In contrast,* T can *distinguish the two cases, and will report $\neg p$ in case A but not in B, and p in case B but not in A.*

A *question* Q is a partition of $\mathcal{W}$. That is, a question is a set of disjoint *answers* $A \in Q$, with each world W appearing in a unique cell $Q[W]$ – the correct answer at W.

Example 6.2.2.

1. *Any formula $\Phi \in \mathcal{L}$ and case c defines a question $Q_{\Phi,c}$, whose two cells consist of the worlds satisfying Φ, respectively $\neg\Phi$, in case c. Intuitively, this question asks whether Φ is true or false in case c.*

2. *The finest question $Q_\perp = \{\{W\} \mid W \in \mathcal{W}\}$ asks: what is the "actual" world?*

3. *More generally, for any set X and function $f : \mathcal{W} \to X$, the equivalence relation given by $W \simeq_f W'$ iff $f(W) = f(W')$ defines a question Q_f.*

   In this way any data associated with a world gives rise to a question. For example, if $f(W) = \{i \in \mathcal{S} \mid \Pi_i^W[p] = \|p\|\}$ we ask for the set of sources with expertise on p; if $f(W) = |\{c \in \mathcal{C} \mid W, c \models p\}|$ we ask for the number of cases where p holds, etc.

   In fact, all questions are of this form: given Q we may define $f : \mathcal{W} \to Q$ by $f(W) = Q[W]$; then $Q_f = Q$.

A method solves Q if it eventually believes the correct answer when given any stream.

[3] Alternatively, we can consider statements of the form "φ is sound for i in case c" as a higher-order "proposition"; a stream then enumerates all true propositions of this kind.

Definition 6.2.2. *A method L* solves *a question Q if for all worlds W and all streams ρ for W, there is $n \in \mathbb{N}$ such that $L(\rho[m]) \subseteq Q[W]$ for all $m \geq n$. A question Q is* solvable *if there is some consistent method L which solves Q.*

Note that we do not require $W \in L(\rho[m])$. Since we work in a finite framework, solvability can be also expressed in terms of eliminating incorrect worlds.

Proposition 6.2.1. *A method L solves Q if and only if for all W, all streams ρ for W, and all $W' \notin Q[W]$, there is $n_{W'} \in \mathbb{N}$ such that $W' \notin L(\rho[m])$ for all $m \geq n_{W'}$.*

Proof. "if": Taking $n = \max\{n_{W'} \mid W' \notin Q[W]\}$, which exists since $\mathcal{W}$ is finite, $L(\rho[m]) \subseteq Q[W]$ for $m \geq n$.

"only if": Taking n from the definition of L solving Q, we may simply take $n_{W'} = n$ for all $W' \notin Q[W]$. □

6.3 Characterising Solvable Questions

In this section we explore solvability of questions, finding that there is a unique "hardest" question which subsumes all solvable questions. We show this is itself solvable, and thus obtain a precise characterisation of solvability.

Questions are partially ordered by partition refinement: $Q \preceq Q'$ iff each $A' \in Q'$ can be written as a union of answers from Q. Equivalently, $Q[W] \subseteq Q'[W]$ for all W. This can be interpreted as a *difficulty ordering*: if $Q \preceq Q'$ then each answer of Q' is just a disjunction of answers of Q, and thus Q' is *easier* than Q. Naturally, if Q is solvable then so too is any easier question.

Proposition 6.3.1. *If Q is solvable and $Q \preceq Q'$, then Q' is solvable.*

Proof. The method which solves Q also solves Q'. □

Since question solving is based on streams of sound reports, worlds satisfying the same soundness statements cannot be distinguished by any solvable question. To formalise this, define a preorder $\sqsubseteq$ on $\mathcal{W}$ by

$$W \sqsubseteq W' \iff \forall i, c, \varphi : \; W, c \models \mathsf{S}_i\varphi \implies W', c \models \mathsf{S}_i\varphi.$$

Thus, $W \sqsubseteq W'$ iff any report sound for W is also sound for W'. We denote by $\sqsubset$ and $\approx$ the strict and symmetric parts of $\sqsubseteq$, respectively.[4]

Lemma 6.3.1. *$W \sqsubseteq W'$ if and only if for all $i \in \mathcal{S}$ and $c \in \mathcal{C}$, $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$.*

Proof. "if": Suppose $W, c \models \mathsf{S}_i\varphi$. Then $v_c^W \in \Pi_i^W[\varphi]$, so there is $u \in \|\varphi\|$ such that $v_c^W \in \Pi_i^W[u]$. Consequently $u \in \Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$, which means $v_c^{W'} \in \Pi_i^{W'}[u] \subseteq \Pi_i^{W'}[\varphi]$. Hence $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$.

"only if": Let $u \in \Pi_i^W[v_c^W]$. Let φ be any formula with $\|\varphi\| = \{u\}$. Then $W, c \models \mathsf{S}_i\varphi$, so $W \sqsubseteq W'$ gives $W', c \models \mathsf{S}_i\varphi$, i.e. $v_c^{W'} \in \Pi_i^{W'}[u]$, so $u \in \Pi_i^{W'}[v_c^{W'}]$. Hence $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. □

[4] Baltag, Gierasimczuk, and Smets [5] explore *topological* interpretations of solvability by considering the topology on the set of worlds generated by observable propositions. In our setting, this is the topology generated by sets of the form $\{W \mid W, c \models \mathsf{S}_i\varphi\}$. In this topology, $\sqsubseteq$ is the *specialisation preorder*.

Note that $\Pi_i[v_c]$ is the set of valuations indistinguishable from the "actual" valuation in case c, for source i. In light of Lemma 6.3.1, we can interpret $W \sqsubseteq W'$ as saying that all sources are *more knowledgeable* in each case c in world W than in W'. However, $W \sqsubseteq W'$ does not say anything about the partition cells not containing some v_c.

Proposition 6.3.2. *The following are equivalent.*

1. *W and W' have exactly the same streams.*
2. *$W \approx W'$.*
3. *For all $i \in \mathcal{S}$ and $c \in \mathcal{C}$, $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$.*

Proof. (2) and (3) are easily seen to be equivalent in light of Lemma 6.3.1. To show (1) is equivalent to (2), first suppose W and W' have the same streams, and suppose $W, c \models \mathsf{S}_i\varphi$. Taking an arbitrary stream ρ for W, completeness gives $\langle i, c, \varphi\rangle \in \rho$. But ρ is a stream for W' too, and soundness gives $W', c \models \mathsf{S}_i\varphi$. Hence $W \sqsubseteq W'$. A symmetrical argument shows $W' \sqsubseteq W$.

On the other hand, if $W \approx W'$ then W and W' satisfy exactly the same soundness statements, so it is clear that any sequence ρ is a stream for W iff it is a stream for W'. □

Since it will play a special role throughout, we denote by Q^* the question formed by the equivalence relation $\approx$. Then $Q^*[W]$ is the set of W' with $W \approx W'$. Since no solvable question can distinguish $\approx$-equivalent worlds, we have the following.

Lemma 6.3.2. *If Q is solvable then $Q^* \preceq Q$.*

Proof. Suppose L is a consistent method solving Q. We show $Q^*[W] \subseteq Q[W]$ for all W. Indeed, let $W' \in Q^*[W]$. Then $W' \approx W$. Taking any stream ρ for W, there is n such that $L(\rho[m]) \subseteq Q[W]$ for $m \geq n$. On the other hand ρ is also a stream for W' by Proposition 6.3.2, so there is n' such that $L(\rho[m]) \subseteq Q[W']$ for $m \geq n'$. Setting $m = \max\{n, n'\}$ and using the fact that L is consistent, we find $\emptyset \subset L(\rho[m]) \subseteq Q[W] \cap Q[W']$. Since Q is a partition, this means $Q[W] = Q[W']$, i.e. $W' \in Q[W]$. □

So, any solvable question is coarser than Q^*. Fortunately, Q^* itself is solvable since we work in a finite framework. For a sequence σ, write $\mathcal{X}_\sigma^{\mathsf{snd}}$ for the set of worlds W such that $W, c \models \mathsf{S}_i\varphi$ for all $\langle i, c, \varphi\rangle \in \sigma$. To solve Q^* it suffices to conjecture the $\sqsubseteq$-minimal worlds in $\mathcal{X}_\sigma^{\mathsf{snd}}$.

Proposition 6.3.3. *Q^* is solvable.*

Proof. Set $L(\sigma) = \min_\sqsubseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ if $\mathcal{X}_\sigma^{\mathsf{snd}} \neq \emptyset$, and $L(\sigma) = \mathcal{W}$ otherwise (where $W \in \min_\sqsubseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ iff $W \in \mathcal{X}_\sigma^{\mathsf{snd}}$ and there is no $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ with $W' \sqsubset W$). Note that L is consistent since $\mathcal{W}$ is finite and non-empty. We show that L solves Q^* by Proposition 6.2.1. Take any world W and a stream ρ. First note that, by soundness of ρ, $W \in \mathcal{X}_{\rho[n]}^{\mathsf{snd}}$ for all $n \in \mathbb{N}$, so we are always in the first case in the definition of L.

