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 referencesbeneath each citation?]

- Joseph Singleton and Richard Booth. "An Axiomatic Approach to Truth Discovery". In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '20. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2020, pp. 2011–2013. ISBN: 9781450375184
- Joseph Singleton and Richard Booth. "Rankings for Bipartite Tournaments via Chain Editing". In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '21. Virtual Event, United Kingdom: International Foundation for Autonomous Agents and Multiagent Systems, 2021, pp. 1236–1244. ISBN: 9781450383073
- Joseph Singleton. "A Logic of Expertise". In: *ESSLLI 2021 Student Session* (2021). URL: https://arxiv.org/abs/2107.10832
- Joseph Singleton and Richard Booth. *Who's the Expert? On Multi-source Belief Change*. 2022. DOI: 10.48550/ARXIV.2205.00077. URL: https://arxiv.org/abs/2205.00077

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 [37, 54, 47, 16, 56], in general truth discovery methods also handle continuous values [29, 51].

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.

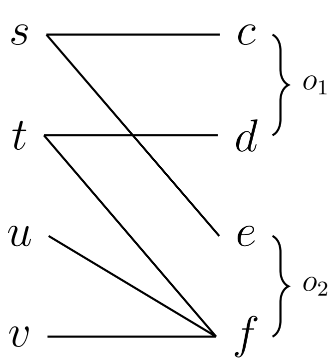


Figure 2.1: Illustrative example of a truth discovery problem, with sources s, t, u, v, object o_1 with associated claims c and d, and o_2 with claims e and f.

- $D = \{D_o\}_{o \in O}$ *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 $\text{obj}(c) = o$ 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* [18] and, specifically (since sources do not necessarily provide a claim for each object) to the setting of *binary aggregation with abstentions* [9, 15]. 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., [9]). 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 [29, 13, 53], whereas others give an assessment of the believability (or confidence, probability etc…) of *each claim* (o, v) [54, 37, 21, 57, 55, 56]. 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 [1], 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

$$
\begin{aligned}
s \sqsubseteq_N^T s' &\iff T_N(s) \leq T_N(s'), \\
c \preceq_N^T c' &\iff T_N(c) \leq T_N(c').
\end{aligned}
$$

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* [37], *TruthFinder* [54] and *CRH* [29]. 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*

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

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*

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

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*

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

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.

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.**]**

[1]This is often called *majority voting* in the truth discovery literature (e.g. [28, 50, 29]), but using the terminology of social choice theory it is better described as *plurality voting*.

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 [37, 54, 52, 17, 56, 29, 21, 57]. 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 [37] is a simple and well-known operator adapted from the *Hubs and Authorities* [27] 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 $T_N^{\text{sums}}(s) = 0$, $T_N^{\text{sums}}(t) = 1$ and $T_N^{\text{sums}}(u) = T_N^{\text{sums}}(v) = \sqrt{2}/2 \approx 0.7071$, giving a source ranking $s \sqsubset u \simeq v \sqsubset t$. For claims, we have $T_N^{\text{sums}}(c) = T_N^{\text{sums}}(e) = 0$, $T_N^{\text{sums}}(d) = \sqrt{2} - 1 \approx 0.4142$ 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 [54] 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 $\text{imp}(c \to c') \in [-1, 1]$, 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 $\text{imp}(c \to c')$ to be $-\lambda$ if c and c' have the same object and 0 otherwise, for some fixed parameter $0 \le \lambda \le 1$.

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 $\text{cl}_N(s) \neq \emptyset$, $T^0 \equiv 0.9$, and $U(T) = T'$, with*

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

$$T'_N(s) = \sum_{c \in \text{cl}_N(s)} \frac{T'_N(c)}{|\,\text{cl}_N(s)|}. \tag{2.2}$$

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

We refer the reader to the original TruthFinder paper [54] 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 [54] 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 [54]) and $\lambda = 0.5$. Assuming that TruthFinder does indeed converge on this network – as it appears to do empirically – we have $T_N^{\text{tf}}(s) \approx 0.5067$, $T_n^{\text{tf}}(t) \approx 0.6590$ and $T_N^{\text{tf}}(u) = T_N^{\text{tf}}(v) = 0.7510$, which gives the ranking $s \sqsubset t \sqsubset u \simeq v$ on the sources. We have $T_N^{\text{tf}}(c) \approx 0.5328$, $T_N^{\text{tf}}(d) \approx 0.5670$, $T_N^{\text{tf}}(e) \approx 0.4807$ and $T_N^{\text{tf}}(f) \approx 0.7510$, 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 [29]. 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 [29, §2.4.1], where each categorical value is represented as a one-hot vector, and the source weight regularisation from [29, 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 $0 \le T_N(c) \le 1$ 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]

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 $\sum_{c \in \operatorname{cl}_N(o)} T'_N(c) = 1$. 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 $T^{\text{crh-}\varepsilon}_N(s) \approx 0.2577$, $T^{\text{crh-}\varepsilon}_N(t) \approx 1.4827$ and $T^{\text{crh-}\varepsilon}_N(u) = T^{\text{crh-}\varepsilon}_N(v) \approx 9.3567$, 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 $T^{\text{crh-}\varepsilon}_N(c) = T^{\text{crh-}\varepsilon}_N(e) \approx 0.0126$, $T^{\text{crh-}\varepsilon}_N(d) \approx 0.0725$ and $T^{\text{crh-}\varepsilon}_N(f) \approx 0.9874$, 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] 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	$s \sqsubset t \sqsubset u \simeq v$	$c \approx e \prec d \prec f$

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 [58] and ranking systems [1]), 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 [1] 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 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 $\text{src}_N(c)$ is less trustworthy (with respect to $\sqsubseteq$) than $\text{src}_N(d)$. 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.

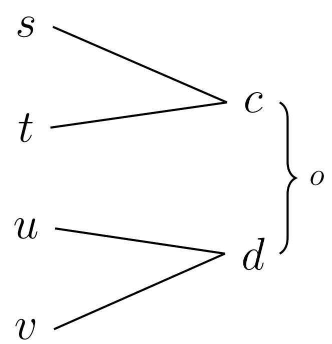


Figure 2.2: A network illustrating ***Claim-coherence****.*

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 $f : A \xrightarrow{\leq} B$. 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 $f : A \xrightarrow{\leq} B$ and $g : B \xrightarrow{\leq} C$, then $g \circ f : A \xrightarrow{\leq} C$). 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 $f : A \xrightarrow{\leq} B$ 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 $g : B \xrightarrow{\leq} A$. 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 $f : A \xrightarrow{\leq} B$. 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. $\square$

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 $\sqsubseteq_N^T$, then $c \prec_N^T c'$.
>
> **Source-coherence**. If $\mathrm{cl}_N(s)$ strictly precedes $\mathrm{cl}_N(s')$ pairwise with respect to $\preceq_N^T$, then $s \sqsubset_N^T s'$.

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.]

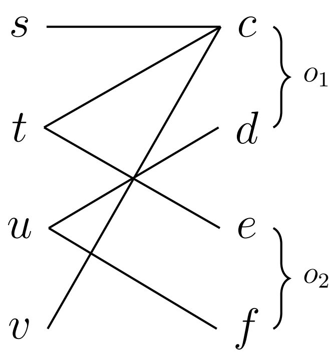


Figure 2.3: A network isomorphic to the one shown in Fig. 2.1.

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 $c \in C$, $\text{obj}(F(c)) = F(\text{obj}(c))$.*

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 $F(\text{src}_N(c)) = \text{src}_{N'}(F(c))$ and $F(\text{cl}_N(s)) = \text{cl}_{N'}(F(s))$. 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 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 [58]. A naive version of positive responsiveness for

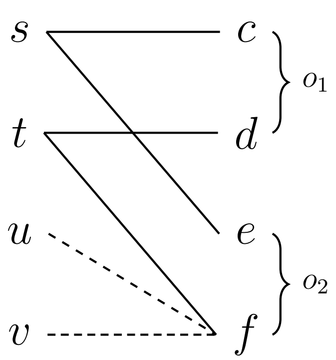


*Figure 2.4: Networks N_0 (solid edges only), $N_1 = N_0 + (u, f)$ and $N_2 = N_1 + (v, f)$ illustrating **Fresh-pos-resp** and **Source-pos-resp**.*

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 $d \in C \setminus \{c\}$, $d \preceq_N^T c$ implies $d \prec_{N+(s,c)}^T c$.

