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 (old)

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

Resolving such conflicts and determining the true facts is therefore an important task. Truth discovery has emerged as a set of techniques to achieve this by considering the *trustworthiness* of sources [27, 19, 7]. The general principle is that true facts are those claimed by trustworthy sources, and trustworthy sources are those that claim believable facts. Application areas include real-time traffic navigation [14], drug side-effect discovery [30], crowdsourcing and social sensing [52, 42, 29].

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

Example 2.0.1. *Let o and p represent two images for which crowdsourcing workers are asked to provide labels (in the truth discovery terminology, o and p are called* objects*). Consider workers (the data sources) s, t, u and v who put forward potential labels f, g for o, and h, i for p, as shown in Fig. 2.1; such potential answers are termed* facts*. In the graphical representation, sources, facts and objects are shown from left to right, and the edges indicate claims made by sources and the objects to which facts relate.*

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

Taking a trust-aware approach, however, we can look beyond object o to consider the trustworthiness *of s and t. Indeed, when it comes to object p, t agrees with two extra sources*

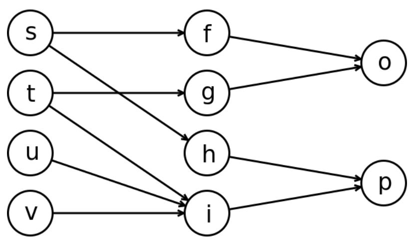


Figure 2.1: Illustrative example of a dataset to which truth discovery can be applied with data sources $\{s, t, u, v\}$, facts $\{f, g, h, i\}$ and objects $\{o, p\}$

u and v, whereas s disagrees with everyone. In principle there could be hundreds *of extra sources here instead of just two, in which case the effect would be even more striking. We may conclude that s is* less trustworthy *than t. Returning to o, we see that g is supported by a more trustworthy source, and conclude that it should be accepted over f.*

Many truth discovery algorithms have been proposed in the literature with a wide range of techniques used, e.g. iterative heuristic-based methods [35, 17], probabilistic models [50], maximum likelihood estimation and optimisation-based methods [28], and neural network models [24, 31, 43]. It is common for such algorithms to be evaluated empirically by running them against real-world or synthetic datasets for which the true facts are already known; this allows *accuracy* and other metrics to be calculated, and permits comparison between algorithms (see [41] for a systematic empirical evaluation of this kind). This may be accompanied by some theoretical analysis, such as calculating run-time complexity [19], proving convergence of an iterative algorithm [51], or proving convergence to the 'true' facts under certain assumptions on the distribution of source trustworthiness [47, 45, 18].

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

In this work we take first steps towards a more general approach, in which we aim to study truth discovery without reference to any specific methodology or probabilistic framework. To do so we note the similarities between truth discovery and problems such as judgment aggregation [15], voting theory [55] ranking and recommendation systems [1, 2, 3, 40] in which the *axiomatic approach* of social choice has been successfully applied. In taking the axiomatic approach one aims to formulate *axioms* that encode intuitively desirable properties that an algorithm may possess. The interaction between these axioms can then be studied; typical results include *impossibility results*, where it is shown that a set of axioms cannot hold simultaneously, and *characterisation results*, where it is shown that a set of axioms are uniquely satisfied by a particular algorithm.

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

With this in mind, we develop a simplified framework for truth discovery in which axioms can be formulated, and go on to give both an impossibility result and an axiomatic characterisation of a baseline voting algorithm. We also analyse the class of *recursive* truth discovery algorithms, which includes many existing examples from the literature. In particular, we analyse the well-known algorithm *Sums* [35] with respect to the axioms.

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

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

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

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

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

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

Chapter outline. Our framework is introduced and formally defined in Section 2.1. Section 2.2 provides examples of truth discovery algorithms from the literature expressed in the framework. In Section 2.3 we formally introduce the axioms and present an impossibility result showing a subset of these cannot all be satisfied simultaneously. The examples of Section 2.2 are then revisited in Section 2.4, where we analyse them with respect to the axioms and propose modifications to resolve some axiom failures. In Section 2.5 we extend the framework to allow variable domains of sources, facts and objects, and give an impossibility result similar to that of Section 2.3. We discuss the axioms in Section 2.6 and related work in Section 2.7. We conclude in Section 2.8. Missing proofs are given in Appendix A.

2.1 An idealised framework for truth discovery

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

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

2.1.1 Truth discovery networks

A core definition of the framework is that of a *truth discovery network*, which represents the input to a truth discovery problem. We model this as a tripartite graph

[2] We generalise to variable domains in Section 2.5.

with certain constraints on its structure, in keeping with approaches taken throughout the truth discovery literature [50, 19].

Definition 2.1.1. *A* truth discovery network *(hereafter a* TD network*) is a directed graph $N = (V, E)$ where $V = \mathcal{S} \cup \mathcal{F} \cup \mathcal{O}$, and $E \subseteq (\mathcal{S} \times \mathcal{F}) \cup (\mathcal{F} \times \mathcal{O})$ has the following properties:*

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

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

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

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

Figure 2.1 (page 3) provides an example of a TD network. Note that there is no requirement that a source makes a claim for *every* object, or even that a source makes any claims at all. This reflects the fact that truth discovery datasets are in practise extremely sparse, i.e. each individual source makes few claims. Conversely, we allow for facts that receive no claims from any sources.

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

A special case of our framework is the binary case in which every object has exactly two associated facts. This brings us close to the setting studied in *judgment aggregation* [15] and, specifically (since sources do not necessarily claim a fact associated to every object) to the setting of *binary aggregation with abstentions* [9, 11]. An important difference, however, is that for simplicity we do not assume any *constraints* on the possible configurations of true facts across *different* objects. That is, any combination of facts is feasible. In judgment aggregation such an assumption has the effect of neutralising the impossibility results that arise in that domain (see, e.g., [9]). We shall see that that is not the case in our setting.

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

2.1.2 Truth discovery operators

Having defined the input to a truth discovery problem, the output must be defined. Contrary to many approaches in the truth discovery literature which output numeric *trust scores* for sources and *belief scores* for facts [50, 35, 17, 54, 52, 53], we consider the primary output to be *rankings* of the sources and facts. To the extent

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

that we do consider numeric scores, it is only to induce a ranking. This is because we are chiefly interested in *ordinal properties* rather than quantitative values. Indeed, for the theoretical analysis we wish to perform it is only important that a source is *more trustworthy* than another; the particular numeric scores produced by an algorithm are irrelevant.

Moreover, the scores produced by existing algorithms may have no semantic meaning [35], and so referring to numeric values is not meaningful when comparing across algorithms. In this case it is only the rankings of sources and facts that can be compared, which is further motivation for our choice. This point of view is also common across the social choice literature.

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

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

Definition 2.1.2. *An* ordinal truth discovery operator T *(hereafter* TD operator*) is a mapping $T : \mathcal{N} \to \mathcal{L}(\mathcal{S}) \times \mathcal{L}(\mathcal{F})$. We shall write $T(N) = (\sqsubseteq_N^T, \preceq_N^T)$, i.e. $\sqsubseteq_N^T$ is a total preorder on $\mathcal{S}$ and $\preceq_N^T$ is a total preorder on $\mathcal{F}$.*

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

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

Definition 2.1.3. *A* numerical TD operator *is a mapping $T : \mathcal{N} \to \mathbb{R}^{\mathcal{S} \cup \mathcal{F}}$, i.e. T assigns to each TD network N a function $T(N) = T_N : \mathcal{S} \cup \mathcal{F} \to \mathbb{R}$. For $s \in \mathcal{S}$, $T_N(s)$ is the* trust score *for s in the network N according to T; for $f \in \mathcal{F}$, $T_N(f)$ is the* belief score *for f. The set of all numerical TD operators will be denoted by $\mathcal{T}_{Num}$.*

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

It is worth noting that yet other truth discovery algorithms output neither rankings nor numeric scores for facts, but only a single 'true' fact for each object [28, 10, 49]. This is also the approach taken in judgment aggregation, where an *aggregation rule* selects which formulas are to be taken as true. In the case of finitely many possible facts, such algorithms can be modelled in our framework as numerical operators where $T_N(f) = 1$ for each identified 'true' fact f, and $T_N(g) = 0$ for other facts g. To go in the reverse direction and obtain the 'true' facts according to an operator, one may simply select the set of facts for each object that rank maximally.

2.2 Examples of truth discovery operators

Our framework can capture some operators that have been proposed in the truth discovery literature. In this section we provide two concrete examples: *Voting*, which is a simple approach commonly used as a baseline method, and *Sums* [35]. We go on to outline the class of *recursive operators* – of which *Sums* is an instance – which contains many more examples from the literature.

2.2.1 Voting

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

Definition 2.2.1. Voting *is the numerical operator defined as follows: for any network $N \in \mathcal{N}$, $s \in \mathcal{S}$ and $f \in \mathcal{F}$, $T_N(s) = 1$ and $T_N(f) = |\mathsf{src}_N(f)|$.*

Consider the network N shown in Fig. 2.1. Facts f, g and h each receive one vote, whereas i receives 3. The fact ranking induced by *Voting* is therefore $f \approx g \approx h \prec i$. On the other hand, all sources receive a trust score of 1 and therefore rank equally.

2.2.2 Sums

Sums [35] is a simple and well-known operator adapted from the *Hubs and Authorities* [21] algorithm for ranking web pages. The algorithm operates iteratively and recursively, assigning each source and fact a sequences of scores, with the final scores taken as the limit of the sequence.

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

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

$$T'_N(s) = \sum_{f \in \mathsf{facts}_N(s)} T_N(f); \quad T'_N(f) = \sum_{s \in \mathsf{src}_N(f)} T'_N(s) \tag{2.1}$$

Consider again the network N shown in Fig. 2.1. It can be shown that, with T denoting the limiting scores from *Sums* with normalisation, we have $T_N(s) = 0$,

$T_N(t) = 1$, and $T_N(u) = T_N(v) = \sqrt{2}/2$. The induced ranking of sources is therefore $s \sqsubset u \simeq v \sqsubset t$.

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

2.2.3 Recursive truth discovery operators

The iterative and recursive aspect of *Sums* is hoped to result in the desired mutual dependence between trust and belief scores: namely that sources claiming high-belief facts are seen as trustworthy, and vice versa. In fact, this recursive approach is near universal across the truth discovery literature (see for instance [48, 14, 53, 28, 17, 54]). As such it is appropriate to identify the class of *recursive operators* as an important subset of $\mathcal{T}_{Num}$. To make a formal definition we first define an *iterative operator*.