Take $W' \notin Q^*[W]$. Then $W \not\approx W'$. Consider two cases:

- **Case 1:** $W \not\sqsubseteq W'$. By definition, there are i, c, φ such that $W, c \models \mathsf{S}_i\varphi$ but $W', c \not\models \mathsf{S}_i\varphi$. By completeness of ρ for W, there is n such that $\rho_n = \langle i, c, \varphi\rangle$. Consequently $W' \notin \mathcal{X}^{\mathsf{snd}}_{\rho[m]}$ for all $m \geq n$. Since $L(\rho[m]) \subseteq \mathcal{X}^{\mathsf{snd}}_{\rho[m]}$, we have $W' \notin L(\rho[m])$ as required.
- **Case 2:** $W \sqsubset W'$. Since $W \in \mathcal{X}^{\mathsf{snd}}_{\rho[n]}$ for all n, W' can never be $\sqsubseteq$-minimal. Thus $W' \notin L(\rho[n])$ for all n.

Note that these cases are exhaustive since $W \not\approx W'$. This completes the proof. □

Putting Propositions 6.3.1 and 6.3.3 and Lemma 6.3.2 together we obtain a characterisation of solvable questions.

Theorem 6.3.1. *Q is solvable if and only if $Q^* \preceq Q$.*

Given this result, Q^* is the only question that really matters: any other question is either unsolvable or formed by coarsening Q^*. With this in mind, we make the following definition.

Definition 6.3.1. *A method is* truth-tracking *if it solves Q^*.*

Example 6.3.1. *We refer back to the questions of Example 6.2.2.*

1. *The question $Q_{\varphi,c}$, for any propositional formula $\varphi \in \mathcal{L}_0$, is solvable if and only if either φ is a tautology or a contradiction. To see the "only if" part, consider the contrapositive. For any contingent formula φ, take worlds W_1, W_2 where no source has any expertise (i.e. $\Pi_i^{W_k} = \{\mathcal{V}\}$) but where $v_c^{W_1} \Vdash \varphi$, $v_c^{W_2} \Vdash \neg\varphi$. Then $W_1 \approx W_2$ (e.g. by Proposition 6.3.2) but $W_1 \notin Q_{\varphi,c}[W_2]$.*

   Similarly, $Q_{\mathsf{E}_i\varphi,c}$ is solvable iff either φ is a tautology or contradiction, when $|\mathsf{Prop}| \geq 2$.

2. *The finest question $Q_\perp$ is not solvable, since there are always distinct W, W' with $W \approx W'$.*

3. *In general, Q_f is solvable iff $W \approx W'$ implies $f(W) = f(W')$, i.e. iff f takes a unique value on each equivalence class of $\approx$.*

6.4 What Information can be Learned?

Solving a question Q has a *global* character: we must find the correct answer $Q[W]$ starting from *any* world W. As we saw in Example 6.3.1, this rules out the possibility of solving many interesting questions due to the presence of "abnormal" worlds (e.g. those in which no sources have any expertise). In this section we take a more fine-grained approach by looking *locally*: given some *particular* world W, what can we learn about W via truth-tracking methods? Concretely, what properties of W are uniquely defined across $Q^*[W]$?

Clearly this depends on W. If no sources have expertise then source partitions are uniquely defined (since *all* consistent formulas are sound, and only the trivial partitions have this property), but any combination of valuations is possible. On the other hand if all sources have total expertise then valuations are uniquely defined,

but there may not be enough cases to uniquely identify the source partitions. Of particular interest is the case where $Q^*[W]$ contains only W; starting in such a world, truth-tracking methods are able to find the true world exactly.

In what follows, say S *decides* Φ *in case* c iff either $S, c \models \Phi$ or $S, c \models \neg\Phi$. That is, the truth value of Φ in case c is unambiguously defined across S. If Φ does not depend on the case (e.g. if $\Phi = \mathsf{E}_i\varphi$) we simply say S decides Φ.

6.4.1 Valuations

We start by considering when $Q^*[W]$ decides a propositional formula φ in case c, i.e. when truth-tracking methods are guaranteed to successfully determine whether or not φ holds in the "actual" world. This leads to a precise characterisation of when $Q^*[W]$ contains a *unique* valuation in case c, so that v_c^W can be found exactly.

We need a notion of *group expertise*. For $\mathcal{S}' \subseteq \mathcal{S}$ and $\Gamma \subseteq \mathcal{L}_0$, write $W \models \mathsf{E}_{\mathcal{S}'}\Gamma$ if for each $\psi \in \Gamma$ there is $i \in \mathcal{S}'$ such that $W \models \mathsf{E}_i\psi$. Then the group $\mathcal{S}'$ have expertise on Γ in a collective sense, even if no single source has expertise on *all* formulas in Γ. We have that φ is decided if $\mathcal{S}$ have group expertise on a set of true formulas $\Gamma \subseteq \mathcal{L}_0$ such that either $\Gamma \Vdash \varphi$ or $\Gamma \Vdash \neg\varphi$.

Theorem 6.4.1. *$Q^*[W]$ decides $\varphi \in \mathcal{L}_0$ in case c if and only if there is $\Gamma \subseteq \mathcal{L}_0$ such that (i) $W, c \models \Gamma$; (ii) $W \models \mathsf{E}_{\mathcal{S}}\Gamma$; and (iii) either $\Gamma \Vdash \varphi$ or $\Gamma \Vdash \neg\varphi$.*

$Q^*[W]$ decides *all* propositional formulas – and thus determines the c-valuation v_c^W exactly – iff $\mathcal{S}$ have group expertise on a maximally consistent set of true formulas. For $S \subseteq \mathcal{W}$ and $c \in \mathcal{C}$, write $\mathcal{V}_c^S = \{v_c^W \mid W \in S\}$ for the c-valuations appearing in S.

Theorem 6.4.2. *The following are equivalent.*

1. $\mathcal{V}_c^{Q^*[W]} = \{v_c^W\}$.
2. *$Q^*[W]$ decides φ in case c, for all $\varphi \in \mathcal{L}_0$.*
3. *There is $\Gamma \subseteq \mathcal{L}_0$ such that (i) $W, c \models \Gamma$; (ii) $W \models \mathsf{E}_{\mathcal{S}}\Gamma$; and (iii) $\mathrm{Cn}_0(\Gamma)$ is a maximally consistent set.*

We illustrate Theorem 6.4.2 with an example.

Example 6.4.1. *Consider W from Fig. 6.1. Then one can show $\mathcal{V}_A^{Q^*[W]} = \{\bar{p}q\} = \{v_A^W\}$, and $\mathcal{V}_B^{Q^*[W]} = \{pq, p\bar{q}\} \neq \{v_B^W\}$. That is, W's A valuation is uniquely determined by truth-tracking methods, but its B valuation is not: there is some world $W' \approx W$ whose B-valuation differs from W's. This matches the informal reasoning in Example 6.0.1, in which patient A could be successfully diagnosed on both p and q but B could not.*

Formally, take $\Gamma = \{p \vee q, \neg p\}$. Then $W, A \models \Gamma$, $W \models \mathsf{E}_{\mathcal{S}}\Gamma$ (since D has expertise on $p \vee q$ and T has expertise on $\neg p$), and $\mathrm{Cn}_0(\Gamma) = \mathrm{Cn}_0(\neg p \wedge q)$, which is maximally consistent. This example shows how the expertise of multiple sources can be combined to find valuations uniquely, but that this is not necessarily possible in all cases.

The remainder of this section proves Theorems 6.4.1 and 6.4.2.

Lemma 6.4.1. *For $W \approx W'$, $i \in \mathcal{S}$ and $\varphi \in \mathcal{L}_0$,*

$$W, c \models \varphi \wedge \mathsf{E}_i\varphi \implies W', c \models \varphi.$$

Proof. From $W, c \models \varphi$ we have $v_c^W \in \|\varphi\|$, so $\Pi_i^W[v_c^W] \subseteq \Pi_i^W[\varphi]$. But $W, c \models \mathsf{E}_i\varphi$ means $\Pi_i^W[\varphi] = \|\varphi\|$, so in fact $\Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Now using $W \approx W'$, we find $v_c^{W'} \in \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Hence $W', c \models \varphi$. □

Lemma 6.4.2. $\mathcal{V}_c^{Q^*[W]} = \bigcap_{i\in\mathcal{S}} \Pi_i^W[v_c^W]$.