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 $s \sqsubseteq_N^T t$ implies $s \sqsubset_{N+(u,c)}^T t$.

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 **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*

[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.

*$s \in \mathrm{src}_{N_0}(e) \subseteq \mathrm{antisrc}_{N_0}(f)$ 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* [33], which characterises the majority voting rule in two-candidate elections. To prove it, we need a preliminary result.

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

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

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

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 $\mathrm{obj}(F(e)) = F(\mathrm{obj}(e))$ for all claims e since $F(o) = o$ and $\mathrm{obj}(c) = \mathrm{obj}(d)$. 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 $s \in \mathrm{src}_N(c)$. Then $F(s) = \hat{\varphi}(s) \in \mathrm{src}_N(d)$, so $(F(s),d) \in R$. By assumption we have $|\,\mathrm{cl}_N(s)| = 1$, 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 $s \in \mathrm{src}_N(d)$ follows by a near-identical argument. Finally, if $s \notin \mathrm{src}_N(c) \cup \mathrm{src}_N(d)$ 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 $c, d \in \mathrm{cl}_N(o)$,

$$|\,\mathrm{src}_N(c)| \leq |\,\mathrm{src}_N(d)| \implies c \preceq^T_N d \tag{2.4}$$

$$|\,\mathrm{src}_N(c)| < |\,\mathrm{src}_N(d)| \implies c \prec^T_N d. \tag{2.5}$$

First we show (2.4). Suppose $|\,\mathrm{src}_N(c)| \leq |\,\mathrm{src}_N(d)|$. Assume without loss of generality that $c \neq d$. Write $k = |\,\mathrm{src}_N(d)| - |\,\mathrm{src}_N(c)| \geq 0$. Let $X = \{s_1, \ldots, s_k\}$ be an arbitrary subset of $\mathrm{src}_N(d)$ 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 $c \preceq_N^T d$. 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 $d \preceq_N^T c$.*

These properties express something about the trustworthiness of sources via the *claim* ranking $\preceq_N^T$, 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 $s \sqsubseteq_N^T t$.

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, $c \preceq_N^T d$.
> 2. If each $s \in Z$ is marginally untrustworthy, $d \preceq_N^T c$.

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 $\preceq_N^T$, 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 [2], 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 [18].

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 [9].

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

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

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

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

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.*

[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.

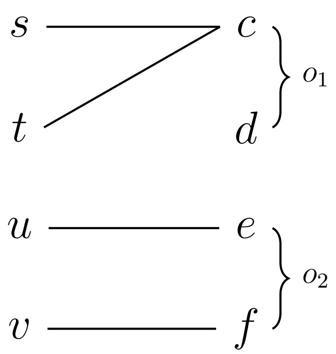


Figure 2.5: A network illustrating ***Disjoint-independence****.*

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 $c \approx_N^T d$ and $e \prec_N^T f$. Considering sources s and t, **Source-coherence** gives $s \sqsubset_N^T t$. But now **Claim-coherence** gives $c \prec_N^T d$: contradiction.

$\square$

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]

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.

[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.

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 $\mathrm{conflict}_N(c)$ strictly precedes $\mathrm{conflict}_N(c')$ pairwise with respect to $\preceq^T_N$, then $c' \prec^T_N c$.

Anti-coherence. If $\mathrm{antisrc}_N(c)$ strictly precedes $\mathrm{antisrc}_N(c')$ 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.*

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**

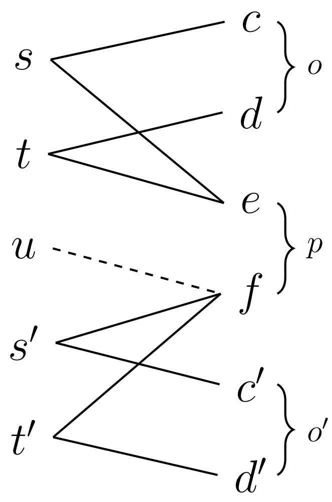


Figure 2.6: Network used to illustrate the impossibility results of Theorem 2.3.2.

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** together 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.

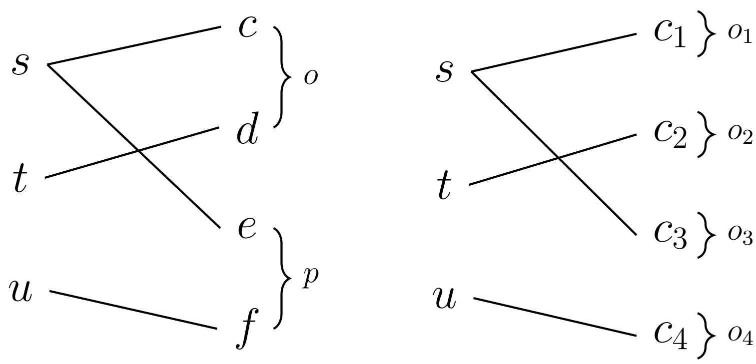
3. Suppose T additionally satisfies **Source-pos-resp** and **Anti-coherence**. Again, $s \sqsubset^T_{N_1} s'$. **Claim-coherence** implies $c \prec^T_{N_1} c'$. Since $\text{antisrc}_{N_1}(d) = \{s\}$ and $\text{antisrc}_{N_1}(d') = \{s'\}$, **Anti-coherence** gives $d' \prec^T_{N_1} d$. But recall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



(a) N *(b) N′*

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

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 $s, s' \in S$, $s \simeq_N^T s'$.

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*

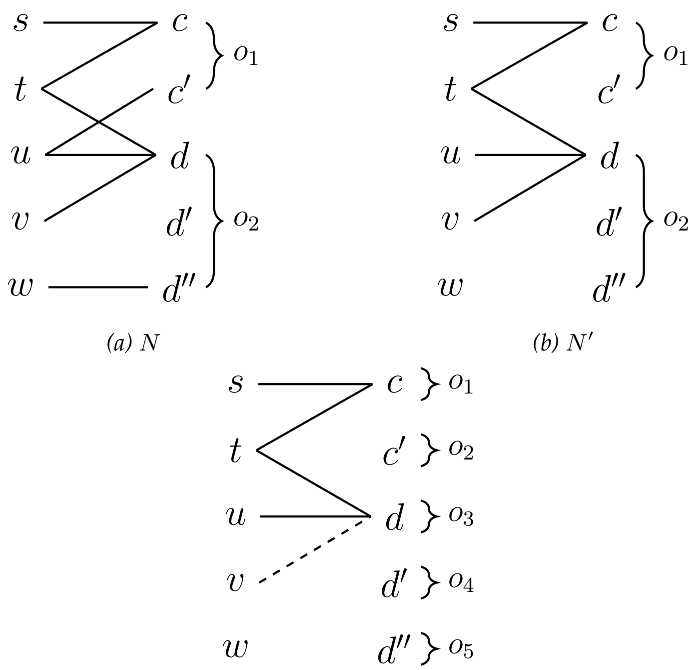
1. *$S' = S$.*
2. *f is a bijection $C_N \to C_{N'}$ such that $(s, c) \in R$ iff $(s, f(c)) \in R'$.*
3. *For al $o \in O'$, $|D'_o| = 1$.*

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.

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.