Definition 2.2.2. *An* iterative operator *is a sequence* $(T^n)_{n\in\mathbb{N}}$ *of numerical operators. An iterative operator is said to* converge *to a numerical operator* T^* *if* $\lim_{n\to\infty} T^n_N(z) = T^*_N(z)$ *for all networks* N *and* $z \in \mathcal{S} \cup \mathcal{F}$*. In such case the iterative operator can be identified with the ordinal operator induced by its limit* T^**.*

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

Definition 2.2.3. *An iterative operator* $(T^n)_{n\in\mathbb{N}}$ *is said to be* recursive *if there is a function* $U : \mathcal{T}_{Num} \to \mathcal{T}_{Num}$ *such that* $T^{n+1} = U(T^n)$ *for all* $n \in \mathbb{N}$*.*

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

Returning to *Sums*, we see that (2.1) defines a mapping $\mathcal{T}_{Num} \to \mathcal{T}_{Num}$ and consequently an update function U^{Sums}. The normalisation step can be considered a separate update function norm which maps any numerical operator T to T', where[4]

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

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

Many other existing algorithms proposed in the literature can also be realised as recursive operators in the framework, such as *Investment*, *PooledInvestment* [35], *TruthFinder* [50], LDT [53] and others.

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

2.3 Axioms for truth discovery

Having laid out the formal framework, we now introduce axioms for truth discovery. To start with, we consider axioms which encode a desirable theoretical property that we believe any 'reasonable' operator T should satisfy. Several properties of this nature can be obtained by adapting existing axioms from the social choice literature (e.g. from voting [8], ranking systems [40, 1] and judgement aggregation [15]), to our framework.

However, the correspondence between truth discovery and classical social choice problems – such as voting – has its limits. To show this, we translate the famous Independence of Irrelevant Alternatives (IIA) axiom [4] to our setting, and argue that it is actually an *undesirable* property. Indeed, it will be seen that this translated axiom, in combination with two basic desirable axioms, leads to *Voting*-like behaviour in every network, which is undesirable for the reasons given in Section 2.2.1. Furthermore, a slight strengthening of the IIA axiom completely characterises the fact ranking component of *Voting*. These results formalise the intuition that truth discovery's consideration of source-trustworthiness leads to fundamental differences from classical social choice problems.

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

2.3.1 Coherence

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

Whilst this intuition is difficult to formalise in general, it is possible to do so in particular cases where there are obvious means by which to compare the set of facts for two sources (and vice versa). This situation is considered in the axiomatic analysis of ranking and reputation systems under the name *Transitivity* [40, 1], and we adapt it to truth discovery in this section. A preliminary definition is required.

Definition 2.3.1. *Let T be a TD operator, N be a TD network and $Y, Y' \subseteq \mathcal{F}$. We say Y is* less believable *than Y' with respect to N and T if there is a bijection $\varphi : Y \to Y'$ such that $f \preceq_N^T \varphi(f)$ for each $f \in Y$, and $\hat{f} \prec_N^T \varphi(\hat{f})$ for some $\hat{f} \in Y$.*

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

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

Coherence.

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

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

2.3.2 Symmetry

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

Definition 2.3.2. *Two TD networks N and N' are* equivalent *if there is a graph isomorphism π between them that preserves sources, facts and objects, i.e., $\pi(s) \in \mathcal{S}$, $\pi(f) \in \mathcal{F}$ and $\pi(o) \in \mathcal{O}$ for all $s \in \mathcal{S}$, $f \in \mathcal{F}$ and $o \in \mathcal{O}$. In such case we write $\pi(N)$ for N'.*

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

In the theory of voting in social choice, Symmetry as above is expressed as two axioms: *Anonymity*, where output is insensitive to the names of voters, and *Neutrality*, where output is insensitive to the names of alternatives [55]. Analogous axioms are also used in judgment aggregation.

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

2.3.3 Fact ranking axioms

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

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

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

That is, f cannot do better than to be claimed by all sources when T satisfies Unanimity, and cannot do worse than to be claimed by none when T satisfies Groundedness. Unanimity here is a truth discovery rendition of the same axiom in judgment aggregation, and can also be compared to the *weak Paretian* property in voting [8].

Groundedness is a version of the same axiom studied in the analysis of collective annotation [25].

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

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

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

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

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

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

2.3.4 Independence axioms

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

An analogous truth discovery axiom states that the ranking of facts f_1 and f_2 for some object o depends only on the claims relating to o. Intuitively, this is *not* a desirable property. Indeed, we have already seen in Example 2.0.1 that the claims for object p in the network from Fig. 2.1 can play an important role in determining the ranking of f and g for object o, but the adapted IIA axiom precludes this.

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

Per-object Independence (POI).

Let $o \in \mathcal{O}$. Suppose N_1, N_2 are networks such that $F_o = \mathsf{obj}_{N_1}^{-1}(o) = \mathsf{obj}_{N_2}^{-1}(o)$ and $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$ for each $f \in F_o$. Then the restrictions

[5] One could also consider the weak version, in which we only require $g \preceq_{N'}^T f$ in the consequent; we discuss this in Section 2.6.

[6] Note that collusion has been studied in the truth discovery literature (e.g. [12, 6, 13]).

of $\preceq^T_{N_1}$ and $\preceq^T_{N_2}$ to F_o are equal; that is, $f_1 \preceq^T_{N_1} f_2$ iff $f_1 \preceq^T_{N_2} f_2$ for all $f_1, f_2 \in F_o$.

Considering Fig. 2.1 again, POI implies that the ranking of f and g remains the same if the claims for h and i are removed. But in this case, Symmetry implies $f \approx g$. Similarly, the ranking of h and i remains the same if the claims for f and g are removed. In this case, Symmetry together with Monotonicity implies $h \prec i$, since $|\mathsf{src}_N(h)| < |\mathsf{src}_N(i)|$.

This observation forms the basis of the following result, which formalises the undesirability of POI: in the presence of our less controversial requirements of Symmetry and Monotonicity, it forces *Voting*-like behaviour within $\mathsf{obj}_N^{-1}(o)$ for each $o \in \mathcal{O}$. We note that, for the special case of binary networks, similar results have been shown in the literature on binary aggregation with abstentions [9].

Theorem 2.3.1. *Let T be any operator satisfying Symmetry, Monotonicity and POI. Then for any $N \in \mathcal{N}$, $o \in \mathcal{O}$ and $f_1, f_2 \in \mathsf{obj}_N^{-1}(o)$ we have $f_1 \preceq^T_N f_2$ iff $|\mathsf{src}_N(f_1)| \leq |\mathsf{src}_N(f_2)|$.*

Proof (sketch). We will sketch the main ideas of the proof here with some technical details omitted; see Appendix A for the full proof. Let N be a network, o be an object and $f_1, f_2 \in \mathsf{obj}_N^{-1}(o)$. Consider N' obtained by removing from N all claims for objects other than o. By POI, we have $f_1 \preceq^T_N f_2$ iff $f_1 \preceq^T_{N'} f_2$. Since $|\mathsf{src}_N(f_j)| = |\mathsf{src}_{N'}(f_j)|$ also ($j \in \{1, 2\}$), it is sufficient for the proof to show that $f_1 \preceq^T_{N'} f_2$ iff $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$.

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

Otherwise, $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Consider N'' where k sources from $\mathsf{src}_{N'}(f_2)$ are removed, and all other claims remain. By Symmetry as above, $f_1 \approx^T_{N''} f_2$. Applying Monotonicity k times we can produce N' from N'' and get $f_1 \prec^T_{N'} f_2$ as desired.

For the 'only if' statement, suppose $f_1 \preceq^T_{N'} f_2$ but, for contradiction, $|\mathsf{src}_{N'}(f_1)| > |\mathsf{src}_{N'}(f_2)|$. Applying Monotonicity again as above we get $f_1 \succ^T_{N'} f_2$ and the required contradiction. $\square$

Recall that Coherence formalises the idea that source-trustworthiness should inform the fact ranking, and vice versa. Clearly *Voting* does not conform to this idea, and in fact even the object-wise voting patterns in Theorem 2.3.1 are incompatible with Coherence. This can easily be seen in the network in Fig. 2.1 where, regarding object p, we have $|\mathsf{src}_N(h)| < |\mathsf{src}_N(i)|$ (hence $h \prec^T_N i$) and, regarding object o, we have $|\mathsf{src}_N(f)| = |\mathsf{src}_N(g)|$ (hence $f \approx^T_N g$). Hence $\mathsf{facts}_N(s)$ is less believable than $\mathsf{facts}_N(t)$. If Coherence held this would give $s \sqsubset^T_N t$, but then $\mathsf{src}_N(f)$ is less trustworthy than $\mathsf{src}_N(g)$, giving $f \prec^T_N g$ – a contradiction. From this discussion and Theorem 2.3.1 we obtain as a corollary the following first impossibility result for truth discovery.

Theorem 2.3.2. *There is no TD operator satisfying Coherence, Symmetry, Monotonicity and POI.*

Given that Theorem 2.3.1 characterises the fact ranking of *Voting* for facts relating to a single object, it is natural to ask if there is a stronger form of independence that guarantees this behaviour across *all* facts. As our next result shows, the answer is *yes*, and the necessary axiom is obtained by ignoring the role of objects altogether for fact ranking.

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

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

Theorem 2.3.3. *Suppose $|\mathcal{O}| \geq 3$. Then an operator T satisfies Strong Independence, Monotonicity and Symmetry if and only if for any network N and $f_1, f_2 \in \mathcal{F}$ we have*

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

Theorem 2.3.3 can be seen as a characterisation of the class of TD operators that rank facts in the same way as *Voting*. The proof is similar to that of Theorem 2.3.1, but uses a different transformation to obtain a modified network N' in the first step.

We have established that neither POI nor Strong Independence are satisfactory axioms for truth discovery, and a weaker independence property is required. Figure 2.1 can help us once again in this regard. Whereas POI and Strong Independence would say that facts h and i are irrelevant to f, the argument with Coherence for Theorem 2.3.2 suggests otherwise due the indirect links via the sources. We therefore propose that only when there is no (undirected) path between two nodes can we consider them to be truly irrelevant to each other. That is, nodes are relevant to each other iff they lie in the same *connected component* of the network.

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

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

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

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

2.4 Satisfaction of the axioms

With the axioms formally defined, we can now consider whether they are satisfied by the example operators of Section 2.2. *Voting* can be analysed outright; for *Sums* we require some preliminary results giving sufficient conditions for iterative and recursive operators to satisfy various axioms. It will be seen that neither *Voting* nor *Sums* satisfy all our desirable axioms, but it is possible to modify each operator to gain some improvement with respect to the axioms.