Proof. "$\subseteq$": Suppose $u \in \mathcal{V}_c^{Q^*[W]}$. Then there is $W' \approx W$ such that $u = v_c^{W'}$. Let $i \in \mathcal{S}$. Then $u \in \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W]$ by Proposition 6.3.2, as required.

"$\supseteq$": Suppose $u \in \bigcap_{i\in\mathcal{S}} \Pi_i^W[v_c^W]$. Let W' be the world obtained from W by setting the c-valuation to u, keeping partitions and other valuations the same. We need to show $W' \approx W$. We do so via Proposition 6.3.2, by showing condition (3). Take any $i \in \mathcal{S}$ and $d \in \mathcal{C}$. If $d \neq c$ then $v_d^{W'} = v_d^W$; since partitions are the same in W' as in W we get $\Pi_i^W[v_d^W] = \Pi_i^{W'}[v_d^{W'}]$. For $c = d$, note $\Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[u]$. By assumption $u \in \Pi_i^W[v_c^W]$, so $\Pi_i^W[u] = \Pi_i^W[v_c^W]$. Hence $\Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W]$ as required. □

Proof of Theorem 6.4.1. "if": Take $W' \in Q^*[W]$. Note that since $W, c \models \Gamma$ and $W, c \models \mathsf{E}_{\mathcal{S}}\Gamma$, we may apply Lemma 6.4.1 to each formula in Γ in turn to find $W', c \models \Gamma$. Now, if $W, c \models \varphi$ then we must have $\Gamma \Vdash \varphi$, so $W', c \models \varphi$ too. Otherwise $W, c \not\models \varphi$, so we must have $\Gamma \Vdash \neg\varphi$ and $W', c \not\models \varphi$. This shows $W', c \models \varphi$ if and only if $W, c \models \varphi$. Since $W' \in Q^*[W]$ was arbitrary, $Q^*[W]$ decides φ in case c.

"only if": Suppose $Q^*[W]$ decides φ in case c. For each $i \in \mathcal{S}$, take some $\psi_i \in \mathcal{L}_0$ such that $\|\psi_i\| = \Pi_i^W[v_c^W]$. Then $W \models \mathsf{E}_i\psi_i$. Set $\Gamma = \{\psi_i\}_{i\in\mathcal{S}}$. Clearly $W, c \models \Gamma$ and $W \models \mathsf{E}_{\mathcal{S}}\Gamma$. Now, take any $u \in \|\Gamma\|$. By Lemma 6.4.2, $\|\Gamma\| = \bigcap_{i\in\mathcal{S}} \Pi_i^W[v_c^W] = \mathcal{V}_c^{Q^*[W]}$. Hence there is some $W' \in Q^*[W]$ such that $u = v_c^{W'}$. But $Q^*[W]$ decides φ in case c, so $W', c \models \varphi$ iff $W, c \models \varphi$. Thus $u \Vdash \varphi$ iff $W, c \models \varphi$. Since $u \in \|\Gamma\|$ was arbitrary, we have $\Gamma \Vdash \varphi$ if $W, c \models \varphi$, and $\Gamma \Vdash \neg\varphi$ otherwise. □

Proof of Theorem 6.4.2. (1) implies (2): If $W' \in Q^*[W]$ then W and W' share the same c-valuation by (1), so clearly $W, c \models \varphi$ iff $W', c \models \varphi$, for any φ. Hene $Q^*[W]$ decides φ in case c.

(2) implies (1): Clearly $v_c^W \in \mathcal{V}_c^{Q^*[W]}$. Suppose $u \in \mathcal{V}_c^{Q^*[W]}$. Then there is $W' \in Q^*[W]$ such that $u = v_c^{W'}$. Let $p \in \mathsf{Prop}$. Since $W, W' \in Q^*[W]$ and $Q^*[W]$ decides p in case c, we have $u \Vdash p$ iff $v_c^W \Vdash p$. Since p was arbitrary, $u = v_c^W$.

(2) implies (3): Applying Theorem 6.4.1 to each $\varphi \in \mathcal{L}_0$, there is a set $\Gamma_\varphi \subseteq \mathcal{L}_0$ such that $W, c \models \Gamma_\varphi$, $W \models \mathsf{E}_{\mathcal{S}}\Gamma_\varphi$, and either $\Gamma_\varphi \Vdash \varphi$ or $\Gamma_\varphi \Vdash \neg\varphi$. Set $\Gamma = \bigcup_{\varphi\in\mathcal{L}_0} \Gamma_\varphi$. Clearly $W, c \models \Gamma$ – so Γ is consistent – and $W \models \mathsf{E}_{\mathcal{S}}\Gamma$. To show $\mathrm{Cn}_0(\Gamma)$ is *maximally* consistent, suppose $\varphi \notin \mathrm{Cn}_0(\Gamma)$. From monotonicity of classical consequence and $\Gamma_\varphi \subseteq \Gamma$, we get $\varphi \notin \mathrm{Cn}_0(\Gamma_\varphi)$. Hence $\Gamma_\varphi \Vdash \neg\varphi$, and $\Gamma \Vdash \neg\varphi$ too. This means $\mathrm{Cn}_0(\Gamma) \cup \{\varphi\}$ is inconsistent, and we are done.

(3) implies (2): Take $\varphi \in \mathcal{L}_0$. Then we may apply Theorem 6.4.1 with Γ from (3) – noting that the maximal consistency property ensure either $\Gamma \Vdash \varphi$ or $\Gamma \models \neg\varphi$ – to see that $Q^*[W]$ decides φ in case c. □

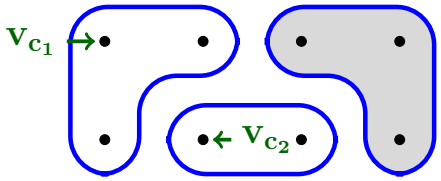


Figure 6.2: World W from Example 6.4.2. Note that for brevity we do not label the valuations.

6.4.2 Source Partitions

We now apply the analysis of the previous section to the set of source partitions $\{\Pi_i^W\}_{i\in\mathcal{S}}$. For $S \subseteq \mathcal{W}$ and $i \in \mathcal{S}$, write $\mathcal{P}_i^S = \{\Pi_i^W \mid S \in W\}$ for the i-partitions appearing in S. When $S = Q^*[W]$, these are exactly those partitions which agree with Π_i^W at each valuation v_c^W.

Lemma 6.4.3. $\Pi \in \mathcal{P}_i^{Q^*[W]}$ *if and only if* $\{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}} \subseteq \Pi$.

Proof. "if": Suppose $\{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}} \subseteq \Pi$. Let W' be obtained from W by setting i's partition to Π, keeping valuations and other source partitions the same. We claim $W' \approx W$. Indeed, take any $j \in \mathcal{S}$ and $c \in \mathcal{C}$. If $j \neq i$ then $\Pi_j^{W'} = \Pi_i^W$; since valuations are the same we get $\Pi_j^W[v_c^W] = \Pi_j^{W'}[v_c^{W'}]$. For $j = i$, note that since $\Pi_i^W[v_c^W] \in \Pi$ by assumption, and $v_c^W \in \Pi_i^W[v_c^W]$, we have $\Pi[v_c^W] = \Pi_i^W[v_c^W]$. By construction of W', this means $\Pi_i^W[v_c^W] = \Pi[v_c^{W'}] = \Pi_i^{W'}[v_c^{W'}]$. By Proposition 6.3.2, $W' \approx W$. Hence $\Pi \in \mathcal{P}_i^{Q^*[W]}$.

"only if": This is clear from Proposition 6.3.2. □

Example 6.4.2. *Suppose* $|\mathsf{Prop}| = 3$, $\mathcal{C} = \{c_1, c_2\}$ *and* $i \in \mathcal{S}$. *Consider a world W whose i-partition is shown in Fig. 6.2. By Lemma 6.4.3, a partition Π appears as $\Pi_i^{W'}$ for some $W' \approx W$ if and only if it contains the leftmost and bottommost sets. Any such Π consists of these cells together with a partition of the shaded area. Since there are 5 possible partitions of a 3-element set, it follows that* $|\mathcal{P}_i^{Q^*[W]}| = 5$.

Example 6.4.2 hints that if the cells containing the valuations v_c^W cover the whole space of valuations $\mathcal{V}$, or just omit a single valuation, then i's partition is uniquely defined in $Q^*[W]$. That is, truth-tracking methods can determine the full extent of i's expertise if the "actual" world is W. Indeed, we have the following analogue of Theorem 6.4.2 for partitions.

Theorem 6.4.3. *The following are equivalent.*

1. $\mathcal{P}_i^{Q^*[W]} = \{\Pi_i^W\}$.
2. $Q^*[W]$ *decides* $\mathsf{E}_i\varphi$ *for all* $\varphi \in \mathcal{L}_0$.
3. $|\mathcal{V} \setminus R| \leq 1$, *where* $R = \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]$.