(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.*

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 $|D''_{\operatorname{obj}(c)}| = |D''_{\operatorname{obj}(d)}| = 1$ ensures we are able to choose an automorphism on N_0

[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			

which swaps c and d (and swaps $\mathrm{src}_{N_0}(c) \setminus \mathrm{src}_{N_0}(d)$ with $\mathrm{src}_{N_0}(d) \setminus \mathrm{src}_{N_0}(c)$). By **Symmetry**, $c \approx^T_{N_0} d$.

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 $c \prec^T_{N''} d$. 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.

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 $T^{w^{\text{agg}}}$. **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 $\text{src}_N(c) = \{F^{-1}(s') \mid s' \in \text{src}_{N'}(F(c))\}$ and $\text{cl}_N(s) = \{F^{-1}(c') \mid c' \in \text{cl}_{N'}(F(s))\}$. 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* $\text{cl}_N(u) = \emptyset$ *and let* c *be a claim. Then for all* $s \neq u$ *with* $\text{cl}_N(s) \neq \emptyset$*,*

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

Proof. First, note that for any claim d,

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

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

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

□

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

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

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

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

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

where we use the fact that $s \in \mathrm{antisrc}_N(c)$ means $c \notin \mathrm{cl}_N(s)$. 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 $\mathrm{cl}_{N \sqcup N'}(s) = \mathrm{cl}_N(s)$ and $\mathrm{src}_{N \sqcup N'}(c) = \mathrm{src}_N(c)$.

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.

[7] While Pasternack and Roth [37] do not consider convergence, Sums is an adaptation of the *Hubs and Authorities* algorithm, for which Kleinberg [27] proves convergence: phrased in our terminology,

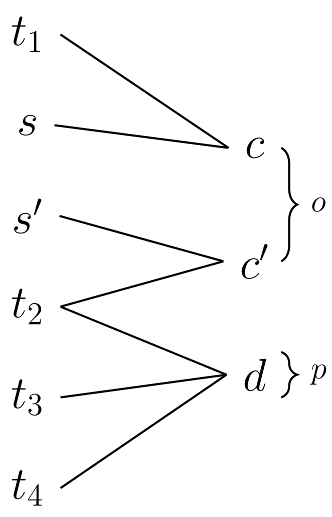


Figure 2.9: Counterexample for ***Source-coherence*** *for Weighted Agreement.*

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

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

$$\lim_{n\to\infty} \max_i |x_n^i| = \max_i |\lim_{n\to\infty} x_n^i|.$$

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

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 $M_{sc} = \mathbb{1}[s \in \mathrm{src}_N(c)]$. Similarly, claim scores converge to a unit eigenvector of $M^{\mathrm{T}}M$. **[TODO:** possibly signpost that we also take linear algebra approach for unbounded sums?]

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 $T^{\text{sums}}_N \equiv 0$ and $U(T^{\text{sums}})_N = T^{\text{sums}}_N$. 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

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

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 $T^{\text{sums}}_N(c) = 0$ for all c, so there is some c_1 with $T^{\text{sums}}_N(c_1) > 0$. Furthermore, $\text{src}_n(c_1) \neq \emptyset$ (otherwise one can easily show $T^{\text{sums}}_N(c_1) = 0$). Likewise, there is some s_1 such that $T^{\text{sums}}_N(s_1) > 0$. Now using the fact that $T^n_N(c) \to T^{\text{sums}}_N(c)$ for each c and taking the limit of the sum, Lemma 2.5.2 gives

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

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

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

as required. One can show $T^{\text{sums}}_N(c) = U(T^{\text{sums}})_N(c)$ for any claim c by a near-identical argument, and thus $U(T^{\text{sums}})_N = T^{\text{sums}}_N$. 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

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

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^{sums}_N 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 $T_N^{\text{sums}}(c) \le T_N^{\text{sums}}(f(c))$ 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

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

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

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

By assumption $T_N^{\text{sums}}(c) - T_N^{\text{sums}}(f(c)) \le 0$ for each c, and the inequality is strict for $c = c_0$. Hence $T_N^{\text{sums}}(s) < T_N^{\text{sums}}(s')$, and $s \sqsubset_N^{T^{\text{sums}}} s'$ 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 $s \simeq t \simeq u \simeq v \sqsubset x_1 \simeq x_2 \simeq x_3 \simeq x_4$

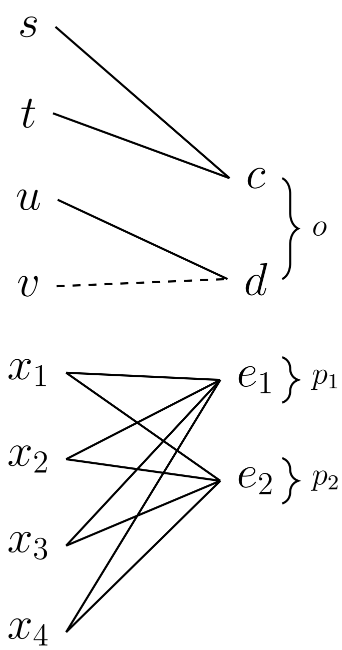


Figure 2.10: Networks used as counterexamples for Sums axiom failures.

and $c \approx d \prec e_1 \approx e_2$. This violates **Fresh-pos-resp**, since $c \preceq^{T^{\text{sums}}}_{N_0} d$ but $c \not\prec^{T^{\text{sums}}}_{N_0+(v,d)} d$. It also violates **Source-pos-resp**, since $s \in \text{antisrc}_{N_0}(d)$, $u \in \text{src}_n(d)$ and $s \sqsubseteq^{T^{\text{sums}}}_{N_0} u$, but $s \not\sqsubset^{T^{\text{sums}}}_{N_0+(v,d)} u$.

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 $c \approx^{T^{\text{sums}}}_{N_1 \sqcup N_2} d$. However, it is straightforward to check that in the network N_1 alone we have $d \prec^{T^{\text{sums}}}_{N_1} c$; 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 $d \prec^{T^{\text{sums}}}_{N'_0} c$ and $c \approx^{T^{\text{sums}}}_{N_0} d$. 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 $T^{\text{sums}}_N(s) = T^{\text{sums}}_N(t) = 0$, $T^{\text{sums}}_N(u) = \sqrt{3} - 1 \approx 0.7321$, $T^{\text{sums}}_N(s') = T^{\text{sums}}_N(t') = 1$ and $T^{\text{sums}}_N(c) = T^{\text{sums}}_N(d) = T^{\text{sums}}_N(e) = 0$, $T^{\text{sums}}_N(f) = 1$, $T^{\text{sums}}_N(c') = T^{\text{sums}}_N(d') = \frac{1}{2}(\sqrt{3} - 1) \approx 0.3660$, 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 $c' \not\prec^{T^{\text{sums}}}_N c$. It also violates **Anti-coherence**, since $\text{antisrc}_N(c) = \{t\}$ strictly precedes $\text{antisrc}_N(c') = \{t'\}$ but $c' \not\prec^{T^{\text{sums}}}_N c$. $\square$

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 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 [31, 12]. There have also been high-profile instances of experts going beyond their area of expertise to comment on issues of public health [49], highlighting the importance of *domain-specific* notions of expertise. Identifying experts is also an important task for *liquid democracy* [6], 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 [8, 19], sociology [10], and philosophy [26, 48, 22], among other fields. In this work we study the *logical* content of expertise, and its relation to truthfulness of information.

Specifically, we generalise the *modal logic* setting of Singleton [40]. The two core notions of the framework are *expertise* and *soundness of information*. Intuitively, a source has expertise on φ if they are able to correctly refute φ in any situation where it is false.[1] Thus, our notion of expertise *does not depend on the "actual" state of affairs*, but only on the source's epistemic state.

It is *sound* for a source to report φ if φ is true *up to lack of expertise*: if φ is logically weakened to a proposition ψ on which the source has expertise, then ψ must be true. That is, the consequences of φ on which the source has expertise are true. This formalises the idea of "filtering out" parts of a statement within a source's expertise. For example, suppose $\varphi = p \wedge q$, and the source has expertise on p but not q. Supposing p is true but q is false, φ is false. However, if we discard information by ignoring q (on which the source has no expertise), we obtain the weaker formula 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

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

have expertise on any statement strictly stronger than p), then $p \wedge q$ is *false* but *sound* for the source to report.