2.4.1 Voting

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

Theorem 2.4.1. Voting *satisfies Symmetry, Unanimity, Groundedness, Monotonicity, POI, Strong Independence and PCI.* Voting *does not satisfy Coherence.*

The proof is straightforward, and is deferred to Appendix A. Note that once Symmetry, Monotonicity and POI are shown, the fact that *Voting* fails Coherence follows from our impossibility result (Theorem 2.3.2), and Fig. 2.1 serves as an explicit counterexample.

2.4.2 Iterative and recursive operators

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

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

Definition 2.4.1. *Let T be a numerical TD operator, N a network, $Y, Y' \subseteq \mathcal{F}$ and $\varepsilon, \rho > 0$. We say Y is (ε, ρ)-*less believable *than Y' with respect to N and T if there is a bijection $\varphi : Y \to Y'$ such that $T_N(f) - T_N(\varphi(f)) \leq \varepsilon$ for all $f \in Y$, and $T_N(\hat{f}) - T_N(\varphi(\hat{f})) \leq \varepsilon - \rho$ for some $\hat{f} \in Y$.*

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

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

This generalises Definition 2.3.1 by relaxing the requirement that $f \preceq_N^T \varphi(f)$, and instead requiring that f can only be more believable than $\varphi(f)$ by some threshold $\varepsilon > 0$. Definition 2.3.1 is recovered in the limiting case $\varepsilon \to 0$. We obtain a sufficient condition on the update function U for a recursive operator to satisfy Source-Coherence.

Lemma 2.4.1. *Let $U : \mathcal{T}_{Num} \to \mathcal{T}_{Num}$. For any prior operator T^{prior}, $\mathsf{rec}(T^{prior}, U)$ satisfies Source-Coherence if the following condition is satisfied: there exist $C, D > 0$ such that for all networks N and numerical operators T it holds that if $\mathsf{facts}_N(s_1)$ is (ε, ρ)-less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T, then $T'_N(s_1) - T'_N(s_2) \leq C\varepsilon - D\rho$, where $T' = U(T)$.*

The proof of Lemma 2.4.1 uses the following result, the proof of which is a straightforward application of the definition of the limit.

Lemma 2.4.2. *Let N be a truth discovery network and $(T^n)_{n\in\mathbb{N}}$ be a convergent iterative operator with limit T^*. Then for $f_1, f_2 \in \mathcal{F}$, $f_1 \preceq_N^{T^*} f_2$ if and only if*

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

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

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

Analogous statements for source rankings also hold.

Proof of Lemma 2.4.1. Let N be a network. Suppose U has the stated property and that $\mathsf{rec}(T^{\text{prior}}, U) = (T^n)_{n\in\mathbb{N}}$ converges to T^*. Suppose $\mathsf{facts}_N(s_1)$ is less trustworthy than $\mathsf{facts}_N(s_2)$ with respect to N and T^* under a bijection φ. We must show that $s_1 \sqsubset_N^{T^*} s_2$.

Now, there is some $\hat{f} \in \mathsf{facts}_N(s_1)$ with $\hat{f} \prec_N^{T^*} \varphi(\hat{f})$. The second part of Lemma 2.4.2 therefore applies; let ρ be as given there. Now let $\varepsilon > 0$. Since $f \preceq_N^{T^*} \varphi(f)$ for each $f \in \mathsf{facts}_N(s_1)$, we may apply Lemma 2.4.2 with $f, \varphi(f)$ and $\bar{\varepsilon} = \varepsilon/C$ to get that there is $K \in \mathbb{N}$ such that

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

and

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

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

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

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

Since $D\rho$ is positive and does not depend on ε, we get $s_1 \sqsubset_N^{T^*} s_2$ by Lemma 2.4.2. This shows that T^* satisfies Source-Coherence. $\square$

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

Lemma 2.4.3. $\mathsf{rec}(T^{prior}, U)$ *satisfies Fact-Coherence if there exist* $E, F > 0$ *such that for all networks* N *and numerical operators* T *it holds that if* $\mathsf{src}_N(f_1)$ *is* (ε, ρ)*-less trustworthy than* $\mathsf{src}_N(f_2)$ *with respect to* N *and* T'*, then* $T'_N(f_1) - T'_N(f_2) \leq E\varepsilon - F\rho$*, where* $T' = U(T)$*.*

Proof. The proof proceeds in an identical way to Lemma 2.4.1; the only difference is that we may simply take $K' = K$ in the final step. □

Note that there is asymmetry between Lemma 2.4.1 and Lemma 2.4.3 – in the condition on U in Lemma 2.4.1 we have $\mathsf{facts}_N(s_1)$ (ε, ρ)-less trustworthy than $\mathsf{facts}_N(s_2)$ with respect to T, whereas in Lemma 2.4.3 the corresponding condition is with respect to $T' = U(T)$. This reflects the manner in which *Sums* and other TD operators are typically defined: source trust scores are updated based on the fact scores of the previous iteration, whereas fact belief scores are updated based on the (new) trust scores in the *current* iteration.

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

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

Example 2.4.1. *Fix some* $\hat{f} \in \mathcal{F}$*, and define an iterative operator by*

$$T^n_N(s) = 1$$

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

That is, each T^n *is a modification of* Voting *in which we boost the score of all facts tied with* $\hat{f}$ *under* Voting *by* $1 - \frac{1}{n+1}$*. Since this additional weight is (strictly) less than 1 for each* n*, the ordinal operator induced by* T^n *is simply* Voting*, and therefore satisfies Symmetry. However, it is easy to see that the limit operator* T^* *has* $T^*_N(\hat{f}) = |\mathsf{src}_N(\hat{f})| + 1$*; this means* T^* *uses extra information beyond the structure of the network* N *in its ranking (namely, the identity of a selected fact* $\hat{f}$*) which violates Symmetry.*

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

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

Lemma 2.4.4. *Suppose* $(T^n)_{n\in\mathbb{N}}$ *converges to* T^**. Then*

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

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

As a consequence of Lemma 2.4.4, any recursive operator $\mathsf{rec}(T^{\text{prior}}, U)$ satisfies Symmetry whenever T^{prior} satisfies numerical Symmetry and U *preserves* numerical Symmetry, in the sense that $U(T)$ satisfies numerical Symmetry whenever T does (and similarly for PCI).

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

Lemma 2.4.5. *Suppose $(T^n)_{n\in\mathbb{N}}$ converges to T^*. Then*

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

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

Definition 2.4.2. *A numerical operator T satisfies* Improvement *if for each N, N' and f as in the statement of Monotonicity, we have $\delta(f) > \delta(g)$ for all $g \neq f$, where*

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

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

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

Proposition 2.4.1. *Suppose $(T^n)_{n\in\mathbb{N}}$ converges to T^*, and T^n satisfies Improvement for each $n \in \mathbb{N}$. Suppose also that $\inf_{n\in\mathbb{N}} \rho^n_{N,N'} > 0$ for each N, N' arising in the statement of Monotonicity. Then T^* satisfies Monotonicity.*

Proof. Let N, N' and f be as in the statement of Monotonicity, and suppose $g \preceq^{T^*}_N f$ for some $g \neq f$. We will show $g \prec^{T^*}_{N'} f$ using Lemma 2.4.2.

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

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

By Lemma 2.4.2, we have $g \prec^{T^*}_{N'} f$ as required. □

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

2.4.3 Sums

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

Theorem 2.4.2. Sums *satisfies Coherence, Symmetry, Unanimity and Groundedness.* Sums *does not satisfy POI, Strong Independence, PCI or Monotonicity.*

Proof (sketch). Symmetry, Unanimity and Groundedness can be easily shown from Lemma 2.4.4 and Lemma 2.4.5; the details can be found in the appendix. In the remainder of the proof, $(T^n)_{n\in\mathbb{N}}$ will denote the iterative operator *Sums*, T^* will denote the limit operator, and $U = \mathsf{norm} \circ U^{\mathsf{Sums}}$ will denote the update function for *Sums*.

Coherence. We will show Source-Coherence using Lemma 2.4.1. The argument for Fact-Coherence is similar (using Lemma 2.4.3) and can be found in the appendix.

Suppose $N \in \mathcal{N}$, $T \in \mathcal{T}_{Num}$, $\varepsilon, \rho > 0$, and $\mathsf{facts}_N(s_1)$ is (ε, ρ)-less believable than $\mathsf{facts}_N(s_2)$ with respect to N and T under a bijection $\varphi : \mathsf{facts}_N(s_1) \to \mathsf{facts}_N(s_2)$. By definition there is $\hat{f} \in \mathsf{facts}_N(s_1)$ such that $T_N(\hat{f}) - T_N(\varphi(\hat{f})) \leq \varepsilon - \rho$. By the remark after the proof of Lemma 2.4.1, we may assume without loss of generality that $\varepsilon < \frac{1}{|\mathcal{F}|}\rho$.

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

Note at this stage that it is possible to further weaken the hypotheses of Lemma 2.4.1: the result follows if U has the stated property not for *all* operators T, but only for those such that $T = T^n$ for some $n \in \mathbb{N}$. Next, note that if $T'_N(x) = 0$ for all $x \in \mathcal{S}$ then trust and belief scores are 0 in all subsequent iterations, and thus all sources rank equally in the limit T^*. But this means the hypothesis for Source-Coherence cannot be satisfied (there are no strict inequalities). We may therefore assume without loss of generality that $T'_N(x) \neq 0$ for at least one $x \in \mathcal{S}$. Therefore, by definition of norm,

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

where

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

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

via φ, we have

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

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

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

and so

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

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

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

Taking $C = 1$ and $D = \frac{1}{|\mathcal{F}|}$, the hypotheses of Lemma 2.4.1 are satisfied; thus *Sums* satisfies Source-Coherence.

POI, Strong Independence, PCI and Monotonicity. The remaining axioms are handled by counterexamples derived from the network shown in Fig. 2.2. It can be shown that, if N denotes this network, we have $T^*_N(f) = T^*_N(g) = 0$, so $f \approx^{T^*}_N g$.

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

For Monotonicity, consider the network N'' obtained from N by removing the edge (u, g). Then we still have $T^*_{N''}(f) = T^*_{N''}(g) = 0$, and in particular $f \preceq^{T^*}_{N''} g$.

Returning to N amounts to adding extra support for the fact g. Monotonicity would give $f \prec_N^{T^*} g$ here, but this is clearly false. Hence Monotonicity is not satisfied by *Sums*.