Note that $R = \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]$ is the set of valuations indistinguishable from the actual state at some case c. Theorem 6.4.3 (3) says this set needs to essentially cover the whole space $\mathcal{V}$, omitting at most a single point. In this sense, it is easier to find Π_i^W uniquely when i has *less expertise*, since the cells $\Pi_i^W[v_c^W]$ will be larger. In the

extreme case where i has total expertise, i.e. $\Pi_i^W = \{\{v\} \mid v \in \mathcal{V}\}$, we need at least $2^{|\mathsf{Prop}|} - 1$ cases with distinct valuations in order to find Π_i^W exactly.

Example 6.4.3. *In Example 6.4.2 we have already seen an example of a world W for which $\mathcal{P}_i^{Q^*[W]}$ does not contain a unique partition. For a positive example, consider the world W from Fig. 6.1. Then $\mathcal{V} \setminus R_{\mathsf{D}} = \{\overline{pq}\}$ and $\mathcal{V} \setminus R_{\mathsf{T}} = \emptyset$, so both the partitions of D and T can be found uniquely by truth-tracking methods.*

The remainder of this section proves Theorem 6.4.3.

Lemma 6.4.4. *Let $i \in \mathcal{S}$ and $U \subseteq \mathcal{V}$. Then $U \subseteq \bigcup_{c \in \mathcal{C}} \Pi_i^W[v_c^W]$ and $W \approx W'$ implies $\Pi_i^W[U] = \Pi_i^{W'}[U]$.*

Proof. It suffices to show that for all $u \in U$ we have $\Pi_i^W[u] = \Pi_i^{W'}[u]$, since by definition $\Pi[U] = \bigcup_{u \in U} \Pi[u]$. Let $u \in U$. Then there is $c \in \mathcal{C}$ such that $u \in \Pi_i^W[v_c^W]$. Hence $\Pi_i^W[u] = \Pi_i^W[v_c^W]$. But since $W \approx W'$, $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$. This means $u \in \Pi_i^{W'}[v_c^{W'}]$, so $\Pi_i^{W'}[u] = \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W] = \Pi_i^W[u]$, as required, □

Lemma 6.4.5. *$Q^*[W]$ decides $\mathsf{E}_i\varphi$ if and only if, writing $R = \bigcup_{c \in \mathcal{C}} \Pi_i^W[v_c^W]$, either (i) $\|\varphi\| \subseteq R$; (ii) $\|\neg\varphi\| \subseteq R$; or (iii) there is some $c \in \mathcal{C}$ such that $\Pi_i^W[v_c^W]$ intersects with both $\|\varphi\|$ and $\|\neg\varphi\|$.*

Proof. "if": First suppose (i) holds. Take $W' \in Q^*[W]$. From $\|\varphi\| \subseteq R$, $W \approx W'$ and Lemma 6.4.4 we get $\Pi_i^W[\varphi] = \Pi_i^{W'}[\varphi]$. Consequently, $W' \models \mathsf{E}_i\varphi$ iff $W \models \mathsf{E}_i\varphi$. Since W' was arbitrary, either all worlds in $Q^*[W]$ satisfy $\mathsf{E}_i\varphi$, or all do not. Hence $Q^*[W]$ decides $\mathsf{E}_i\varphi$.

If (ii) holds, a similar argument shows that $Q^*[W]$ decides $\mathsf{E}_i\neg\varphi$. But it is easily checked that $\mathsf{E}_i\varphi \equiv \mathsf{E}_i\neg\varphi$, so $Q^*[W]$ also decides $\mathsf{E}_i\varphi$.

Finally, suppose (iii) holds. Then there is $c \in \mathcal{C}$ and $u \in \|\varphi\|$, $v \in \|\neg\varphi\|$ such that $u, v \in \Pi_i^W[v_c^W]$. We claim $Q^*[W] \models \neg\mathsf{E}_i\varphi$. Indeed, take $W' \in Q^*[W]$. Then $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$, so $u, v \in \Pi_i^{W'}[v_c^{W'}]$. In particular, u and v differ on φ but are contained in the same cell in $\Pi_i^{W'}$. Hence $W' \models \neg\mathsf{E}_i\varphi$.

"only if": We show the contrapositive. Suppose none of (i), (ii), (iii) hold. Then there is $u \in \|\varphi\| \setminus R$ and $v \in \|\neg\varphi\| \setminus R$. Let us define two worlds W_1, W_2 from W by modifying i's partition:

$$\Pi_i^{W_1} = \{\Pi_i^W[v_c^W]\}_{c \in \mathcal{C}} \cup \{\mathcal{V} \setminus R\},$$
$$\Pi_i^{W_2} = \{\Pi_i^W[v_c^W]\}_{c \in \mathcal{C}} \cup \{\{w\} \mid w \in \mathcal{V} \setminus R\}.$$

Then $W_1, W_2 \in Q^*[W]$ by Lemma 6.4.3. We claim that $W_1 \models \neg\mathsf{E}_i\varphi$ but $W_2 \models \mathsf{E}_i\varphi$, which will show $Q^*[W]$ does not decide $\mathsf{E}_i\varphi$.

First, note that since $u, v \notin R$, we have $\Pi_i^{W_1}[u] = \Pi_i^{W_1}[v] = \mathcal{V} \setminus R$. Since u and v differ on φ but share the same partition cell, $W_1 \models \neg\mathsf{E}_i\varphi$.

To show $W_2 \models \mathsf{E}_i\varphi$, take $w \in \|\varphi\|$. If $w \notin R$ then $\Pi_i^{W_2}[w] = \{w\} \subseteq \|\varphi\|$. Otherwise there is $c \in \mathcal{C}$ such that $w \in \Pi_i^W[v_c^W]$. Thus $\Pi_i^W[v_c^W]$ intersects with $\|\varphi\|$. Since (iii) does not hold, this in fact implies $\Pi_i^W[v_c^W] \subseteq \|\varphi\|$, and consequently $\Pi_i^{W_2}[w] = \Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Since $w \in \|\varphi\|$ was arbitrary, we have shown $\Pi_i^{W_2}[\varphi] = \bigcup_{w \in \|\varphi\|} \Pi_i^{W_2}[w] \subseteq \|\varphi\|$. Since the reverse inclusion always holds, this shows $W_2 \models \mathsf{E}_i\varphi$, and we are done. □

Proof of Theorem 6.4.3. The implication (1) to (2) is clear since if $W' \in Q^*[W]$ then $\Pi_i^{W'} = \Pi_i^W$ by (1), so $W' \models \mathsf{E}_i\varphi$ iff $W \models \mathsf{E}_i\varphi$, and thus $Q^*[W]$ decides $\mathsf{E}_i\varphi$.

To show (2) implies (3) we show the contrapositive. Suppose $|\mathcal{V} \setminus R| > 1$. Then there are distinct $u, v \in \mathcal{V} \setminus R$. Let φ be any propositional formula with $\|\varphi\| = \{u\}$. We show by Lemma 6.4.5 that $Q^*[W]$ does not decide $\mathsf{E}_i\varphi$. Indeed, all three conditions fail: $\|\varphi\| \not\subseteq R$ (since $u \notin R$), $\|\neg\varphi\| \not\subseteq R$ (since $v \in \|\neg\varphi\| \setminus R$) and no $\Pi_i^W[v_c^W]$ intersects with $\|\varphi\|$ (otherwise $u \in \Pi_i^W[v_c^W] \subseteq R$).

Finally, for (3) implies (1) we also show the contrapositive. Suppose there is $\Pi \in \mathcal{P}_i^{Q^*[W]} \setminus \{\Pi_i^W\}$. Write $\mathcal{R} = \{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}}$, so that $\mathcal{R}$ is a partition of R. By Lemma 6.4.3, $\mathcal{R} \subseteq \Pi$. Note that $\mathcal{R} \subseteq \Pi_i^W$ too. Since $\Pi \neq \Pi_i^W$, we in fact have $\mathcal{R} \subset \Pi$ and $\mathcal{R} \subset \Pi_i^W$. Hence $\Pi \setminus \mathcal{R}$ and $\Pi_i^W \setminus \mathcal{R}$ are distinct partitions of $\mathcal{V} \setminus R$. Since a one-element set has a unique partition, $\mathcal{V} \setminus R$ must contain at least two elements. □

6.4.3 Learning the Actual World Exactly

Putting Theorems 6.4.2 and 6.4.3, we obtain a precise characterisation of when W can be found *exactly* by truth-tracking methods, i.e when $Q^*[W] = \{W\}$.