In terms of refutation, φ is sound if the source cannot refute $\neg\varphi$. That is, either φ is in fact true, or the source does not possess sufficient expertise to rule out φ.

This informal picture of expertise already suggests a close connection between expertise, soundness and *knowledge*. Indeed, we will see that, under certain conditions, expertise can be equivalently interpreted in terms of *S4 or S5 knowledge*, familiar from epistemic logic.

Beyond the individual expertise of a single source, one can also consider the *collective expertise* of a group. For example, 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* [20]. Just as individual expertise (and soundness) can be expressed in terms of knowledge, we will see that collective expertise can be expressed in terms of collective knowledge.

[TODO: mention dynamic stuff.]

Contributions. On the conceptual side, we extend the modal framework of expertise of Singleton [40] to reason about the expertise of sources and soundness of information. We generalise this framework by working with a more general semantics, introducing collective expertise among multiple sources, and considering how expertise may evolve via learning and announcements. On the technical side we obtain axiomatisations for the more general semantics, and axiomatise several new sub-classes of models with additional axioms.

Chapter outline. In Section 3.1 we give a motivating example and define the syntax and semantics. Section 3.2 looks at how expertise may be closed under certain operations (e.g. conjunction, negation). The core connection with epistemic logic is given in Section 3.3. We turn to axiomatics in Section 3.4, and give sound and complete logics for various classes of expertise models. In Section 3.5 we generalise to multiple sources. Section 3.6 introduces the dynamic extension of the logic, and we conclude in Section 3.7. Several of the main proofs have also been formalised with the Lean theorem prover.[3] **[TODO:** reference appendix instead.]

3.1 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.

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

[3] `https://github.com/anonymous-logician/expertise`

Example 3.1.1. *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 [23] 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 3.1.1. *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.

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 $M, x \models \varphi$.

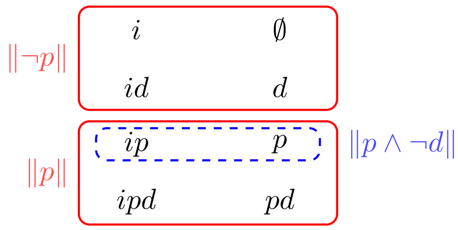


Figure 3.1: Expertise model from Example 3.1.2, which formalises Example 3.1.1.

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* [39, 34, 36], 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 3.1.2. *To formalise Example 3.1.1, consider the model $M = (X, P, V)$ shown in Fig. 3.1, 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 3.1.1.*

We further illustrate the semantics by listing some valid formulas.

Proposition 3.1.1. *The following formulas are valid:*

1. $\varphi \to \mathsf{S}\varphi$
2. $\mathsf{E}\varphi \leftrightarrow \mathsf{AE}\varphi$
3. $\mathsf{A}(\varphi \to \psi) \to (\mathsf{S}\varphi \wedge \mathsf{E}\psi \to \psi)$

[4] 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 3.2.

[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 $\|\varphi\|$.

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.

3.2 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 3.2.1. *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 3.2.1 (1) and Proposition 3.1.1 (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 3.2.1. *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 3.2.2. *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 $x \notin B_2$.

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 $M, x \models \hat{\mathsf{S}}\varphi$.

[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 3.2.1. 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.

Then $M, x \not\models \mathsf{S}\neg(\varphi \to \psi)$ and $M, x \not\models \mathsf{S}\neg\varphi$. Hence there is $A \in P$ such that $\|\neg(\varphi \to \psi)\|_M \subseteq A$ but $x \notin A$, and $B \in P$ such that $\|\neg\varphi\|_M \subseteq B$ but $x \notin B$. Note

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

Since $x \notin A \cup B$ and $A \cup B \in P$ by closure under finite unions, this shows $M, x \not\models \mathsf{S}\neg\psi$, i.e. $M, x \models \hat{\mathsf{S}}\psi$. This completes the proof of $F \models \hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi)$. $\square$

3.3 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 3.3.1. *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 [4, 35].[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 3.3.2. *For a set X and $P \subseteq 2^X$, let R_P be the binary relation on X given by*

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

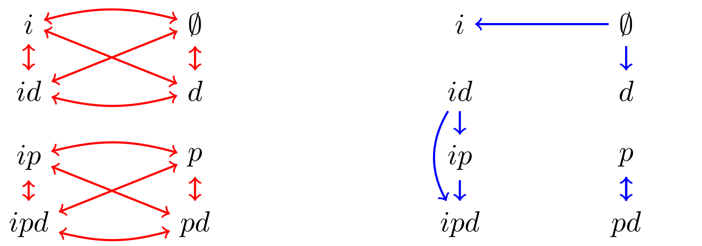


Figure 3.2: Left: the relation R_P corresponding to X and P from Example 3.1.2 (with reflexive edges omitted). Note that R_P is an equivalence relation, with equivalence classes $\|p\|$ and $\|\neg p\|$. Right: an example of a non-symmetric relation R_P, corresponding to $P = \{\emptyset, X, \{id, ip, ipd\}, \{id, ip\}, \{id\}, \{i, \emptyset\}, \{\emptyset, d\}, \{p, pd\}\}$.

In the case where P is the collection of closed sets of a topology on X, R_P is the *specialisation preorder*. Fig. 3.2 shows an example of R_P for X and P from Example 3.1.2. 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 3.3.1. *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 3.3.2. *Let X be a set.*

1. *For any $P \subseteq 2^X$, R_P is reflexive and transitive.*
2. *If $P \subseteq 2^X$ is closed under unions and intersections, then for all $A \subseteq X$:*

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

3. *If R is a reflexive and transitive relation on X, there is $P \subseteq 2^X$ closed under unions and intersections such that $R_P = R$.*

[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* [35, pp. 24].

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 3.3.1, 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 3.3.1. *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 3.3.2 (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 3.3.2 (2), since P is fully determined by R_P for expertise models closed under unions and intersections. Finally, Lemma 3.3.2 (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 3.3.2. *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 3.3.1 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 3.3.2 (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 3.3.1. 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 3.3.2 (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 3.3.3. *Let $f : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}} \to \mathbb{M}^*_{\mathsf{S4}}$ be the bijection from Theorem 3.3.1. Then for all $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{unions}}$, $x \in X$ and $\varphi \in \mathcal{L}$:*

$$M, x \models \varphi \iff f(M), x \models t(\varphi) \tag{3.1}$$

Moreover, if $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$ is the bijection from Theorem 3.3.2, then for all $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$:

$$M, x \models \varphi \iff g(M), x \models t(\varphi) \tag{3.2}$$

Proof. Note that since g is defined as the restriction of f to $\mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}}$, (3.2) follows from (3.1). We show (3.1) 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 3.3.2 (2), we have

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

We show (3.1) 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 $f(M), x \models \mathsf{A}(\neg t(\varphi) \rightarrow \mathsf{K}\neg t(\varphi))$, i.e. $f(M), x \models t(\mathsf{E}\varphi)$.

  Now suppose $f(M), x \models t(\mathsf{E}\varphi)$, i.e. $f(M), x \models \mathsf{A}(\neg t(\varphi) \rightarrow \mathsf{K}\neg t(\varphi))$. 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 $\neg t(\varphi) \rightarrow \mathsf{K}\neg t(\varphi)$ 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.

3.4 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 3.2.

The General Case Let L be the extension of propositional logic generated by the axioms and inference rules shown in Table 3.1. Note that we treat A as a "box" and S as a "diamond" modality. Some of the axioms were already seen in Proposition 3.1.1; 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 3.4.1. L *is sound with respect to* $\mathbb{M}$.