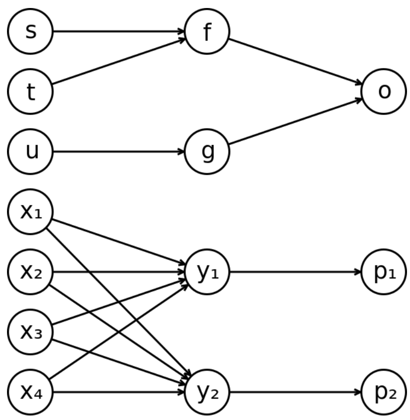


Figure 2.2: Network which yields counterexamples for POI, Strong Independence, PCI and Monotonicity for Sums

□

The key to the counterexamples derived from Fig. 2.2 in the above proof lies in the lower connected component, which – restricted to $\mathcal{S} \cup \mathcal{F}$ – is a *connected* bipartite graph. That is, each source x_i claims all facts in the component, and each fact y_j is claimed by all sources in the component. Moreover, sources elsewhere in the network claim fewer facts than the x_i, and facts elsewhere are claimed by fewer sources than the y_j.

Since *Sums* assigns scores by a simple sum, this results in the scores for the x_i and y_j dominating those of the other sources and facts. The normalisation step then divides these scores by the (comparatively large) maximum. As the next result shows, under certain conditions this causes scores to decrease *exponentially* and become 0 in the limit. In particular, we can generate pathological examples such as Fig. 2.2 where a whole connected component receives scores of 0, which leads to failure of Monotonicity and the independence axioms.

Proposition 2.4.2. *Let N be a network. Suppose there is $X \subseteq \mathcal{S}$, $Y \subseteq \mathcal{F}$ such that*

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

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

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

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

$$T_N^n(x) = 1 \quad (x \in X)$$
$$T_N^n(y) = 1 \quad (y \in Y)$$

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

Proof. We proceed by induction. The result is easy to show in the base case $n = 2$ since $|\mathsf{facts}_N(s)| \leq \frac{1}{2}|\mathsf{facts}_N(x)|$ for any $x \in X$ and $s \notin X$ (and similarly for facts). Assume the result holds for some $n > 1$. Write $T' = U^{\mathsf{Sums}}(T^n)$, so that $T^{n+1} = \mathsf{norm}(T')$. If $s \notin X$ then $\mathsf{facts}_N(s) \subseteq \mathcal{F} \setminus Y$, so

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

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

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

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

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

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

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

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

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

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

2.4.4 Modifying *Voting* and *Sums*

So far we have seen that neither of the basic operators *Voting* or *Sums* are completely satisfactory with respect to the axioms of Section 2.3. Armed with the knowledge of how and why certain axioms fail, one may wonder whether it is possible to modify the operators accordingly so that the axioms *are* satisfied. Presently we shall show that this is partially possible both in the case of *Voting* and *Sums*.

2.4.4.1 Voting

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

Definition 2.4.3. *For a network N, define a binary relation $\lhd_N$ on $\mathcal{S}$ by $s_1 \lhd_N s_2$ iff* $\mathsf{facts}_N(s_1)$ *is less believable than* $\mathsf{facts}_N(s_2)$ *with respect to* Voting. *The numerical operator* SC-Voting *(Source-Coherence Voting) is defined by*

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

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

Theorem 2.4.3. SC-Voting *satisfies Source-Coherence, Symmetry, Unanimity, Groundedness, Monotonicity, Fact-PCI, POI and Strong Independence. It does not satisfy Fact-Coherence or Source-PCI.*

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

Lemma 2.4.6. *$\lhd_N$ is transitive and irreflexive.*

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

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

$$\hat{f} \prec_N^T \varphi(\hat{f}) \preceq_N^T \varphi(\varphi(\hat{f}) \preceq_N^T \cdots \preceq_N^T \varphi^n(\hat{f}) \tag{2.2}$$

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

Since F is finite, the sequence $\varphi(\hat{f}), \varphi(\varphi(\hat{f})), \ldots$ must repeat at some point, i.e. there is $i < j$ such that $\varphi^i(\hat{f}) = \varphi^j(\hat{f})$. But then injectivity of φ implies that $\hat{f} = \varphi^{j-i}(\hat{f})$. Taking $n = j - i$ in (2.2) we get $\hat{f} \prec_N^T \hat{f}$ – a contradiction. ☐

Proof of Theorem 2.4.3 (sketch). Note that *SC-Voting* inherits Unanimity, Groundedness, Monotonicity, Fact-PCI, POI and Strong Independence from *Voting*, since these axioms only refer to the rankings of facts (which is the same for *SC-Voting* as for *Voting*).

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

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

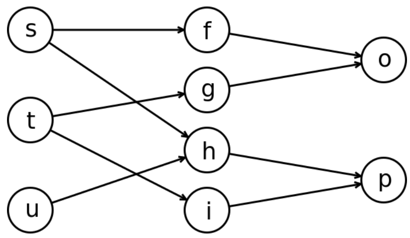


Figure 2.3: Fact-Coherence counterexample for SC-Voting

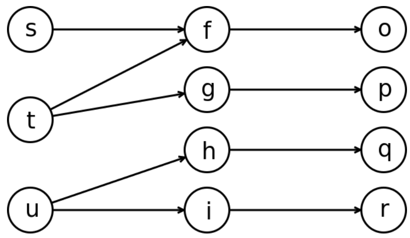


Figure 2.4: Source-PCI counterexample for SC-Voting

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

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

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

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

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

as required for Symmetry.

Fact-Coherence. Consider the network shown in Fig. 2.3. We have $f \approx g \approx i \prec h$. Source-Coherence between s and t gives $t \sqsubset s$. If Fact-Coherence held we would then get $g \prec f$, but this is not the case.

Source-PCI. Let N_1 denote the top connected component of the network shown in Fig. 2.4, and let N_2 denote the network as a whole. The fact ranking is the same in both networks: $g \approx h \approx i \prec f$. In N_1 sources s and t claim a different number of facts, so neither $s \lhd_{N_1} t$ nor $t \lhd_{N_1} s$. Consequently $W_{N_1}(s) = W_{N_1}(t) = \emptyset$ and $s \simeq_{N_1}^{T^{SCV}} t$.

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

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

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

2.4.4.2 Sums

Our main concern with *Sums* is the failure of PCI and Monotonicity. We have seen that this is in some sense caused by the normalisation step: in Fig. 2.2 the scores of f, g etc go to 0 in the limit after dividing the 'global' maximum score across the network, which happens to come from a different connected component.

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

However, this approach does not negate the undesirable effects of Proposition 2.4.2. Indeed, suppose the network in Fig. 2.2 was modified so that fact y_1 is associated with object o instead of p_1. Clearly Proposition 2.4.2 still applies after this change, and all sources and facts shown now belong to the same connected component. Therefore the 'per-component *Sums*' operator gives the same result as *Sums* itself, and in particular our Monotonicity counterexample still applies. Perhaps even worse, one can show that Coherence fails for this operator. We consider the loss of Coherence too high a price to pay for PCI.

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

Definition 2.4.4. UnboundedSums *is the recursive operator* $\mathsf{rec}(T^{prior}, U^{Sums})$ *where* $T_N^{prior}(s) = 1$, $T_N^{prior}(f) = |\mathsf{src}_N(f)|$ *and* U^{Sums} *is defined as in Section 2.2.2.*[9]

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

Theorem 2.4.4. UnboundedSums *is ordinally convergent in the following sense: there is an ordinal operator T^* such that for each network N there exists $J_N \in \mathbb{N}$ such that $T^n_N(s_1) \leq T^n_N(s_2)$ iff $s_1 \sqsubseteq^{T^*}_N s_2$ for all $n \geq J_N$ and $s_1, s_2 \in \mathcal{S}$ (and similarly for facts).*

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

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

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

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

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

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

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

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

So, fix $s_p, s_q \in \mathcal{S}$. Write $\delta_n = [v_n]_p - [v_n]_q$. First note that $A = MM^\top$ is symmetric, so the *spectral theorem* gives the existence of k orthogonal eigenvectors $x_1, \ldots, x_k$ for A [5, Theorem 7.29]. Let $\lambda_1, \ldots, \lambda_k$ be the corresponding eigenvalues. Form a $(k \times k)$-matrix P whose i-th column is x_i, and let $D = \text{diag}(\lambda_1, \ldots, \lambda_k)$. Then A can be diagonalised as $A = PDP^{-1}$. It follows that for any $n \in \mathbb{N}$, $A^n = PD^nP^{-1}$.

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

$$PD^n = \left[x_1 \mid \ldots \mid x_k\right] \begin{bmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_k^n \end{bmatrix} = \left[\lambda_1^n x_1 \mid \ldots \mid \lambda_k^n x_k\right]$$

and

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

which means

$$v_{n+1} = A^n v_1 = PD^n P^\top v_1 = \left[\lambda_1^n x_1 \mid \ldots \mid \lambda_k^n x_k\right] \begin{bmatrix} x_1 \cdot v_1 \\ \vdots \\ x_k \cdot v_1 \end{bmatrix} = \sum_{i=1}^{k} (x_i \cdot v_1)\lambda_i^n x_i$$

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

$$\delta_{n+1} = [v_n]_p - [v_n]_q = \sum_{i=1}^{k} (x_i \cdot v_1)\lambda_i^n \left([x_i]_p - [x_i]_q\right) = \sum_{i=1}^{k} r_i \lambda_i^n \tag{2.3}$$

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

Now, it is easy to see that $A = MM^\top$ is *positive semi-definite*, which means its eigenvalues $\lambda_1, \ldots, \lambda_k$ are all non-negative. We re-index the sum in (2.3) by grouping together the λ_i which are equal, to get

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

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

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

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

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

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

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

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

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

In light of Theorem 2.4.4, we may consider *UnboundedSums* itself as an ordinal operator T^*, where $s \sqsubseteq_N^{T^*} t$ iff $s \sqsubseteq_N^{T^{J_n}} t$ for each network N (and similarly for the fact ranking). Moreover, with the normalisation problems aside, *UnboundedSums* provides an example of a principled operator satisfying our two key axioms – Coherence and PCI. In particular, we escape the undesirable behaviour of *Sums* in Fig. 2.2; whereas *Sums* trivialises the ranking of sources and facts in the upper connected component, *UnboundedSums* allows meaningful ranking (e.g. we have $g \prec f$). In particular, the counterexample for Monotonicity for *Sums* is no longer a counterexample for *UnboundedSums*. We conjecture that *UnboundedSums* also satisfies Monotonicity, but this remains to be proven. For example, we have experimentally verified that *UnboundedSums* satisfies all the specific instances of Monotonicity with the starting network N as in Fig. 2.1.