Corollary 6.4.1. *$Q^*[W] = \{W\}$ if and only if*

1. *There is a collection $\{\Gamma_c\}_{c\in\mathcal{C}} \subseteq \mathcal{L}_0^{\mathcal{C}}$ such that for each c, (i) $W, c \models \Gamma_c$; (ii) $W \models \mathsf{E}_{\mathcal{S}}\Gamma_c$; (iii) $\mathrm{Cn}_0(\Gamma_c)$ is maximally consistent; and*
2. *For each each $i \in \mathcal{S}$, $|\mathcal{V} \setminus \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]| \leq 1$.*

6.5 Truth-Tracking Methods

So far we have focussed on solvable questions, and the extent to which they reveal information about the actual world. We now turn to the methods which solve them. We give a general characterisation of truth-tracking methods under mild assumptions, before discussing the family of *conditioning* methods from Singleton and Booth [83].

6.5.1 A General Characterisation

For sequences σ, δ, write $\sigma \equiv \delta$ iff δ is obtained from σ by replacing each report $\langle i, c, \varphi\rangle$ with $\langle i, c, \psi\rangle$, for some $\psi \equiv \varphi$. For $k \in \mathbb{N}$, let σ^k denote the k-fold repetition of σ. Consider the following properties which may hold of a learning method L.

Equivalence. If $\sigma \equiv \delta$ then $L(\sigma) = L(\delta)$.

Repetition. $L(\sigma^k) = L(\sigma)$.

Soundness. $L(\sigma) \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$.

Equivalence says that L should not care about the syntactic form of the input. **Repetition** says that beliefs should not change if each source repeats their reports k times. **Soundness** says that all reports in σ are believed to be sound.

For methods satisfying these properties, we have a precise characterisation of truth-tracking, i.e. necessary and sufficient conditions for L to solve Q^*. First, some

new notation is required. Write $\delta \preceq \sigma$ iff for each $\langle i, c, \varphi\rangle \in \delta$ there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. That is, σ contains everything δ does, up to logical equivalence. Set

$$T_\sigma = \mathcal{X}^{\mathsf{snd}}_\sigma \setminus \bigcup \left\{ \mathcal{X}^{\mathsf{snd}}_\delta \mid \delta \not\preceq \sigma \right\} \subseteq \mathcal{W}.$$

Then $W \in T_\sigma$ iff σ is sound for W and any δ sound for W has $\delta \preceq \sigma$. In this sense σ contains *all* soundness statements for W – up to equivalence – so can be seen as a finite version of a stream. Let us call σ a *psuedo-stream* for W whenever $W \in T_\sigma$.

Theorem 6.5.1. *A method L satisfying **Equivalence**, **Repetition** and **Soundness** is truth-tracking if and only if it satisfies the following property.*

Credulity. $T_\sigma, c \not\models \mathsf{S}_i\varphi \implies L(\sigma), c \models \neg\mathsf{S}_i\varphi$.

Before the proof, we comment on our interpretation of **Credulity**. It says that whenever $\neg\mathsf{S}_i\varphi$ is consistent with T_σ – those W for which σ is a psuedo-stream – L should believe $\neg\mathsf{S}_i\varphi$. Since the number of sound statements *decreases* with increasing expertise, this is a principle of *maximal trust*: we should believe i has the expertise to rule out φ in case c, whenever this is consistent with T_σ. That is, some amount of *credulity* is required to find the truth. Our assumption that learning methods receive complete streams ensures that, if a source in fact lacks this expertise, they will eventually report φ and this belief can be be retracted. A stronger version of **Credulity** spells this out explicitly in terms of expertise:

$$\forall \sigma, i, c, \varphi : \; T_\sigma, c \not\models \neg\mathsf{E}_i\varphi \implies L(\sigma), c \models \mathsf{E}_i\varphi. \tag{6.1}$$

(6.1) implies **Credulity** in the presence of **Soundness**, and is thus a sufficient condition for truth-tracking (when also taken with **Equivalence** and **Repetition**).[5]

Theorem 6.5.1 also shows truth-tracking cannot be performed *deductively*: the method $L(\sigma) = \mathcal{X}^{\mathsf{snd}}_\sigma$, which does not go beyond the mere information that each report is sound, fails **Credulity**. Some amount of *inductive* or *non-monotonic* reasoning, as captured by **Credulity**, is necessary.

The rest of this section works towards the proof of Theorem 6.5.1. We collect some useful properties of psuedo-streams. First, psuedo-streams provide a way of accessing Q^* via a finite sequence: T_σ is a cell in Q^* whenever it is non-empty.

Lemma 6.5.1. *If $W \in T_\sigma$, then (i) $W' \in \mathcal{X}^{\mathsf{snd}}_\sigma$ iff $W \sqsubseteq W'$; and (ii) $T_\sigma = Q^*[W]$.*

Proof. Suppose $W \in T_\sigma$. For (i), first suppose $W' \in \mathcal{X}^{\mathsf{snd}}_\sigma$ and $W, c \models \mathsf{S}_i\varphi$. Considering the singleton sequence $\delta = \langle i, c, \varphi\rangle$ we have $W \in \mathcal{X}^{\mathsf{snd}}_\delta$. From $W \in T_\sigma$ we get $\delta \preceq \sigma$, i.e. there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. From $W' \in \mathcal{X}^{\mathsf{snd}}_\sigma$ and $\mathsf{S}_i\varphi \equiv \mathsf{S}_i\psi$ we get $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$.

Now suppose $W \sqsubseteq W'$ and let $\langle i, c, \varphi\rangle \in \sigma$. Then since $W \in T_\sigma \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$ we have $W, c \models \mathsf{S}_i\varphi$, and $W \sqsubseteq W'$ gives $W', c \models \mathsf{S}_i\varphi$. Consequently $W' \in \mathcal{X}^{\mathsf{snd}}_\sigma$.

Now for (ii), first suppose $W' \in Q^*[W]$. Then W and W' satisfy exactly the same soundness statements, so $W' \in T_\sigma$ also. Conversely, suppose $W' \in T_\sigma$. Then $W' \in \mathcal{X}^{\mathsf{snd}}_\sigma$, so (i) gives $W \sqsubseteq W'$. But we also have $W' \in T_\sigma$ and $W \in \mathcal{X}^{\mathsf{snd}}_\sigma$, so (i) again gives $W' \sqsubseteq W$. Hence $W \approx W'$, i.e. $W' \in Q^*[W]$. □

[5] We conjecture (6.1) is strictly stronger than **Credulity**.

The next two results show that initial segments of streams are (eventually) psuedo-streams, and that any psuedo-stream gives rise to a stream.

Lemma 6.5.2. *If ρ is a stream for W, there is n such that $W \in T_{\rho[m]}$ for all $m \geq n$.*

Proof. Let $\widehat{\cdot}$ be a function which selects a representative formula for each equivalence class of $\mathcal{L}_0/\equiv$, so that $\varphi \equiv \widehat{\varphi}$ and $\varphi \equiv \psi$ implies $\widehat{\varphi}$ is equal to $\widehat{\psi}$. Note that since Prop is finite, and since $\mathcal{S}$ and $\mathcal{C}$ are also finite, there are only finitely many reports of the form $\langle i, c, \widehat{\varphi}\rangle$. By completeness of ρ for W, we may take n sufficiently large so that $W, c \models \mathsf{S}_i\widehat{\varphi}$ implies $\langle i, c, \widehat{\varphi}\rangle \in \rho[n]$, for all i, c, φ. Now, take $m \geq n$. We need to show $W \in T_{\rho[m]}$. Clearly $W \in \mathcal{X}^{\mathsf{snd}}_{\rho[m]}$, since ρ is sound for W. Suppose $W \in \mathcal{X}^{\mathsf{snd}}_{\delta}$. We need to show $\delta \preceq \rho[m]$. Indeed, take $\langle i, c, \varphi\rangle \in \delta$. Then $W, c \models \mathsf{S}_i\varphi$. Since $\mathsf{S}_i\varphi \equiv \mathsf{S}_i\widehat{\varphi}$, we have $W, c \models \mathsf{S}_i\widehat{\varphi}$. Hence $\langle i, c, \widehat{\varphi}\rangle$ appears in $\rho[n]$, and consequently in $\rho[m]$ too. Since $\varphi \equiv \widehat{\varphi}$, this shows $\delta \preceq \rho[m]$. □

Lemma 6.5.3. *If $W \in T_\sigma$ and $N = |\sigma|$, there is a stream ρ for W such that $\rho[Nk] \equiv \sigma^k$ for all $k \in \mathbb{N}$.*

Proof. First note that $W \in T_\sigma$ implies $\sigma \neq \emptyset$, so $N > 0$. Since $\mathcal{L}_0$ is countable, we may index the set of $\mathcal{L}_0$ formulas equivalent to $\varphi \in \mathcal{L}_0$ as $\{\varphi_n\}_{n\in\mathbb{N}}$. Let σ_n be obtained from σ by replacing each report $\langle i, c, \varphi\rangle$ with $\langle i, c, \varphi_n\rangle$. Then $\sigma \equiv \sigma_n$. Let ρ be the sequence obtained as the infinite concatenation $\sigma_1 \circ \sigma_2 \circ \sigma_3 \circ \cdots$ (this is possible since σ is of positive finite length). Then $\rho[Nk] = \sigma_1 \circ \cdots \circ \sigma_k$, and consequently $\rho[Nk] \equiv \sigma^k$.