Table 3.1: 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)

Proof. The inference rules are clearly sound. All axioms were either shown to be sound in Proposition 3.1.1 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 3.4.2. *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 3.4.3 (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 3.4.4. *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 3.4.5. *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 ($\mathrm{Nec}_{\mathsf{A}}$), $\vdash \mathsf{A}(\psi_0 \to \cdots \to \psi_n \to \varphi)$. 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 3.4.3), 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 3.4.1. *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 3.4.5 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 3.4.5. 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 3.4.6. *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 3.4.5 and Corollary 3.4.1. We treat the cases of E and S formulas.

- E: First suppose $\mathsf{E}\varphi \in \Gamma$. By Corollary 3.4.1, $\mathsf{E}\varphi \in \Sigma$. 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 3.4.5 then gives $\mathsf{A}(\varphi \leftrightarrow \psi) \in \Sigma$. By Corollary 3.4.1, 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 3.4.5, $\mathsf{A}(\varphi \to \psi) \in \Sigma$. By Corollary 3.4.1 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 ($\mathsf{T}_\mathsf{S}$) and ($\mathsf{Nec}_\mathsf{A}$), we have $\mathsf{A}(\varphi \to \mathsf{S}\varphi) \in \Sigma$, and consequently $|\varphi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$ by Lemma 3.4.5. 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 3.4.5, $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Sigma$.

  Now, take any $(\Lambda, u) \in |\mathsf{S}\psi|_\Sigma \times \mathbb{R}$. Since $\Lambda \in X_\Sigma$, Corollary 3.4.1 gives $\mathsf{A}(\psi \to \mathsf{S}\varphi) \in \Lambda$. By ($\mathsf{W}_\mathsf{S}$), $\mathsf{S}\psi \to \mathsf{S}\mathsf{S}\varphi \in \Lambda$. Since $\Lambda \in |\mathsf{S}\psi|_\Sigma$, we get $\mathsf{S}\mathsf{S}\varphi \in \Lambda$. But then ($4_\mathsf{S}$) gives $\mathsf{S}\varphi \in \Lambda$. That is, $(\Lambda, u) \in |\mathsf{S}\varphi|_\Sigma \times \mathbb{R} \subseteq A$. This shows $|\mathsf{S}\psi|_\Sigma \times \mathbb{R} \subseteq A$, so A is S-closed.

  Finally, we show that for all $\psi \in \mathcal{L}$, $A \neq |\psi|_\Sigma \times \mathbb{R}$. For contradiction, suppose there is ψ with $A = |\psi|_\Sigma \times \mathbb{R}$. Then since $(\Gamma, s) \in A$, we have $\Gamma \in |\psi|_\Sigma$. But then $(\Gamma, t) \in |\psi|_\Sigma \times \mathbb{R} = A$: contradiction.

  This completes the proof that $A \in \widehat{P}^1_\Sigma$. Thus $\widehat{M}_\Sigma, (\Gamma, t) \not\models \mathsf{S}\varphi$, and we are done.

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

□

Theorem 3.4.1. 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 3.4.6 (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 subclasses of $\mathbb{M}$ corresponding to closure conditions.

To start, consider closure under intersections. It was shown in Proposition 3.2.1 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 3.4.2. $\mathsf{L}_{\mathsf{int}}$ *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 3.2.1 (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 3.4.5 and Corollary 3.4.1.

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

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

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 $\varphi \in \mathcal{L}$,

$$M_\Sigma, \Gamma \models \varphi \iff \varphi \in \Gamma,$$

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

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 3.4.6). 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), $\mathsf{S}\psi \to \mathsf{S}\mathsf{S}\varphi \in \Delta$. Now $\mathsf{S}\psi \in \Delta$ implies $\mathsf{S}\mathsf{S}\varphi \in \Delta$, and ($4_\mathsf{S}$) gives $\mathsf{S}\varphi \in \Delta$, i.e. $\Delta \in |\mathsf{S}\varphi|_\Sigma$. This shows $|\mathsf{S}\psi|_\Sigma \subseteq |\mathsf{S}\varphi|_\Sigma$, and thus $|\mathsf{S}\varphi|_\Sigma$ is S-closed.

  Hence $|\mathsf{S}\varphi|_\Sigma$ is a set in P_Σ not containing Γ. Moreover, $\|\varphi\|_{M_\Sigma} \subseteq |\mathsf{S}\varphi|_\Sigma$ by the induction hypothesis and (T_S). Hence $M_\Sigma, \Gamma \not\models \mathsf{S}\varphi$.

Strong completeness now follows. If $\Gamma \not\vdash_{\mathsf{L_{int}}} \varphi$, then $\Gamma \cup \{\neg\varphi\}$ is consistent, so by Lindenbaum's Lemma there is $\Sigma \in X_{\mathsf{L_{int}}}$ 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 3.2.2 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 $\mathbb{M}_{\mathsf{top}} = \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{finite-unions}}$ 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:

$$\hat{\mathsf{S}}(\varphi \to \psi) \to (\hat{\mathsf{S}}\varphi \to \hat{\mathsf{S}}\psi) \qquad (\mathrm{K}_\mathsf{S})$$

$$\mathsf{A}\varphi \to \hat{\mathsf{S}}\varphi \qquad (\mathrm{Inc})$$

Let $\mathsf{L_{top}}$ be the extension of $\mathsf{L_{int}}$ by (K_S) and (Inc). Note that $\mathsf{L_{top}}$ contains the **KT4** axioms for $\hat{\mathsf{S}}$ (recalling that (T_S) and (4_S) are the "diamond" versions of **T** and **4**).

[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.

Since **KT4** together with the bridge axiom (Inc) is complete for the class of relational models $\mathbb{M}^*_{\mathsf{S4}}$, we can exploit Theorem 3.3.3 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 3.4.3. $\mathsf{L}_{\mathsf{top}}$ *is sound and strongly complete with respect to* $\mathbb{M}_{\mathsf{top}}$.

Proof. Soundness of (K_{S}) for $\mathbb{M}_{\mathsf{top}}$ follows from Proposition 3.2.2. 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 3.3.3. 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 3.2. It is well known that $\mathsf{L}_{\mathsf{S4A}}$ is strongly complete with respect to $\mathbb{M}^*_{\mathsf{S4}}$ [5, Theorem 7.2].

Table 3.2: Axioms and inference rules for $\mathsf{L}_{\mathsf{S4A}}$.

$\mathsf{K}(\varphi \to \psi) \to (\mathsf{K}\varphi \to \mathsf{K}\psi)$	(K_{K})
$\mathsf{K}\varphi \to \varphi$	(T_{K})
$\mathsf{K}\varphi \to \mathsf{K}\mathsf{K}\varphi$	(4_{K})
$\mathsf{A}(\varphi \to \psi) \to (\mathsf{A}\varphi \to \mathsf{A}\psi)$	(K_{A})
$\mathsf{A}\varphi \to \varphi$	(T_{A})
$\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi$	(5_{A})
$\mathsf{A}\varphi \to \mathsf{K}\varphi$	($\mathrm{Inc}_{\mathsf{K}}$)
From φ infer $\mathsf{A}\varphi$	($\mathrm{Nec}_{\mathsf{A}}$)
From $\varphi \to \psi$ and φ infer ψ	(MP)

Now, define a translation $u : \mathcal{L}_{\mathsf{KA}} \to \mathcal{L}$ as follows:

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

Recall the translation $t : \mathcal{L} \to \mathcal{L}_{\mathsf{KA}}$ from Section 3.3. 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 $u(t(\varphi))$.

Claim 3.4.1. *Let* $\varphi \in \mathcal{L}$. *Then* $\vdash_{\mathsf{L}_{\mathsf{top}}} \varphi \leftrightarrow u(t(\varphi))$.

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 $\mathsf{L}_{\mathsf{top}}$) for S, E and A:

$$\text{From } \varphi \leftrightarrow \psi \text{ infer } \bigcirc\varphi \leftrightarrow \bigcirc\psi \quad (\bigcirc \in \{\mathsf{S}, \mathsf{E}, \mathsf{A}\}).$$

[11] Note that **KT4** is also complete for topological spaces with respect to the interior semantics [4].