Theorem 2.4.5. UnboundedSums *satisfies Coherence, Symmetry, Unanimity, Groundedness and PCI.* UnboundedSums *does not satisfy POI and Strong Independence.*

Proof (sketch). The proof that *UnboundedSums* satisfies Symmetry, PCI, Unanimity and Groundedness is routine, and we leave the details to the appendix. In what follows, let $(T^n)_{n\in\mathbb{N}}$ denote *UnboundedSums*, T^* denote the ordinal limit of *UnboundedSums*, and for a network N let J_N be as in Theorem 2.4.4. Then the rankings in N induced by T^n for $n \geq J_N$ are the same as T^*.

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

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

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

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

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

Table 2.1: Satisfaction of the axioms for the various operators. Recall that POI and Strong Independence are undesirable properties.

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

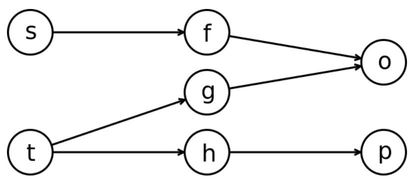


Figure 2.5: Counterexample for POI and Strong Independence for UnboundedSums

POI and Strong Independence. To show POI and Strong Independence are not satisfied, consider the network N shown in Fig. 2.5. It can be seen (e.g. by induction) that

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

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

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

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

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

To summarise this section, Table 2.1 shows which axioms are satisfied by each of the operators.

2.5 Variable domain truth discovery

So far, we have considered an arbitrary but fixed (finite) domain of sources, facts and objects $(\mathcal{S}, \mathcal{F}, \mathcal{O})$. Our operators and axioms were defined with respect to this

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

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

Formally, let $\mathbb{S}$, $\mathbb{F}$ and $\mathbb{O}$ be countably infinite sets of sources, facts and objects respectively. A *domain* is a triple $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, where $\mathcal{S} \subseteq \mathbb{S}$, $\mathcal{F} \subseteq \mathbb{F}$ and $\mathcal{O} \subseteq \mathbb{O}$ are finite, non-empty sets. We think of $\mathbb{S}$, $\mathbb{F}$ and $\mathbb{O}$ as being the 'universe' of possible sources, facts and objects, and a domain as the (finite) sets of entities under consideration in a particular TD problem. Given a domain $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, we define $\mathcal{D}$-networks and $\mathcal{D}$-operators as in Definitions 2.1.1 and 2.1.2.

Definition 2.5.1. *A* variable domain operator T *is a mapping which maps each domain $\mathcal{D}$ to a $\mathcal{D}$-operator $T_{\mathcal{D}}$.*

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

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

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

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

Like Symmetry, Isomorphism simply says that operators only care about the *structure* of the network, not the particular 'names' chosen for sources, facts and objects. Symmetry is obtained as the special case where N and N' are equivalent when seen as networks in a common domain $\mathcal{D}$. All the operators of Sections 2.2 and 2.4.4 satisfy Isomorphism.

Next we introduce a new monotonicity property. In what follows, for a network $N = (V, E)$ in domain $(\mathcal{S}, \mathcal{F}, \mathcal{O})$, $f \in \mathcal{F}$ and $\mathcal{S}' \subseteq \mathbb{S}$ finite and disjoint from $\mathcal{S}$, write $N + (\mathcal{S}', f)$ for the network in domain $(\mathcal{S} \cup \mathcal{S}', \mathcal{F}, \mathcal{O})$ with edge set $E \cup \{(s, f) \mid s \in \mathcal{S}'\}$, i.e. the extension of N where a collection of 'fresh' sources $\mathcal{S}'$ each claim f. For example, Fig. 2.6 shows $N + (\mathcal{S}', h)$ for the network N from Fig. 2.1 and new sources $\mathcal{S}' = \{x_1, \ldots, x_4\}$.

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

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

When paired with Isomorphism, Eventual Monotonicity takes on a form similar to postulates for *Improvement* and *Majority* operators in belief merging [22, 23]: there is a threshold $n \in \mathbb{N}$ such that f becomes strictly more believable than g after n new claims are added for f. That is, the identities of the new sources $\mathcal{S}'$ are irrelevant; it is just the *size* of $\mathcal{S}'$ that matters. We require a preliminary lemma.

Lemma 2.5.1. *Suppose a variable domain operator T satisfies Isomorphism. Let $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ be a domain, N a network in $\mathcal{D}$ and $f \in \mathcal{F}$. Then for all non-empty, finite sets $\mathcal{S}_1', \mathcal{S}_2' \subseteq \mathbb{S}$ disjoint from $\mathcal{S}$ with $|\mathcal{S}_1'| = |\mathcal{S}_2'|$,*

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

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

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

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

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

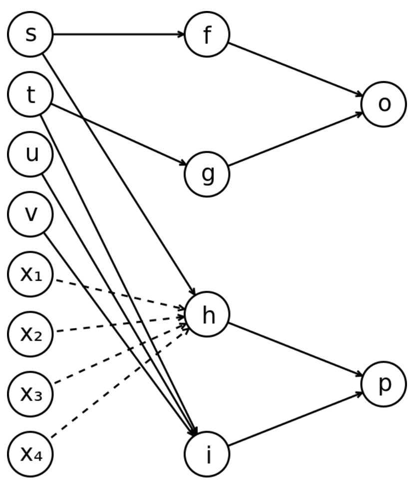


Figure 2.6: $N + (\mathcal{S}', h)$, where N is the network from Fig. 2.1 and $\mathcal{S}' = \{x_1, \ldots, x_4\}$.

operators T satisfying Isomorphism, write $\preceq^T_{N+(n\times f)}$ for $\preceq^T_{N+(\mathcal{S}',f)}$; Lemma 2.5.1 guarantees this is well-defined (i.e. does not depend on the particular choice of $\mathcal{S}'$). That is, $\preceq^T_{N+(n\times f)}$ is the fact ranking resulting from adding n new claims for f from fresh sources. We obtain an equivalent characterisation of Eventual Monotonicity, whose proof is almost immediate given Lemma 2.5.1.

Proposition 2.5.1. *Suppose T satisfies Isomorphism. Then T satisfies Eventual Monotonicity if and only if for all domains $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$, all networks N in $\mathcal{D}$ and distinct $f, g \in \mathcal{F}$, there is $n \in \mathbb{N}$ such that $g \prec^T_{N+(n\times f)} f$.*

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

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

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

Theorem 2.5.1. Voting, Sums, SC-Voting *and* UnboundedSums *satisfy Eventual Monotonicity.*

Proof (sketch). Let $\mathcal{D} = (\mathcal{S}, \mathcal{F}, \mathcal{O})$ be a domain, N a network in $\mathcal{D}$ and $f, g \in \mathcal{F}$. Given that Isomorphism holds for each operator, we sketch the proof via Proposition 2.5.1.

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

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

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

To conclude this section, we show that the impossibility result of Theorem 2.3.2 holds in the variable domain case if one replaces Monotonicity with Eventual Monotonicity and Symmetry with Isomorphism.

Theorem 2.5.2. *There is no variable domain operator satisfying Coherence, Isomorphism, Eventual Monotonicity and POI.*

Proof. For contradiction, suppose T is an operator satisfying the stated axioms. Let N be the network from Fig. 2.1, viewed as a network in domain $(\{s,t,u,v\}, \{f,g,h,i\}, \{o,p\})$. Applying Eventual Monotonicity with i and h, we have that there is N' with $i \prec^T_{N'} h$, where $N' = N + (\mathcal{S}', h)$ for some $\mathcal{S}' \subseteq \mathbb{S} \setminus \{s,t,u,v\}$. Since N' only adds claims for p-facts, POI applied to object o and Isomorphism give $f \approx^T_{N'} g$ (e.g. consider π which simply swaps s with t and f with g). From Source-Coherence we get $t \sqsubset^T_{N'} s$. But $\mathsf{src}_{N'}(f) = \{s\}$ and $\mathsf{src}_{N'}(g) = \{t\}$, so Fact-Coherence gives $g \prec^T_{N'} f$: contradiction! □

2.6 Discussion

In this section we discuss the axioms and their limitations. First, the version of Monotonicity we consider is a strict one: a new claim for f gives f a *strictly* positive boost in the fact believability ranking. This is also the case for Eventual Monotonicity in the variable domain case, where we require that some number of new claims make f strictly more believable than any other fact g. As noted in Section 2.3.3, this assumes there is no *collusion* between sources. Indeed, suppose sources s_1, s_2 are in collusion. For example, s_2 may agree to unconditionally back up all claims made by s_1. In this case a claim of f from s_1 alone should carry just as much weight as the claim from both s_1 and s_2. However, Monotonicity requires that f becomes strictly more believable when moving to the latter case.

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

> Weak Monotonicity. Let N, s, f, N' be as in the statement of Monotonicity. Then for all $g \neq f$, $g \preceq^T_N$ implies $g \preceq^T_{N'} f$.

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

In the same spirit, one could consider versions of Coherence only using weak comparisons. Say $\mathsf{facts}_N(s_1)$ is *weakly less believable* than $\mathsf{facts}_N(s_2)$ iff the condition in Definition 2.3.1 holds, but without the requirement that some $\hat{f} \in \mathsf{facts}_N(s_1)$ is strictly less believable than its counterpart $\varphi(\hat{f})$ in $\mathsf{facts}_N(s_2)$, and define $\mathsf{src}_N(f_1)$ weakly less trustworthy than $\mathsf{src}_N(f_2)$ in a similar way. The weak analogue of Coherence is as follows.

> Weak Coherence.
>
> For any network N, $\mathsf{facts}_N(s_1)$ weakly less believable than $\mathsf{facts}_N(s_2)$ implies $s_1 \sqsubseteq^T_N s_2$, and $\mathsf{src}_N(f_1)$ weakly less trustworthy than $\mathsf{src}_N(f_2)$ implies $f_1 \preceq^T_N f_2$.

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

A natural question now arises. Does the impossibility result of Theorem 2.3.2 still hold with these new axioms? We have an answer in the negative: the 'flat' operator, which sets all sources and facts equally ranked in all networks, satisfies all the axioms of the would-be impossibility.