It remains to show ρ is a stream for W. Soundness of ρ follows from $W \in T_\sigma \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$, since every report in ρ is equivalent to some report in σ by construction. For completeness, suppose $W, c \models \mathsf{S}_i\varphi$. As in the proof of Lemma 6.5.1, considering the singleton sequence $\delta = \langle i, c, \varphi\rangle$, we get from $W \in T_\sigma$ that there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. Hence there is $n \in \mathbb{N}$ such that $\varphi = \psi_n$, so $\langle i, c, \varphi\rangle \in \sigma_n$, and thus $\langle i, c, \varphi\rangle \in \rho$. □

Next we obtain an equivalent formulation of **Credulity** which is less transparent as a postulate for learning methods, but easier to work with.

Lemma 6.5.4. *Suppose L satisfies **Soundness**. Then L satisfies **Credulity** if and only if $L(\sigma) \subseteq T_\sigma$ for all σ with $T_\sigma \neq \emptyset$.*

Proof. "if": Suppose $T_\sigma, c \not\models \mathsf{S}_i\varphi$. Then there is $W \in T_\sigma$ such that $W, c \not\models \mathsf{S}_i\varphi$. By our assumption and Lemma 6.5.1, $L(\sigma) \subseteq T_\sigma = Q^*[W]$. Thus every world in $L(\sigma)$ agrees with W on soundness statements, so $L(\sigma), c \models \neg\mathsf{S}_i\varphi$.

"only if": Suppose there is some $W \in T_\sigma$, and take $W' \in L(\sigma)$. We need to show $W' \in T_\sigma$; by Lemma 6.5.1, this is equivalent to $W \approx W'$. First suppose $W, c \models \mathsf{S}_i\varphi$. Then $W \in T_\sigma$ implies there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. By **Soundness** for L, we have $W' \in L(\sigma) \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$. Consequently $W', c \models \mathsf{S}_i\psi$ and thus $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$. Now suppose $W, c \not\models \mathsf{S}_i\varphi$. Then $T_\sigma, c \not\models \mathsf{S}_i\varphi$. By **Credulity**, $L(\sigma), c \models \neg\mathsf{S}_i\varphi$. Hence $W', c \not\models \mathsf{S}_i\varphi$. This shows $W' \sqsubseteq W$. Thus $W \approx W'$ as required. □

Finally, we prove the characterisation of truth-tracking.

Proof of Theorem 6.5.1. Suppose L satisfies **Equivalence**, **Repetition** and **Soundness**.

"if": Suppose **Credulity** holds. We show L solves Q^*. Take any world W and stream ρ for W. By Lemma 6.5.2, there is n such that $W \in T_{\rho[m]}$ for all $m \geq n$. By Lemma 6.5.1, $T_{\rho[m]} = Q^*[W]$ for such m. In particular, $T_{\rho[m]} \neq \emptyset$. By **Credulity** and Lemma 6.5.4, we get $L(\rho[m]) \subseteq T_{\rho[m]} = Q^*[W]$.

"only if": Suppose L solves Q^*. We show **Credulity** via Lemma 6.5.4. Suppose there is some $W \in T_\sigma$, and write $N = |\sigma| > 0$. By Lemma 6.5.3, there is a stream ρ for W such that $\rho[Nk] \equiv \sigma^k$ for all $k \in \mathbb{N}$. By **Repetition** and **Equivalence**, $L(\sigma) = L(\sigma^k) = L(\rho[Nk])$. But L solves Q^*, so for k sufficiently large we have $L(\rho[Nk]) \subseteq Q^*[W] = T_\sigma$. Hence, going via some large k, we obtain $L(\sigma) \subseteq T_\sigma$ as required. □

6.5.2 Conditioning Methods

In this section we turn to the family of *conditioning* methods, proposed in [83] and inspired by similar methods in the belief change literature [87]. While our interpretation of input sequences is different – we read $\langle i, c, \varphi\rangle$ as i reporting φ is *possible* in case c, whereas Singleton and Booth [83] read this as i *believes* φ – this class of methods can still be applied in our setting.

Conditioning methods operate by successively restricting a fixed *plausibility total preorder*[6] to the information corresponding to each new report $\langle i, c, \varphi\rangle$. In this chapter, we take a report $\langle i, c, \varphi\rangle$ to correspond to the fact that $\mathsf{S}_i\varphi$ holds in case c; this fits with our assumption throughout that sources only report sound statements.[7] Thus, the worlds under consideration given a sequence σ are exactly those satisfying all soundness statements in σ, i.e. $\mathcal{X}^{\mathsf{snd}}_\sigma$. Note that $\mathcal{X}^{\mathsf{snd}}_\sigma$ represents the *indefeasible knowledge* given by σ: worlds outside $\mathcal{X}^{\mathsf{snd}}_\sigma$ are eliminated and cannot be recovered with further reports, since $\mathcal{X}^{\mathsf{snd}}_{\sigma\circ\delta} \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$. The plausibility order allows us to represent *defeasible beliefs* about the most plausible worlds within $\mathcal{X}^{\mathsf{snd}}_\sigma$.

Definition 6.5.1. *For a total preorder $\leq$ on $\mathcal{W}$, the* conditioning method $L_\leq$ *is given by* $L_\leq(\sigma) = \min_\leq \mathcal{X}^{\mathsf{snd}}_\sigma$.

Note that since $\mathcal{X}^{\mathsf{snd}}_\sigma \neq \emptyset$ for all σ[8] and $\mathcal{W}$ is finite, $L_\leq$ is consistent. Moreover, $L_\leq$ satisfies **Equivalence**, **Repetition** and **Soundness**.

Example 6.5.1. *We recall two concrete choices of $\leq$ from Singleton and Booth [83].*

1. *Set $W \leq W'$ iff $r(W) \leq r(W')$, where*

$$r(W) = -\sum_{i\in\mathcal{S}} |\{p \in \mathsf{Prop} \mid \Pi^W_i[p] = \|p\|\}|.$$

The most plausible worlds in this order are those in which source have as much expertise on the propositional variables as possible, on aggregate. We denote the corresponding conditioning method by L_{vbc}, standing for variable-based conditioning.

[6] A total preorder is a reflexive, transitive and total relation.

[7] Singleton and Booth [83] consider more general conditioning methods in which this choice is not fixed.

[8] For example, if $\Pi^W_i = \{\mathcal{V}\}$ for all i then $W \in \mathcal{X}^{\mathsf{snd}}_\sigma$ for all σ.

2. *Set* $W \leq W'$ *iff* $r(W) \leq r(W')$*, where*

$$r(W) = -\sum_{i \in \mathcal{S}} |\Pi_i^W|.$$

This order aims to maximise the number of cells in each source's partitions, thereby maximising the number of propositions on which they have expertise. Note that the propositional variables play no special role. We denote the corresponding conditioning operator by L_{pbc}*, for* partition-based conditioning.

A straightforward property of $\leq$ characterises truth-tracking for conditioning methods. For a generic total preorder $\leq$, let $<$ denote its strict part.

Theorem 6.5.2. *$L_{\leq}$ is truth-tracking if and only if*

$$W \sqsubset W' \implies \exists W'' \approx W \text{ such that } W'' < W'. \tag{6.2}$$

Like **Credulity**, (6.2) is a principle of maximising trust in sources. Recall from that Lemma 6.3.1 that $W \sqsubset W'$ means all sources are more knowledgeable in each case in W than in W', and there is at least one source and case for which this holds strictly. If we aim to trust sources as much as possible, we might impose $W < W'$ here; then W' is strictly less plausible and will be ruled out in favour of W. This yields a sufficient condition for truth-tracking, but to obtain a necessary condition we need to allow a "surrogate" world $W'' \approx W$ to take the place of W.

Proof of Theorem 6.5.2. Write $L = L_{\leq}$. Since L satisfies **Equivalence**, **Repetition** and **Soundness**, we may use Theorem 6.5.1. Furthermore, it is sufficient by Lemma 6.5.4 to show that (6.2) holds if and only if $L(\sigma) \subseteq T_\sigma$, whenever $T_\sigma \neq \emptyset$.