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 3.4.2. *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 $\mathsf{L_{top}}$-provably equivalent to $\neg u(\varphi) \to \mathsf{S}\neg u(\varphi)$, which is an instance of (T_{S}).

- (4_{K}): We have
$$u(\mathsf{K}\varphi \to \mathsf{KK}\varphi) = \neg\mathsf{S}\neg u(\varphi) \to \neg\mathsf{S}\neg\neg\mathsf{S}\neg u(\varphi)$$
This is provably equivalent to $\mathsf{SS}\neg u(\varphi) \to \mathsf{S}\neg u(\varphi)$, which is an instance of (4_{S}).

- $(\mathrm{K}_{\mathsf{A}})$: 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 $(\mathrm{K}_{\mathsf{A}})$ in $\mathsf{L}_{\mathsf{top}}$.

- $(\mathrm{T}_{\mathsf{A}})$: We have
$$u(\mathsf{A}\varphi \to \varphi) = \mathsf{A}u(\varphi) \to u(\varphi)$$
which is an instance of $(\mathrm{T}_{\mathsf{A}})$ in $\mathsf{L}_{\mathsf{top}}$.

- (5_{A}): We have
$$u(\neg\mathsf{A}\varphi \to \mathsf{A}\neg\mathsf{A}\varphi) = \neg\mathsf{A}u(\varphi) \to \mathsf{A}\neg\mathsf{A}u(\varphi)$$
which is an instance of (5_{A}) in $\mathsf{L}_{\mathsf{top}}$.

- $(\mathrm{Inc}_{\mathsf{K}})$: We have
$$u(\mathsf{A}\varphi \to \mathsf{K}\varphi) = \mathsf{A}u(\varphi) \to \neg\mathsf{S}\neg u(\varphi) = \mathsf{A}u(\varphi) \to \hat{\mathsf{S}}u(\varphi)$$
which is an instance of (Inc).

For the inductive step, we show that for each inference rule $\frac{\psi_1,\ldots,\psi_n}{\varphi}$, if $\vdash_{\mathsf{L}_{\mathsf{top}}} u(\psi_i)$ for each i then $\vdash_{\mathsf{L}_{\mathsf{top}}} u(\varphi)$.

- $(\mathrm{Nec}_{\mathsf{A}})$: If $\vdash_{\mathsf{L}_{\mathsf{top}}} u(\varphi)$, then from $(\mathrm{Nec}_{\mathsf{A}})$ in $\mathsf{L}_{\mathsf{top}}$ 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 $\mathsf{L}_{\mathsf{top}}$ and the fact that $u(\varphi \to \psi) = u(\varphi) \to u(\psi)$.

$\square$

Claims 3.4.1 and 3.4.2 easily imply the following.

Claim 3.4.3. *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 3.4.2, $\vdash_{\mathsf{L}_{\mathsf{top}}} u(t(\varphi))$. By Claim 3.4.1, $\vdash_{\mathsf{L}_{\mathsf{top}}} \varphi \leftrightarrow u(t(\varphi))$. By (MP), $\vdash_{\mathsf{L}_{\mathsf{top}}} \varphi$. $\square$

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 3.3.3 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 $\mathsf{L}_{\mathsf{S4A}}$ for $\mathbb{M}^*_{\mathsf{S4}}$, we get $t(\Gamma) \vdash_{\mathsf{L}_{\mathsf{S4A}}} t(\varphi)$. That is, there are $\psi_0, \ldots, \psi_n \in \Gamma$ such that $\vdash_{\mathsf{L}_{\mathsf{S4A}}} t(\psi_0) \wedge \cdots \wedge t(\psi_n) \to t(\varphi)$. Since t passes over conjunctions and implications, this means $\vdash_{\mathsf{L}_{\mathsf{S4A}}} t(\psi_0 \wedge \cdots \wedge \psi_n \to \varphi)$. By Claim 3.4.3, $\vdash_{\mathsf{L}_{\mathsf{top}}} \psi_0 \wedge \cdots \wedge \psi_n \to \varphi$. Hence $\Gamma \vdash_{\mathsf{L}_{\mathsf{top}}} \varphi$, and we are done. $\square$

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 $\mathsf{L}_{\mathsf{int-compl}}$ be the extension of $\mathsf{L}_{\mathsf{top}}$ 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 3.4.4. $\mathsf{L}_{\mathsf{int-compl}}$ *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 3.1.1 (4), we get $M, x \models \neg\mathsf{S}\varphi$.

The completeness proof goes in exactly the same way as Theorem 3.4.3. Letting $\mathsf{L}_{\mathsf{S5A}}$ be the extension of $\mathsf{L}_{\mathsf{S4A}}$ with the (5_{K}) axiom $\neg\mathsf{K}\varphi \to \mathsf{K}\neg\mathsf{K}\varphi$, it can be shown that $\mathsf{L}_{\mathsf{S5A}}$ is strongly complete with respect to $\mathbb{M}^*_{\mathsf{S5}}$. With u as in the proof of Theorem 3.4.3, we have that $\vdash_{\mathsf{L}_{\mathsf{S5A}}} \varphi$ implies $\vdash_{\mathsf{L}_{\mathsf{int-compl}}} u(\varphi)$, 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 $\mathsf{L}_{\mathsf{int-compl}}$, but this follows from (5_{S})). The remainder of the proof goes through as before, this time appealing to the bijection $g : \mathbb{M}_{\mathsf{int}} \cap \mathbb{M}_{\mathsf{compl}} \to \mathbb{M}^*_{\mathsf{S5}}$. □

3.5 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 [20], and establish connections between collective expertise and collective knowledge.

3.5.1 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 [20]:

- $\mathsf{K}_j\varphi$ (individual knowledge): for $j \in J$ and a formula φ, set

  $$M^*, x \models \mathsf{K}_j\varphi \iff \forall y \in X : xR_jy \implies M^*, y \models \varphi.$$

  This is the straightforward adaptation of knowledge in the single-source case to the multi-source setting.

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

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

That is, knowledge of φ is distributed among the sources in J if, by combining their accessibility relations R_j, all states possible at x satisfy φ. Here the R_j are combined by taking their intersection: a state y is possible according to the group at x iff *every* source in J considers y possible at x.

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

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

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

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

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

  Here $\mathsf{K}^1_J\varphi$ says that everyone in J knows φ, $\mathsf{K}^2_J\varphi$ says that everybody in J knows that everybody in J knows φ, and so on. There is common knowledge of φ among J if this nesting of "everybody knows" holds for any order n.

In what follows we write $\mathcal{L}^{\mathcal{J}}_{\mathsf{KA}}$ for the language formed from Prop with knowledge operators K_j, $\mathsf{K}^{\mathsf{dist}}_J$, $\mathsf{K}^{\mathsf{sh}}_J$ and $\mathsf{K}^{\mathsf{com}}_J$, for $j \in \mathcal{J}$ and $J \subseteq \mathcal{J}$ non-empty, and the universal modality A.

3.5.2 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 3.3.3) 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.

Distributed Expertise For distributed expertise, the intuition is clear: the sources in a group J should combine their expertise collections P_j to form a larger collection P^{dist}_J. A first candidate for P^{dist}_J would therefore be $\bigcup_{j\in J} P_j$. However, since we assume each P_j is closed under unions and intersections, we suppose that each source j has the cognitive or computational capacity to combine expertise sets $A \in P_j$ by taking unions or intersections. We argue that the same should be possible for the

[12]In Fagin et al. [20], 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.

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 3.3, 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 [44, 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 3.3.3 that our semantics for expertise and soundness is connected to relational semantics via the mapping $P \mapsto R_P$ (Definition 3.3.2). 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 3.5.1. *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 3.3.3. 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 3.5.2. *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. $\square$

Proposition 3.5.2 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. [20]. In what follows, write $R^+ = \bigcup_{n\in\mathbb{N}} R^n$ for the transitive closure of R.