Proposition 2.6.1. *Define an operator T by $s_1 \simeq^T_N s_2$ and $f \approx^T_N f_2$ for all networks N, sources s_1, s_2 and facts f_1, f_2. Then T satisfies Coherence, Weak Coherence, Symmetry, Weak Monotonicity and POI.*

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

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

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

Proposition 2.6.2. Voting, Sums *and* UnboundedSums *satisfy Weak Coherence*

Proof (sketch).

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

Sums. First, one may adapt Definition 2.4.1 to a numerical variant of a set of facts Y being *weakly* less believable than Y', by dropping all references to ρ. We then have an analogue of Lemma 2.4.1, and Weak Coherence for *Sums* follows by an argument similar to the one used to show Coherence using Lemma 2.4.1.

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

Proposition 2.6.2 indicates that Weak Coherence may in fact be too weak to capture the original intuition behind Coherence – namely, that there should be a mutual dependence between trustworthy sources and believable facts – since it does not even rule out *Voting*. Instead, Weak Coherence can be seen as a simple requirement which only rules out undesirable behaviour, and complements (strict) Coherence.

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

2.7 Related work

In this section we discuss related work.

Ranking systems. Altman and Tennenholtz [1] initiated axiomatic study of ranking systems. First we discuss their framework in relation to ours, and then turn to their axioms. In their framework, a ranking system F maps any (finite) directed graph $G = (V, E)$ to a total preorder $\leq_G^F$ on the vertex set V. In their view this is a variation of the classical social choice setting, in which the set of voters and alternatives coincide. Nodes $v \in V$ "vote" on their peers in V by a form of approval voting [26]: an edge $v \to u$ is interpret as a vote for u from v. A ranking system then outputs a ranking of V, following the general intuition that the more "votes" v receives (i.e. the more incoming edges), the higher v should rank. As with the

[12] Indeed, we conjectured in Section 2.4 that the stronger axiom (strict) Monotonicity holds for *UnboundedSums*. As with Monotonicity, experimental evidence from various starting networks N suggests that Weak Monotonicity is likely to hold.

ranking of facts in truth discovery, this does not necessarily mean ranking nodes simply by the *number* of votes received, since the *quality* of the voters should also be taken in account. For example, a ranking system may prioritise nodes which receive few votes from highly ranked nodes over those with many votes from lower ranked nodes.

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

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

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

However, some characteristic aspects of the truth discovery problem are lost in this translation to ranking systems. Notably, the objects play no role at all in G_N. Sources and facts are also treated symmetrically, where they perhaps should not be. For example, a fact f receiving more claims than g is beneficial for f, all else being equal (see Monotonicity), but a source s claiming more facts than t does not tell us anything about the relative trustworthiness of s and t.

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

In [1], Altman and Tennenholtz also introduce axioms for ranking systems. Their first set of axioms deal with the transitive effects of voting when the alternatives are the voters themselves. As mentioned in Section 2.3, these axioms provided the inspiration for Coherence. The core idea is that if the predecessors of a node v are weaker than those of u, then v should be ranked below u. If v additionally has *more* predecessors, v should rank *strictly* below. Coherence applies this idea both in the direction of sources-to-facts (Fact-Coherence) and from facts-to-sources (Source-Coherence). A notable difference is that we only consider the case where the number of sources for two facts (or the number of facts, for two sources) is the same. For example, a source claiming more facts does not give it the strict boost Transitivity would dictate. Under the mapping $N \mapsto G_N$ described above, any ranking system satisfying Transitivity induces a truth discovery operator which satisfies Coherence.

The other axiom in [1] is an independence axiom RIIA (ranked independence of irrelevant alternatives), which adapts the classical IIA axiom from social choice theory to the ranking system setting, although in a different manner to our independence axioms of Section 2.3.4. We describe the axiom in rough terms, deferring to the work itself for the technical details. Suppose the relative ranking of u_1's predecessors compared to u_2's predecessors is the same as that of v_1's compared to v_2's.

Then RIIA requires $u_1 \leq u_2$ iff $v_1 \leq v_2$. Informally, "the relative ranking of two agents must only depend on the pairwise comparison of the ranks of their predecessors" [1]. While we do not have an analogous axiom, the idea can be adapted to truth discovery networks. Intuitively, such an axiom would state that the ranking of two facts depends only on comparisons between their corresponding sources (and similar for the ranking of sources).

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

PageRank. PageRank [34] is a well-known algorithm for ranking web pages based on the hyperlink structure of the web, viewed as a directed graph. It has also been studied through the lens of social choice and characterised axiomatically [2, 44].[13] In [2] the authors propose several *invariance axioms*, each of which requires that the ranking of pages is not affected by a certain transformation of the web graph. For example, the axiom *Self Edge* says that adding a self loop from a page a to itself does not change the relative ranking of other pages, and results in a strictly positive boost for a (c.f. Monotonicity). However, if we identify a truth discovery network N with the graph G_N as described above, most of the transformations involved do not respect the bipartite, symmetric structure of G_N. That is, the transformed graph does not correspond to any $G_{N'}$, for a network N'. Consequently, the PageRank axioms have no truth discovery counterpart in our setting. The only exception is *Isomorphism*, where the transformation in question is graph isomorphism; this axiom is analogous to our Symmetry and Isomorphism axioms. However, since PageRank is similar to the *Hubs and Authorities* [21] algorithm on which Sums is based – which also uses the link structure of the web to rank pages – we expect there may be additional axioms which can be expressed both for general graphs *and* truth discovery networks, satisfied by PageRank and Sums. We leave the task of finding such axioms to future work.

2.8 Conclusion

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

It was possible to adapt many axioms from social choice theory and related areas. In particular, the *Transitivity* axiom studied in the context of ranking systems [40, 1] took on new life in the form of Coherence, which we consider a core axiom for TD operators. We proceeded to establish the differences between our

[13] Wąs and Skibski [44] axiomatise the *numerical scores* of PageRank, whereas Altman and Tennenholtz [2] axiomatise the resulting ranking. Moreover, Wąs and Skibski point out that Altman and Tennenholtz in fact only consider a simplified version of PageRank called *Katz prestige*, defined only for strongly connected graphs.

setting and classical social choice by considering independence axioms. This led to an impossibility result and an axiomatic characterisation of the baseline *Voting* method.

On the practical side, we analysed the existing TD algorithm *Sums* and found that, surprisingly, it fails PCI. This is a serious issue for *Sums* which has not been discussed in the literature to date, and its discovery here highlights the benefits of the axiomatic method. To resolve this, we suggested a modification to *Sums* – which we call *UnboundedSums* – for which PCI *is* satisfied. However, while *UnboundedSums* resolves axiomatic problems with *Sums*, it may introduce computational difficulties (since the numeric scores involved grow without bound). We leave further investigation of such issues to future work.

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

There is also scope for extensions to our model of truth discovery in the framework itself. For example, even in the variable domain setting of Section 2.5, we make the somewhat simplistic assumption that there are only finitely many possible facts for sources to claim. This effectively means we can only consider *categorical values*; modelling an object whose domain is the set of real numbers, for example, is not straightforward in our framework.

Next, our model does not account for any associations or constraints between objects, whereas in reality the belief in a fact for one object may strengthen or weaken our belief in other facts for related objects. These types of constraints or correlations have been studied both on the theoretical side (e.g. in judgment aggregation) and practical side in truth discovery [48].

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

Finally, it may be argued that truth discovery as formulated in this chapter risks simply to find *consensus* among sources, rather than the *truth*. To remedy this, the framework could be extended to model the possible states of the world and thus the *ground truth* (c.f. [33]). Upon doing so one could investigate how well, and under what conditions, an operator is able to recover the truth from a TD network. Such truth-tracking methods have also been studied in judgment aggregation and belief fusion [16, 20].

3 Truth Discovery

[TODO: Introduction]

3.1 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 [35, 50, 42, 13, 53], in general truth discovery methods also handle continuous values [28, 47].

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 3.1.1. *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.

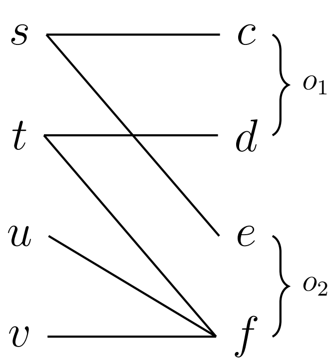


Figure 3.1: Illustrative example of a truth discovery problem, with sources s, t, u, v, object o_1 with associated claims c and d, and o_2 with claims e and f.

- $D = \{D_o\}_{o \in O}$ *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 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 3.1.1. *The network illustrated in Fig. 3.1 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 3.1.1 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* [15] and, specifically (since sources do not necessarily provide a claim for each object) to the setting of *binary aggregation with abstentions* [9, 11]. 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 [28, 10, 49], whereas others give an assessment of the believability (or confidence, probability etc...) of *each claim* (o,v) [50, 35, 17, 54, 52, 53]. 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 3.1.2. *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 3.2 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.

3.2 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* [35], *TruthFinder* [50] and *CRH* [28]. 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*.

3.2.1 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 3.2.1. T^{vote} *is the operator defined by*

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

Applying T^{vote} to the network in Fig. 3.1, 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 3.2.2. *A* weighting w *maps each network N to a function $w_N : S \to \mathbb{R}$. The associated* weighted voting *operator T^w is defined by*

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

Note that 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 that we allow the possibility of "untrustworthy" sources with $w_N(s) < 0$. Reports from such sources *decrease* the credibility of a claim.

Example 3.2.1. *Set*

$$w_N^{\text{agg}}(s) = \frac{1}{|\operatorname{cl}_N(s)|} \sum_{c \in \operatorname{cl}_N(s)} |\operatorname{src}_N(c)|.$$

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. 3.1, 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 3.6. [TODO: Check afterwards.]

[1] This is often called *majority voting* in the truth discovery literature (e.g. [27, 46, 28]), but using the terminology of social choice theory it is better described as *plurality voting*.

3.2.2 Recursive Operators

To capture the mutual dependence between trust in sources and belief in claims, truth discovery methods generally involve recursive computation [35, 50, 48, 14, 53, 28, 17, 54]. 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 3.2.3. *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 3.5. For now, we introduce examples of recursive operators from the literature.

Sums. Sums [35] is a simple and well-known operator adapted from the *Hubs and Authorities* [21] 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 3.2.4. Sums *is the recursive scheme* $(\mathcal{D}, T^0, U)$*, where* $\mathcal{D}$ *is the set of all operators,* $T^0_N \equiv 1/2$*, 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. 3.1, 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 3.2.1. For sources, we see that t moves strictly upwards in the ranking compared to Example 3.2.1. 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 [50] 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 $0 \leq \lambda \leq 1$.