"if": Suppose $W \sqsubset W'$. Let σ be some pseudo-stream for W, so that $W \in T_\sigma$.[9] Note that since $W \in T_\sigma \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ and $W \sqsubset W'$, we have $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ also. By assumption, $L(\sigma) \subseteq T_\sigma = Q^*[W]$. Since $W \not\approx W'$, this means $W' \in \mathcal{X}_\sigma^{\mathsf{snd}} \setminus L(\sigma)$. That is, W' lies in $\mathcal{X}_\sigma^{\mathsf{snd}}$ but is not $\leq$-minimal. Consequently there is $W'' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ such that $W'' < W'$. Since L is consistent, we may assume without loss of generality that $W'' \in L(\sigma)$. Hence $W'' \in Q^*[W]$, so $W'' \approx W$.

"only if": Suppose there is some $W \in T_\sigma$, and let $W' \in L(\sigma)$. We need to show $W' \in T_\sigma = Q^*[W]$, i.e. $W \approx W'$. Since $W' \in L(\sigma) \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$, Lemma 6.5.1 gives $W \sqsubseteq W'$. Suppose for contradiction that $W \not\approx W'$. Then $W \sqsubset W'$. By (6.2), there is $W'' \approx W$ such that $W'' < W'$. But W' is $\leq$-minimal in $\mathcal{X}_\sigma^{\mathsf{snd}}$, so this must mean $W'' \notin \mathcal{X}_\sigma^{\mathsf{snd}}$. On the other hand, $W'' \in Q^*[W] = T_\sigma \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$: contradiction. □

Example 6.5.2. *We revisit the methods of Example 6.5.1.*

1. *The variable-based conditioning method* L_{vbc} *is* not *truth-tracking. Indeed, consider the worlds* W *and* W' *shown in Fig. 6.3, where we assume* $\mathsf{Prop} = \{p, q\}$*,* $\mathcal{S} = \{i\}$ *and* $\mathcal{C} = \{c\}$*. Then* $W \sqsubset W'$ *(e.g. by Lemma 6.3.1). Note that* i *does not have expertise on* p *or* q *in both* W *and* W'*, so* $r(W) = r(W') = 0$*. Moreover,* i*'s partition is uniquely determined in* $Q^*[W]$ *by Theorem 6.4.3, so if* $W'' \approx W$ *then* $r(W'') = 0$ *also. That is, there is no* $W'' \approx W$ *such*

[9] For example, pick some stream ρ and apply Lemma 6.5.2 to obtain a pseudo-stream.

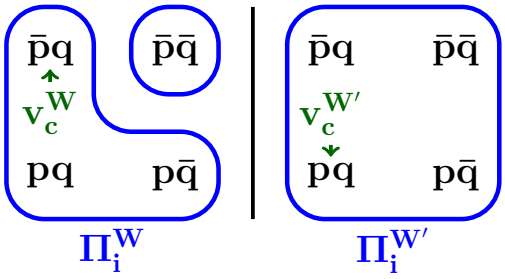


Figure 6.3: Worlds which demonstrate L_{vbc} is not truth-tracking.

that $W'' < W'$. Hence (6.2) fails, and L_{vbc} is not truth-tracking. Intuitively, the problem here is that since i's expertise is not split along the lines of the propositional variables when W is the actual world, L_{vbc} will always maintain W' as a possibility.

2. *The partition-based conditioning method L_{pbc}* is *truth-tracking. Indeed, if $W \sqsubset W'$ we may construct W'' from W by modifying the partition of each source i so that all valuations outside of $\bigcup_{c\in C} \Pi_i^W[v_c^W]$ lie in their own cell. Then $W \approx W''$. One can show that $\Pi_i^{W''}$ refines $\Pi_i^{W'}$ for all $i \in \mathcal{S}$, and there is some i for which the refinement is strict. Hence the partitions in W'' contain strictly more cells, so $W'' < W'$.*

6.6 Conclusion

Summary. In this chapter we studied truth-tracking in the presence of non-expert sources. The model assumes sources report everything true *up to their lack of expertise*, i.e. all that they consider possible. We obtained precise characterisations of when truth-tracking methods can uniquely find the valuations and partitions of a world W. We then gave a postulational characterisation of truth-tracking methods under mild assumptions, before looking specifically at the conditioning methods of Singleton and Booth [83].

Limitations and future work. Conceptually, the assumption that streams are complete is very strong. As seen in Example 6.2.1, completeness requires sources to give jointly inconsistent reports whenever $\Pi_i[v_c]$ contains more than just v_c. Such reports provide information about the source's expertise: if i reports both φ and $\neg\varphi$ we know $\neg\mathsf{E}_i\varphi$. To provide all sound reports sources must also have *negative introspection* over their own knowledge, i.e. they *know* when they do not know something. Indeed, our use of partitions makes expertise closely related to S5 knowledge [83, 79], which has been criticised in the philosophical literature as too strong. In reality, non-expert sources may have *beliefs* about the world, and may prefer to report only that which they believe. A source may even believe a sound report φ is *false*, since soundness only says the source does not *know* $\neg\varphi$. For example, in Example 6.0.1 the doctor D may think it is more likely that A suffers from p than q, but we cannot express this in our framework.

On the technical side, our results on solvability of Q^* and the characterisation of Theorem 6.5.1 rely on the fact that we only consider finitely many worlds. In

a sense this trivialises the problem of induction as studied by Kelly, Schulte, and Hendricks [57] and Baltag, Gierasimczuk, and Smets [5], among others. In future work it would be interesting to see which results can be carried over to the case where Prop is infinite.

7 Conclusion

7.1 Summary

7.2 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 78, 113, 114, 137, 157).

[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 5, 11).

[3] Milica Anđelić et al. "Some new considerations about double nested graphs". In: *Linear Algebra and its Applications* 483 (2015), pp. 323–341 (cited on page 76).

[4] Kenneth J. Arrow. "Social Choice and Individual Values". In: *Ethics* 62.3 (1952), pp. 220–222 (cited on page 20).

[5] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. "On the Solvability of Inductive Problems: A Study in Epistemic Topology". In: *Electronic Proceedings in Theoretical Computer Science* 215 (June 2016), pp. 81–98. ISSN: 2075-2180. DOI: 10.4204/eptcs.215.7. URL: http://dx.doi.org/10.4204/EPTCS.215.7 (cited on pages 159, 161, 174).

[6] Alexandru Baltag, Nina Gierasimczuk, and Sonja Smets. "Truth-Tracking by Belief Revision". In: *Studia Logica* 107.5 (Oct. 2019), pp. 917–947. ISSN: 1572-8730. DOI: 10.1007/s11225-018-9812-x. URL: https://doi.org/10.1007/s11225-018-9812-x (cited on pages 156, 157, 159).

[7] 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 108).

[8] Alexandru Baltag et al. "A dynamic logic for learning theory". In: *Journal of Logical and Algebraic Methods in Programming* 109 (2019), p. 100485 (cited on page 156).

[9] Johan van Benthem and Guram Bezhanishvili. "Modal logics of space". In: *Handbook of spatial logics*. Springer, 2007, pp. 217–298 (cited on pages 87, 98).

[10] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*. Vol. 53. Cambridge University Press, 2002 (cited on page 98).

[11] Christian Blum and Christina Isabel Zuber. “Liquid democracy: Potentials, problems, and perspectives”. In: *Journal of Political Philosophy* 24.2 (2016), pp. 162–182 (cited on page 79).

[12] Richard Booth and Aaron Hunter. “Trust as a Precursor to Belief Revision”. In: *JAIR* 61 (2018), pp. 699–722 (cited on pages 114, 117, 149, 155).

[13] Denis Bouyssou. “Monotonicity of ‘ranking by choosing’: A progress report”. In: *Social Choice and Welfare* 23.2 (2004), pp. 249–273 (cited on pages 61, 65).

[14] 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 35).

[15] 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 pages 40, 41).

[16] 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 35, 37, 39, 76).

[17] Michelene TH Chi, Robert Glaser, and Marshall J Farr. *The nature of expertise*. Psychology Press, 2014 (cited on page 79).

[18] Zoé Christoff and Davide Grossi. “Binary Voting with Delegable Proxy: An Analysis of Liquid Democracy”. In: *Proc. TARK 2017*. 2017 (cited on pages 4, 20).

[19] Harry Collins and Robert Evans. *Rethinking expertise*. University of Chicago Press, 2008 (cited on page 79).

[20] Nicolas de Condorcet. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. 1785 (cited on page 156).

[21] David Cook II. “Nested colorings of graphs”. In: *Australasian Journal of Combinatorics* 62.1 (2015), pp. 100–127 (cited on page 76).

[22] László Csató. “An impossibility theorem for paired comparisons”. In: *Central European Journal of Operations Research* 27.2 (2019), pp. 497–514 (cited on pages 36, 37).

[23] James P Delgrande, Didier Dubois, and Jérôme Lang. “Iterated Revision as Prioritized Merging.” In: *KR* 6 (2006), pp. 210–220 (cited on pages 123, 155).

[24] 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 121, 128, 129).