Lemma 3.5.1 (Fagin et al. [20], 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 3.5.1, 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 3.5.3. *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 3.3.2 (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 3.3.1 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 3.3.2 (2), we have

$$\begin{aligned}
A \text{ downwards closed wrt } R_J^{\mathsf{com}} &\iff A \in P_J^{\mathsf{com}}\\
&\iff \forall j \in J : A \in P_j\\
&\iff \forall j \in J : A \text{ downwards closed wrt } R_j\\
&\iff A \text{ downwards closed wrt } \textstyle\bigcup_{j\in J} R_j\\
&\iff A \text{ downwards closed wrt } R'
\end{aligned}$$

where the last step uses the fact that A is downwards closed with respect to some relation if and only if it is downwards closed with respect to the transitive closure. This completes the proof. □

Collective semantics We now formally define the syntax and semantics of collective expertise. Let $\mathcal{L}^{\mathcal{J}}$ be the language defined by the following grammar:

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid \mathsf{E}_j\varphi \mid \mathsf{S}_j\varphi \mid \mathsf{E}_J^g\varphi \mid \mathsf{S}_J^g\varphi \mid \mathsf{A}\varphi$$

for $p \in \mathsf{Prop}$, $j \in \mathcal{J}$, $g \in \{\mathsf{dist}, \mathsf{com}\}$ and $J \subseteq \mathcal{J}$ non-empty. For a multi-source expertise model $M = (X, \{P_j\}_{j\in\mathcal{J}}, V)$, define the satisfaction relation as before for atomic propositions, propositional connectives and A, and set

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

Note that expertise and soundness are interpreted as before, but with respect to different collections P. Consequently, the interactions shown in Proposition 3.1.1 still hold for individual and collective notions of expertise and soundness.

Example 3.5.1. *Extending Examples 3.1.1 and 3.1.2, 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 3.1.2, 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 3.1.1. 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 3.5.4. *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 3.3. 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 3.3.3.

Theorem 3.5.1. *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 3.3.1 and 3.3.2. For the stated property we proceed by induction on $\mathcal{L}^{\mathcal{J}}$ formulas. As in Theorem 3.3.3, the cases for atomic propositions, propositional connectives and A are straightforward. For expertise and soundness, the argument in the proof of Theorem 3.3.3 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 3.5.1. For common expertise this was shown in Lemma 3.5.1 and Proposition 3.5.3. □

Theorem 3.5.1 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 3.4.3 and 3.4.4.

3.6 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.

3.6.1 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 3.2, 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 3.2 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 3.6.1. *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 [3].

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

Proposition 3.6.1. *The following formulas are valid on* $\mathbb{M}$*:*

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

Proof. The cases for atomic propositions, propositional connectives and A are straightforward. We show the reduction axiom for S. Let $M = (X, P, V)$ be a model and $x \in X$.

"$\to$": Suppose $M, x \models [+\varphi]\mathsf{S}\psi$. Then $M^{+\varphi}, x \models \mathsf{S}\psi$. Hence

$$x \in \bigcap\{A \in P^{+\varphi} \mid \|\psi\|_{M^{+\varphi}} \subseteq A\} \quad (*)$$

Note that $\|\psi\|_{M^{+\varphi}} = \|[+\varphi]\psi\|_M$. Now take $A \in P$ such that $\|[+\varphi]\psi\|_M \subseteq A$. Since $P \subseteq P^{+\varphi}$, we get $x \in A$ from $(*)$. Hence $M, x \models \mathsf{S}[+\varphi]\psi$.

Now suppose $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi)$. Then $\|[+\varphi]\psi\|_M \subseteq \|\varphi\|_M$, so $\|\psi\|_{M^{+\varphi}} \subseteq \|\varphi\|_M$. Since $\|\varphi\|_M \in P^{+\varphi}$, we get $x \in \|\varphi\|_M$ from $(*)$, i.e. $M, x \models \varphi$ as required.

"$\leftarrow$": Suppose $M, x \models \mathsf{S}[+\varphi]\psi$ and $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi) \to \varphi$. Take $A \in P^{+\varphi}$ such that $\|\psi\|_{M^{+\varphi}} \subseteq A$. Then $\|[+\varphi]\psi\|_M \subseteq A$. By definition of $P^{+\varphi}$, there is a collection $\mathcal{A} \subseteq P \cup \{\|\varphi\|_M\}$ such that $A = \bigcap\mathcal{A}$. Let $B \in \mathcal{A}$. If $B \in P$, then $\|[+\varphi]\psi\|_M \subseteq A \subseteq B$ and $M, x \models \mathsf{S}[+\varphi]\psi$ give $x \in B$. Otherwise, $B = \|\varphi\|_M$. Hence $\|[+\varphi]\psi\|_M \subseteq \|\varphi\|_M$, so $M, x \models \mathsf{A}([+\varphi]\psi \to \varphi)$. By the second assumption, we get

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

$M, x \models \varphi$, 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.

3.6.2 Sound Announcements

In logics of public announcement [38, 14], 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 3.6.2. *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 3.6.1. *The report of the economist in Example 3.1.1 can be modelled as* $[?(p \wedge \neg d)]$. *Note that, with* M *as in Example 3.1.2,* $\|\mathsf{S}(p \wedge \neg d)\|_M = \|p\|_M$. *The updated model* $M^{?(p\wedge\neg d)}$ *therefore consists only of the bottom half of* M *as shown in Fig. 3.1. We see that* $M, idp \models [?(p \wedge \neg d)]d$ *– showing that even propositional announcements can "fail" due to lack of expertise – and* $M \models [?(p \wedge \neg d)]\mathsf{A}p$ *– showing that the parts of the report on which the sources does have expertise are* always *true after their announcement.*

As with the expertise increase operator, sound announcements remain sound for purely propositional formulas $\alpha \in \mathcal{L}_0$: $[?\alpha]\mathsf{S}\alpha$ is valid on $\mathbb{M}_{\mathsf{int}}$. This is not true for general formulas $\varphi \in \mathcal{L}$ which may refer to expertise itself. For example, in the model $M = (X, P, V) \in \mathbb{M}_{\mathsf{int}}$ given by $X = \{1, 2, 3, 4\}$, $P = \{\emptyset, X, \{1\}, \{2\}, \{1, 2, 3\}\}$, $V(p) = \{1, 2\}$ and $V(q) = \{2, 4\}$, with $\varphi = p \wedge \neg \mathsf{E}q$ we have $M, 1 \not\models [?\varphi]\mathsf{S}\varphi$.

The following reduction axioms allow formulas involving announcements to be expressed in the static language.

Proposition 3.6.2. *The following formulas are valid on* $\mathbb{M}$*:*

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

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

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

Proof. The cases of atomic propositions, propositional connectives and the universal modality A are straightforward.

For the reduction axiom for S, first note that $\|\psi\|_{M^{?\varphi}} = \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M$. We need to show that $M, x \models [?\varphi]\mathsf{S}\psi$ iff $M, x \models \mathsf{S}\varphi \to \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$. If $M, x \not\models \mathsf{S}\varphi$ this is clear. Otherwise $x \in \|\mathsf{S}\varphi\|_M$, and we have

$$\begin{aligned}
M, x \models [?\varphi]\mathsf{S}\psi &\iff M^{?\varphi}, x \models \mathsf{S}\psi \\
&\iff \forall B \in P^{?\varphi} : \|\psi\|_{M^{?\varphi}} \subseteq B \implies x \in B \\
&\iff \forall A \in P : \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq A \cap \|\mathsf{S}\varphi\|_M \implies x \in A \cap \|\mathsf{S}\varphi\|_M \\
&\iff \forall A \in P : \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq A \implies x \in A \\
&\iff M, x \models \mathsf{S}(\mathsf{S}\varphi \wedge [?\varphi]\psi)
\end{aligned}$$

and the result follows.