Definition 3.2.5. *Given parameters $0 \leq \rho, \lambda \leq 1$, and $0 < \gamma < 1$,* 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*

$$T'_N(c) = \left[1 + \frac{\prod_{s \in \text{src}_N(c)} (1 - T_N(s))^\gamma}{\prod_{t \in \text{antisrc}_N(c)} (1 - T_N(t))^{\gamma\rho\lambda}}\right]^{-1}, \tag{3.1}$$

$$T'_N(s) = \sum_{c \in \text{cl}_N(s)} \frac{T'_N(c)}{|\,\text{cl}_N(s)|}. \tag{3.2}$$

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

We refer the reader to the original TruthFinder paper [50] 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 (3.1) is 1. Note that in (3.1) we have unfolded the definitions of [50] 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. 3.1. We take parameters $\rho = 0.5$ and $\gamma = 0.3$ (as per the experimental setup of Yin, Han, and Yu [50]) 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 3.2.1, and the claim ranking refines that of Example 3.2.1 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 [28]. 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 [28, §2.4.1], where each categorical value is represented as a one-hot vector, and the source weight regularisation from [28, Eq. (4)]. We make a minor modification, however, by adding a small quantity ε to α_s defined below; this ensures the argument to the logarithm in $T'_N(s)$ is non-zero and simplifies analysis of CRH later on.

Definition 3.2.6. *Given* $\varepsilon > 0$, CRH *is the recursive scheme* $(\mathcal{D}, T^0, U)$, *where* $\mathcal{D}$ *is the set of operators with* $0 \leq T_N(c) \leq 1$ *for all* N *and* $c \in C$,

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

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

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

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^{crh}.

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. 3.1 with $\varepsilon = 10^{-5}$, we have $T^{\text{crh}}_N(s) \approx 0.2577$, $T^{\text{crh}}_N(t) \approx 1.4827$ and $T^{\text{crh}}_N(u) = T^{\text{crh}}_N(v) \approx 9.3567$, 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 $T^{\text{crh}}_N(c) = T^{\text{crh}}_N(e) \approx 0.0126$, $T^{\text{crh}}_N(d) \approx 0.0725$ and $T^{\text{crh}}_N(f) \approx 0.9874$, giving the ranking $c \approx e \prec d \prec f$; this is the same as T^{sums}.

Table 3.1 summaries the source and claim rankings for each example operator on the network N from Fig. 3.1.

Table 3.1: Output rankings of the example operators on the network from Fig. 3.1.

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	$s \sqsubset t \sqsubset u \simeq v$	$c \approx e \prec d \prec f$

3.3 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 [55] 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.

3.3.1 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 3.2 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. 3.2. 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.

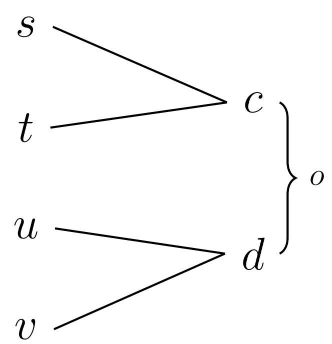


Figure 3.2: A network illustrating ***Claim-coherence****.*

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

$$\exists f : A \to B \text{ bijective s.t. } \forall x \in A : \ x \leq f(x). \tag{3.3}$$

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

If f satisfies the condition in (3.3), 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 (3.3) for which the comparison is strict.

Proposition 3.3.1. *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 3.3.1. *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. $\square$

Proof of Proposition 3.3.1. "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 3.3.1, $x \approx g(f(x))$. 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 $\sqsubseteq^T_N$, then $c \prec^T_N c'$.
>
> **Source-coherence**. If $\mathrm{cl}_N(s)$ strictly precedes $\mathrm{cl}_N(s')$ pairwise with respect to $\preceq^T_N$, then $s \sqsubset^T_N s'$.

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

Example 3.3.1. *Consider the network N from Fig. 3.1 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 3.6.

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. 3.1, 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 3.3.5, 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.]

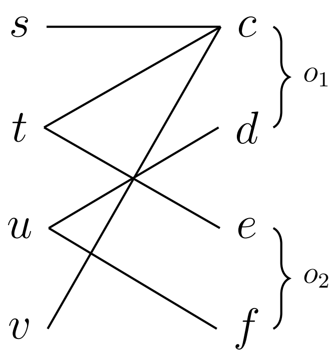


Figure 3.3: A network isomorphic to the one shown in Fig. 3.1.

3.3.2 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 3.3.2. *An* isomorphism *between networks N and N' is mapping $F : S \cup O \cup C \to S' \cup O' \cup C'$ such that*

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

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

> **Symmetry**. If F is an isomorphism between N and N', then $s \sqsubseteq_N^T s'$ iff $F(s) \sqsubseteq_{N'}^T F(s')$ and $c \preceq_N^T c'$ iff $F(c) \preceq_{N'}^T F(c')$.

We illustrate **Symmetry** with an example.

Example 3.3.2. *Consider the network N from Fig. 3.1 and N' from Fig. 3.3, 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 $(suv)(cf)(de)(o_1 o_2)$. 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. 3.1 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 3.3.2. *If F is an automorphism on N and T satisfies **Symmetry**, then $s \simeq^T_N F(s)$ and $c \approx^T_N F(c)$, for all $s \in S$ and $c \in C$.*

Proof. We show $s \simeq^T_N F(s)$ 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 3.3.1 shows there is some $k \geq 1$ such that $s = F^k(s)$.

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

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

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

Proposition 3.3.2 is useful for showing certain sources and claims must rank equally. For example, take the network N from Fig. 3.2. 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 3.3.2, any operator T satisfying **Symmetry** must output flat rankings $s \simeq t \simeq u \simeq v$ and $c \approx d$.

3.3.3 Monotonicity

Given that voting is not a viable truth discovery method, the believability of a claim c should not increase monotonically with $|\text{src}_N(c)|$. Moreover, it should not increase with the *set* of sources $\text{src}_N(c)$, ordered by set inclusion: $\text{src}_N(c) \subseteq \text{src}_N(d)$ should not in general imply $c \preceq d$. Indeed, consider an adversarial source t deliberately making false claims, and suppose $\text{src}_N(c) = \{s\}$ and $\text{src}_N(d) = \{s, t\}$. Then $\text{src}_N(c) \subseteq \text{src}_N(d)$, 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 [55]. A naive version of positive responsiveness for

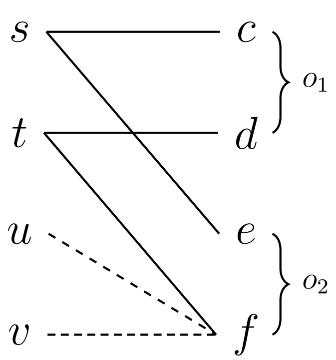


*Figure 3.4: Networks N_0 (solid edges only), $N_1 = N_0 + (u, f)$ and $N_2 = N_1 + (v, f)$ illustrating **Fresh-pos-resp** and **Source-pos-resp**.*

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 $\text{cl}_N(s) = \emptyset$ we write $N + (s, c)$ for the network $(S, O, D, R \cup \{(s, c)\})$.

> **Fresh-pos-resp**. Suppose $\text{cl}_N(s) = \emptyset$. 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.[2] 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 $s \in \text{antisrc}_N(c)$, $t \in \text{src}_N(c)$, and $\text{cl}_N(u) = \emptyset$. 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 3.3.3. *Fig. 3.4 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. 3.1. 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*

[2] 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 3.3.3. *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 3.3.3 only addresses a somewhat trivial case, it will turn out to be useful in characterising voting behaviour more generally in Sections 3.3.4 and 3.4. It can be seen as one of the many generalisations of *May's Theorem* [32], which characterises the majority voting rule in two-candidate elections. To prove it, we need a preliminary result.

Lemma 3.3.2. *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 3.3.2 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 3.3.3. 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 3.3.3. It is sufficient to show that, for all $c, d \in \mathrm{cl}_N(o)$,

$$|\,\mathrm{src}_N(c)| \leq |\,\mathrm{src}_N(d)| \implies c \preceq^T_N d \tag{3.4}$$

$$|\,\mathrm{src}_N(c)| < |\,\mathrm{src}_N(d)| \implies c \prec^T_N d. \tag{3.5}$$

First we show (3.4). 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 3.1.1 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 $0 \leq i \leq k$,

$$c \preceq^T_{N_i} d, \text{ and if } i > 0 \text{ then } c \prec^T_{N_i} d. \tag{3.6}$$

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 3.3.2 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 (3.6) 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, (3.4) follows by taking $i = k$ in (3.6), recalling that $N = N_k$. Moreover, (3.5) 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 (3.6) again. □

Trust-based monotonicity. Suppose $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$. 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 3.3.3. *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 $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$ and $c \preceq_N^T d$. Similarly, s is* marginally untrustworthy *if there are $c, d \in C$ such that $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$ 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 $\text{src}_N(d) = \text{src}_N(c) \cup Z$. 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?]

3.3.4 Independence

Another common class of axioms in social choice theory are *independence* axioms, which require that some aspect of the output is independent of "irrelevant" parts of the input. The original example is Arrow's *Independence of Irrelevant Alternatives* (IIA) in voting theory [4], which says, roughly speaking, that the ranking of candidates A and B should depend only on the individual rankings of A and B, not on any "irrelevant" alternative C. It has been adapted to several settings in which the axiomatic method has been applied. Perhaps closest to our setting is judgment aggregation, where independence requires the collective acceptance of a report φ does not depend on how the individuals accept or reject some other report ψ [15].

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. 3.1. 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 $c \preceq_N^T d$ iff $c \preceq_{N'}^T d$.

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.[3] Note that for the special case of binary networks, similar results have been shown in the literature on binary aggregation with abstentions [9].

Proposition 3.3.4. *Suppose T satisfies **Symmetry**, **Fresh-pos-resp** and **Classical-independence**. Then for all $o \in O$ and $c, d \in \mathrm{cl}_N(o)$,*

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

Proof. Take $c, d \in \mathrm{cl}_N(o)$. Let the network N' have the same sources, objects and domains as N, but with reports $R' = R \cap (S \times \{c, d\})$. That is, N' discards all reports for claims other than c and d. Then we have $\mathrm{src}_{N'}(c) = \mathrm{src}_N(c)$, $\mathrm{src}_{N'}(d) = \mathrm{src}_N(d)$, and $\mathrm{src}_{N'}(e) = \emptyset$ for all $e \notin \{c, d\}$. By **Classical-independence**, $c \preceq_N^T d$ iff $c \preceq_{N'}^T d$.