[25] 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 35).

[26] José van Dijck and Donya Alinejad. “Social Media and Trust in Scientific Expertise: Debating the Covid-19 Pandemic in The Netherlands”. In: *Social Media + Society* 6.4 (2020) (cited on page 79).

[27] 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 5).

[28] 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 pages 80, 109).

[29] Elad Dokow and Ron Holzman. "Aggregation of binary evaluations with abstentions". In: *Journal of Economic Theory* 145 (2010), pp. 544–561 (cited on page 4).

[30] 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 3).

[31] Pål Grønås Drange et al. "On the threshold of intractability". In: *Algorithms-ESA 2015*. Springer, 2015, pp. 411–423 (cited on pages 39, 76).

[32] 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 page 8).

[33] 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 40, 41, 76, 156).

[34] 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 4, 20).

[35] K Anders Ericsson and Tyler J Towne. "Expertise." In: *WIREs Cognitive Science* (2010) (cited on page 79).

[36] 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 40).

[37] 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 156).

[38] Ronald Fagin et al. *Reasoning about knowledge*. MIT press, 2003 (cited on pages 80, 101, 102, 104).

[39] Eduardo L Fermé and Sven Ove Hansson. "Selective revision". In: *Studia Logica* 63.3 (1999), pp. 331–342 (cited on pages 114, 116, 144, 149, 155).

[40] 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 5, 8).

[41] 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 35).

[42] Ebrahim Ghorbani. *Some spectral properties of chain graphs*. 2017. arXiv: 1703.03581 [math.CO] (cited on page 76).

[43] E Mark Gold. “Language identification in the limit”. In: *Information and Control* 10.5 (May 1967), pp. 447–474. DOI: 10.1016/s0019-9958(67)91165-5 (cited on page 156).

[44] Alvin I Goldman. “Expertise”. In: *Topoi* 37.1 (2018), pp. 3–10 (cited on page 79).

[45] Julio González-Díaz, 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 35–37, 40, 41, 46, 76).

[46] 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 81).

[47] 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 154).

[48] Bernard Grofman, Guillermo Owen, and Scott L Feld. “Thirteen theorems in search of the truth”. In: *Theory and decision* 15.3 (1983), pp. 261–278 (cited on page 156).

[49] Adam Grove. “Two modellings for theory change”. In: *Journal of philosophical logic* (1988), pp. 157–170 (cited on page 137).

[50] Sven Ove Hansson. “A survey of non-prioritized belief revision”. In: *Erkenntnis* 50.2 (1999), pp. 413–427 (cited on page 114).

[51] 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 114, 155).

[52] Stephan Hartmann and Jan Sprenger. “Judgment aggregation and the problem of tracking the truth”. In: *Synthese* 187.1 (July 2012), pp. 209–221. ISSN: 1573-0964. DOI: 10.1007/s11229-011-0031-5. URL: https://doi.org/10.1007/s11229-011-0031-5 (cited on page 156).

[53] 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 115, 155).

[54] Sanjay Jain et al. *Systems that learn: an introduction to learning theory*. MIT, 1999 (cited on page 156).

[55] 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 35, 36, 39, 45, 76, 77).

[56] 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 137).

[57] Kevin Kelly, Oliver Schulte, and Vincent Hendricks. "Reliable belief revision". In: *Logic and Scientific Methods*. Springer, 1997, pp. 383–398 (cited on pages 156, 174).

[58] Daniel Kilov. "The brittleness of expertise and why it matters". In: *Synthese* 199.1 (2021), pp. 3431–3455 (cited on page 79).

[59] 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 8, 29).

[60] 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 154).

[61] 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 78, 113, 115, 154, 157).

[62] 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 128).

[63] 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, 35, 156).

[64] 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 3, 5–8, 10).

[65] Sue Llewellyn. "Covid-19: how to be careful with trust and expertise on social media". In: *BMJ* 368 (2020) (cited on page 79).

[66] Kenneth O May. "A set of independent necessary and sufficient conditions for simple majority decision". In: *Econometrica: Journal of the Econometric Society* (1952), pp. 680–684 (cited on page 17).

[67] Richard Montague. "Universal grammar". In: *Theoria* 36.3 (1970), pp. 373–398 (cited on page 82).

[68] 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 77).

[69] 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 76).

[70] Catherine Olsson et al. *Skill Rating for Generative Models*. 2018. arXiv: 1808.04888 [stat.ML] (cited on page 35).

[71] Aybüke Özgün. "Evidence in Epistemic Logic : A Topological Perspective". Theses. Université de Lorraine, Oct. 2017 (cited on page 87).

[72] Eric Pacuit. *Neighborhood semantics for modal logic*. Springer International Publishing, 2017 (cited on page 82).

[73] 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 3, 5, 6, 8, 29).

[74] Jan Plaza. "Logics of public communications". In: *Synthese* 158.2 (2007), pp. 165–179 (cited on page 109).

[75] Theodoros Rekatsinas et al. "SLiMFast: Guaranteed results for data fusion and source reliability". In: *Proceedings of the 2017 ACM International Conference on Management of Data*. 2017, pp. 1399–1414 (cited on page 114).

[76] Ariel Rubinstein. "Ranking the participants in a tournament". In: *SIAM Journal on Applied Mathematics* 38.1 (1980), pp. 108–111 (cited on pages 37, 45, 46).

[77] 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 123).

[78] Dana Scott. "Advice on Modal Logic". In: *Philosophical Problems in Logic: Some Recent Developments*. Ed. by Karel Lambert. Dordrecht: Springer Netherlands, 1970, pp. 143–173. ISBN: 978-94-010-3272-8. DOI: 10.1007/978-94-010-3272-8_7 (cited on page 82).

[79] Joseph Singleton. "A Logic of Expertise". In: *ESSLLI 2021 Student Session* (2021). URL: https://arxiv.org/abs/2107.10832 (cited on pages v, 154, 158, 173).

[80] 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 page 35).

[81] 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 pages v, 36).

[82] Joseph Singleton and Richard Booth. "Towards an axiomatic approach to truth discovery". In: *Autonomous Agents and Multi-Agent Systems* 36.2 (2022), pp. 1–49 (cited on page v).

[83] Joseph Singleton and Richard Booth. "Who's the Expert? On Multi-source Belief Change". In: (2022). https://arxiv.org/abs/2205.00077 (cited on pages 156, 158, 168, 171, 173).

[84] Joseph Singleton and Richard Booth. "Who's the Expert? On Multi-source Belief Change". In: *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*. Aug. 2022, pp. 331–340. DOI: 10.24963/kr.2022/33. URL: https://doi.org/10.24963/kr.2022/33 (cited on page v).

[85] 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 36, 37, 77).

[86] Giora Slutzki and Oscar Volij. "Scoring of web pages and tournaments—axiomatizations". In: *Social Choice and Welfare* 26.1 (2006), pp. 75–92 (cited on page 35).

[87] Wolfgang Spohn. "Ordinal conditional functions: A dynamic theory of epistemic states". In: *Causation in decision, belief change, and statistics*. Springer, 1988, pp. 105–134 (cited on page 171).

[88] Anne K Steiner. "The lattice of topologies: structure and complementation". In: *Transactions of the American Mathematical Society* 122.2 (1966), pp. 379–398 (cited on page 103).

[89] Zoi Terzopoulou and Ulle Endriss. "Optimal Truth-Tracking Rules for the Aggregation of Incomplete Judgments". In: *Proceedings of the 12th International Symposium on Algorithmic Game Theory (SAGT-2019)*. Sept. 2019 (cited on page 156).

[90] Hans van Ditmarsch et al., eds. *Handbook of Epistemic Logic*. English. College Publications, 2015. ISBN: 978-1-84890-158-2 (cited on page 154).

[91] 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 3, 40).

[92] 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 79).

[93] Sergi Xaudiera and Ana S Cardenal. "Ibuprofen narratives in five European countries during the COVID-19 pandemic". In: *Harvard Kennedy School Misinformation Review* 1.3 (2020). URL: https://misinforeview.hks.harvard.edu/article/ibuprofen-narratives-in-five-european-countries-during-the-covid-19-pandemic/ (cited on page 79).

[94] Houping Xiao and Shiyu Wang. "A Joint Maximum Likelihood Estimation Framework for Truth Discovery: A Unified Perspective". In: *IEEE Transactions on Knowledge and Data Engineering* (2015), pp. 1–1. DOI: 10.1109/TKDE.2022.3173911 (cited on page 7).

[95] 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 3).

[96] 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 page 8).

[97] 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 5).

[98] 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 38, 76).

[99] 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 115, 155).

[100] 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 3, 5, 6, 8–10).

[101] 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 114).

[102] 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 page 5).

[103] 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 3, 5, 8).

[104] 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 5, 8).

[105] 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 11, 16).