For the E reduction axiom, take $M \in \mathbb{M}_{\mathsf{int}}$. Again, suppose without loss of generality that $x \in \|\mathsf{S}\varphi\|_M$. Then we have

$$\begin{aligned}
M, x \models [?\varphi]\mathsf{E}\psi &\iff M^{?\varphi}, x \models \mathsf{E}\psi \\
&\iff \|\psi\|_{M^{?\varphi}} \in P^{?\varphi} \\
&\iff \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \in P^{?\varphi} \\
&\iff \|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \in P \\
&\iff M, x \models \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi)
\end{aligned}$$

where the forwards direction of the penultimate equivalence holds since $P^{?\varphi} \subseteq P$ when M is closed under intersections, and the backwards direction holds since $\|\mathsf{S}\varphi \wedge [?\varphi]\psi\|_M \subseteq \|\mathsf{S}\varphi\|_M = X^{?\varphi}$. It follows that $M, x \models [?\varphi]\mathsf{E}\psi$ iff $M, x \models \mathsf{S}\varphi \to \mathsf{E}(\mathsf{S}\varphi \wedge [?\varphi]\psi)$, as required. □

To conclude, we note some interesting validities involving the dynamic operators and their interaction.

Proposition 3.6.3. *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 3.1.1 (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 3.1.1 (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.

3.7 Conclusion

This chapter presented a simple modal logic framework to reason about the expertise of information sources and soundness of information, generalising the framework of Singleton [40]. We investigated both conceptual and technical issues, establishing several completeness for various classes of expertise models. The connection with epistemic logic showed how expertise and soundness may be given precise interpretations in terms of knowledge; if expertise is closed under intersections and unions this results in S4 knowledge, and closure under complements strengthens this to S5. Finally, we extended the framework to handle multiple sources and studied notions of collective expertise.

There are many directions for future work. First, our approach allows one to reason about soundness of information only if the extent of a source's expertise is known up-front. In practical situations it is more likely that one has to *estimate* a source's expertise, e.g. on the basis of previous reports [25, 11]; such approaches could be combined with our logical framework in future work.

Expertise is also not static: it may change over time as sources learn and acquire new evidence. To model this one could introduce *dynamic expertise operators*, as in Dynamic Epistemic Logic. One source of inspiration here is *dynamic evidence logics* [46, 45], which study how evidence (and beliefs formed on the basis of evidence) change over time. Such logics also use neighbourhood semantics to interpret evidence modalities, which is technically (and possibly conceptually) similar to our semantics for expertise.

Finally, there is scope to study the interaction between interaction between expertise and *trust*, which has been extensively studied from a logical perspective [7, 30, 32, 24]. Intuitively, source i should trust j on φ if i believes that j has expertise on φ. "Belief in expertise" in this manner is not particularly meaningful in the current framework, since $\mathsf{E}_j\varphi$ either holds everywhere or nowhere. Future work could extend the semantics to allow the expertise collection P_j to vary between states, so as to model one source's uncertainty about the expertise of another.

Bibliography

[1] 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).

[2] Kenneth J. Arrow. "Social Choice and Individual Values". In: *Ethics* 62.3 (1952), pp. 220–222 (cited on page 19).

[3] 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 64).

[4] Johan van Benthem and Guram Bezhanishvili. "Modal logics of space". In: *Handbook of spatial logics*. Springer, 2007, pp. 217–298 (cited on pages 43, 54).

[5] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal logic*. Vol. 53. Cambridge University Press, 2002 (cited on page 54).

[6] 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 35).

[7] Richard Booth and Aaron Hunter. "Trust as a Precursor to Belief Revision". In: *JAIR* 61 (2018), pp. 699–722 (cited on page 68).

[8] Michelene TH Chi, Robert Glaser, and Marshall J Farr. *The nature of expertise*. Psychology Press, 2014 (cited on page 35).

[9] 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).

[10] Harry Collins and Robert Evans. *Rethinking expertise*. University of Chicago Press, 2008 (cited on page 35).

[11] Mehdi Dastani et al. "Inferring trust". In: *CLIMA*. Springer. 2004, pp. 144–160 (cited on page 68).

[12] 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 35).

[13] 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).

[14] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer Netherlands, 2008. DOI: 10.1007/978-1-4020-5839-4 (cited on page 65).

[15] 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).

[16] 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).

[17] 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).

[18] 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, 19).

[19] K Anders Ericsson and Tyler J Towne. "Expertise." In: *WIREs Cognitive Science* (2010) (cited on page 35).

[20] Ronald Fagin et al. *Reasoning about knowledge*. MIT press, 2003 (cited on pages 36, 57, 58, 60).

[21] 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).

[22] Alvin I Goldman. "Expertise". In: *Topoi* 37.1 (2018), pp. 3–10 (cited on page 35).

[23] 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 37).

[24] Andreas Herzig et al. "A logic of trust and reputation". In: *Logic Journal of the IGPL* 18.1 (2010), pp. 214–244 (cited on page 68).

[25] Aaron Hunter. "Building Trust for Belief Revision". In: *PRICAI 2021: Trends in Artificial Intelligence*. Ed. by Duc Nghia Pham et al. Cham: Springer International Publishing, 2021, pp. 543–555. ISBN: 978-3-030-89188-6 (cited on page 68).

[26] Daniel Kilov. "The brittleness of expertise and why it matters". In: *Synthese* 199.1 (2021), pp. 3431–3455 (cited on page 35).

[27] 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, 28).

[28] 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 page 7).

[29] 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).

[30] Churn-Jung Liau. "Belief, information acquisition, and trust in multi-agent systems – A modal logic formulation". In: *Artificial Intelligence* 149.1 (2003), pp. 31–60 (cited on page 68).

[31] Sue Llewellyn. "Covid-19: how to be careful with trust and expertise on social media". In: *BMJ* 368 (2020) (cited on page 35).

[32] Emiliano Lorini, Guifei Jiang, and Laurent Perrussel. "Trust-Based Belief Change". In: *Proc. ECAI*. 2014, pp. 549–554 (cited on page 68).

[33] 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).

[34] Richard Montague. "Universal grammar". In: *Theoria* 36.3 (1970), pp. 373–398 (cited on page 38).

[35] Aybüke Özgün. "Evidence in Epistemic Logic : A Topological Perspective". Theses. Université de Lorraine, Oct. 2017 (cited on page 43).

[36] Eric Pacuit. *Neighborhood semantics for modal logic*. Springer International Publishing, 2017 (cited on page 38).

[37] 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, 28).

[38] Jan Plaza. "Logics of public communications". In: *Synthese* 158.2 (2007), pp. 165–179 (cited on page 65).

[39] 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 38).

[40] Joseph Singleton. "A Logic of Expertise". In: *ESSLLI 2021 Student Session* (2021). URL: https://arxiv.org/abs/2107.10832 (cited on pages iv, 35, 36, 68).

[41] 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 iv).

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

[43] Joseph Singleton and Richard Booth. *Who's the Expert? On Multi-source Belief Change*. 2022. DOI: 10.48550/ARXIV.2205.00077. URL: https://arxiv.org/abs/2205.00077 (cited on page iv).

[44] 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 59).

[45] Johan van Benthem, David Fernández-Duque, and Eric Pacuit. "Evidence and plausibility in neighborhood structures". In: *Annals of Pure and Applied Logic* 165.1 (2014), pp. 106–133 (cited on page 68).

[46] Johan van Benthem and Eric Pacuit. "Dynamic logics of evidence-based beliefs". In: *Studia Logica* 99.1 (2011), pp. 61–92 (cited on page 68).

[47] 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 page 3).

[48] 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 35).

[49] 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 35).

[50] 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).

[51] 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).

[52] 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).

[53] 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).

[54] 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, 9).

[55] 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).

[56] 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).

[57] 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).

[58] 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, 15).