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 3.3.3 (1) to find $c \preceq_{N'}^T d$ iff $|\,\mathrm{src}_{N'}(c)| \leq |\,\mathrm{src}_{N'}(d)|$. But $|\,\mathrm{src}_{N'}(c)| = |\,\mathrm{src}_N(c)|$, and likewise for d. Consequently

$$c \preceq_N^T d \iff c \preceq_{N'}^T d \iff |\,\mathrm{src}_{N'}(c)| \leq |\,\mathrm{src}_{N'}(d)| \iff |\,\mathrm{src}_N(c)| \leq |\,\mathrm{src}_N(d)|$$

as desired. ∎

While this result appears similar to Proposition 3.3.3, 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 3.3.3 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. 3.1 (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 3.3.1. *Suppose and operator satisfies **Symmetry**, **Claim-coherence** and **Fresh-pos-resp**. Then the following axioms cannot hold simultaneously.*

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

[3] We give a further axiom which implies voting behaviour for claims of *different* objects – and leads to a complete characterisation of voting – in Section 3.4.

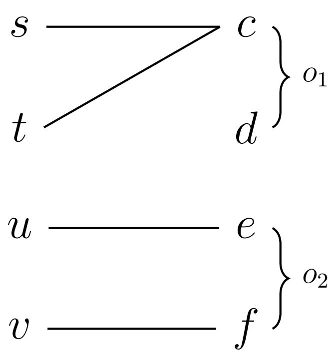


Figure 3.5: A network illustrating ***Disjoint-independence****.*

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 3.3.3 and Proposition 3.3.4, as described above.

2. Let N be as shown in Fig. 3.1. Suppose some operator T satisfies the stated axioms. From Proposition 3.3.4 we get $c \approx^T_N d$ and $e \prec^T_N f$. Considering sources s and t, **Source-coherence** gives $s \sqsubset^T_N t$. But now **Claim-coherence** gives $c \prec^T_N d$: contradiction.

□

By only looking at a claim's sources, **Classical-independence** ignores the indirect interaction with other sources and claims in the network. Our next axiom accounts for such interactions by considering networks with *disjoint sub-networks*, such as the one shown in Fig. 3.5. 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*.[4]

Definition 3.3.4. *Networks N and N' are* disjoint *if $S \cap S' = \emptyset$ and $O \cap O' = \emptyset$. For N, N' disjoint, their* union *is the network $N \sqcup N' = (S \cup S', O \cup O', \hat{D}, R \cup R')$, where $\hat{D}_o = D_o$ for $o \in O$, and $\hat{D}_o = D'_o$ for $o \in O'$.*

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

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

[4]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.

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

3.3.5 Conflicting claims

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

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

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

Theorem 3.3.2. *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. 3.6 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 $\text{src}_{N_0}(u) = \emptyset$, $s \in \text{antisrc}\, N_0(f)$ and $s' \in \text{src}_{N_0}(f)$. 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 $\text{conflict}_{N_1}(d) = \{c\}$ and $\text{conflict}_{N_1}(d') = \{c'\}$, and apply **Conflict-coherence** to get $d \prec^T_{N_1} d'$. But by **Symmetry** as above, we have $c \approx^T_{N_1} d$ and $c' \approx^T_{N_1} d'$. Consequently $c \approx^T_{N_1} d \prec^T_{N_1} d' \approx^T_{N_1} c'$, i.e. $c \prec^T_{N_1} c'$. Clearly this contradicts

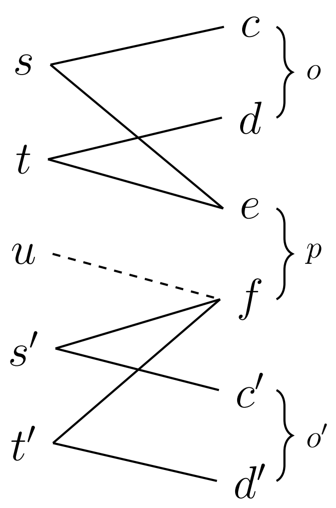


Figure 3.6: Network used to illustrate the impossibility results of Theorem 3.3.2.

$c' \prec^T_{N_1} c$. 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.

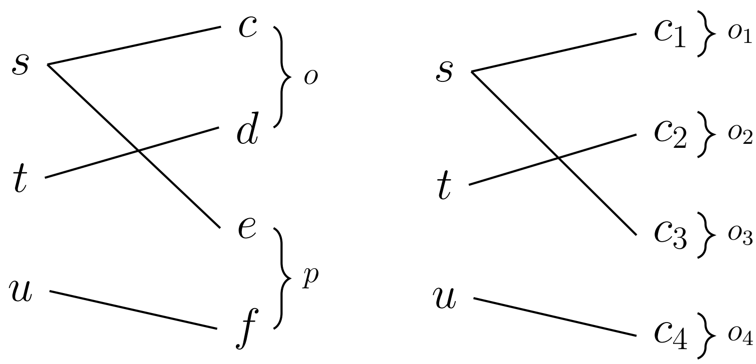
3. Suppose T additionally satisfies **Source-pos-resp** and **Anti-coherence**. Again, $s \sqsubset^T_{N_1} s'$. **Claim-coherence** implies $c \prec^T_{N_1} c'$. Since $\text{antisrc}_{N_1}(d) = \{s\}$ and $\text{antisrc}_{N_1}(d') = \{s'\}$, **Anti-coherence** gives $d' \prec^T_{N_1} d$. But recall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

3.4 A Characterisation of Voting

In the previous section we introduced general axioms for truth discovery operators. We saw in Proposition 3.3.4 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 3.2.1. In this section we give two further axioms which force this ranking even for claims across different objects, and thus



(a) N *(b) N'*

Figure 3.7: Illustration of an object reduction of a network.

characterise T^{vote} completely. Like **Classical-independence**, these axioms are *not* desirable properties, and are introduced only to capture the behaviour of voting. The first axiom simply says that the source ranking is flat.

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

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

Definition 3.4.1. *A network N' is an* object reduction *of N via $f : C_N \to C_{N'}$ if*

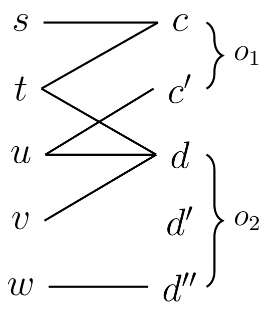
1. *$S' = S$.*
2. *f is a bijection $C_N \to C_{N'}$ such that $(s, c) \in R$ iff $(s, f(c)) \in R'$.*
3. *For al $o \in O'$, $|D'_o| = 1$.*

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. 3.7 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. 3.7 – 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 3.3.5 explicitly require that this information *does* affects the rankings. Voting does not use this information, however, which leads to the following axiom.

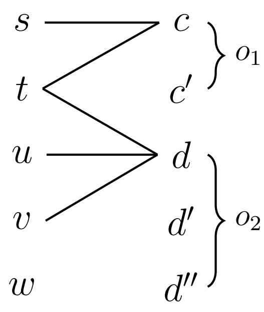
> **Object-irrelevance**. If N' is an object reduction of N via f, then $c \preceq^T_N d$ iff $f(c) \preceq^T_N f(d)$.

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.

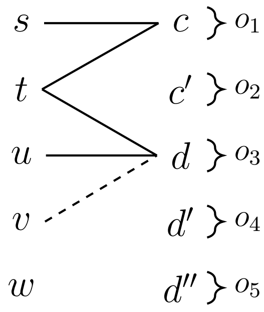
We can now characterise voting, up to ranking equivalence.



(a) N



(b) N'



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

*Figure 3.8: Illustration of the proof of Theorem 3.4.1. 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 3.4.1. *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 3.3.4 and only sketch the argument here. An illustration of the proof is shown in Fig. 3.8.

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,[5] and rankings are again preserved by **Object-irrelevance**. As before, it suffices to show that $|\operatorname{src}_N(c)| \leq |\operatorname{src}_N(d)|$ implies $c \preceq_N^T d$ and $|\operatorname{src}_N(c)| < |\operatorname{src}_N(d)|$ implies $c \prec_N^T d$, since c and d are arbitrary.

Write $k = |\operatorname{src}_N(d)| - |\operatorname{src}_N(c)| \geq 0$. Choosing k sources from $\operatorname{src}_N(d) \setminus \operatorname{src}_N(c)$, let N_0 be the network obtained from N'' in which reports for d from these sources are removed. Note that such sources *only* report d, since reports for other claims were removed in the construction of N'. Then $|\operatorname{src}_{N_0}(c)| = |\operatorname{src}_{N_0}(d)|$. The fact that $|D''_{\text{obj}(c)}| = |D''_{\text{obj}(d)}| = 1$ ensures we are able to choose an automorphism on N_0

[5]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.

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 $w_N(s) \in \{0, 1\}$.]

3.5 Fixed-points for Recursive Operators

Sketch of this section:

- For a recursive operator R, we might not have a closed-form expression for the limit operator T^*.
- But hopefully, the operator converges to a fixed point of U, so that $U(T^*) = T^*$.
- This will simplify analysis of recursive operators (e.g. to see which axioms hold).
- Fixing N, an operator gives us a function $S_N \cup C_N \to \mathbb{R}$. Let $\mathcal{F}_N$ be the set of such functions.
- Basic compatibility condition for a recursive operator: if $T_N = T'_N$ then $U(T)_N = U(T')_N$.
- This implies that U defines a function $\mathcal{F}_N \to \mathcal{F}_N$ (which we also denote by U).
- $\mathcal{F}_N$ is a metric space with the uniform distance

$$d(f, g) = \|f - g\|_\infty = \max_{z \in S \cup C} |f(z) - g(z)|.$$

- Can use results for convergence in metric spaces
- If R converges and $\mathcal{D}$ is closed, then $T^* \in \mathcal{D}$ (i.e. the limit is in the domain of the update function)
- Since $S \cup C$ is finite, pointwise convergence (i.e. our definition) implies uniform convergence
- If U is continuous and R converges, then $U(T^*) = T^*$.
- Banach's fixed-point theorem: if U is a contraction mapping, then U is continuous and R converges to a fixed point. Moreover, T_0 can be chosen arbitrarily.
- We can try these for each of our example operators.
- Show which operators are weightable, using the fixed-point properties.

Having introduced several example operators in Section 3.2 and axioms in Section 3.3, it is natural to ask which axioms are satisfied by each operator. To aid this process, in this section we first collect some results for the limit T^* of a recursive scheme $(\mathcal{D}, T_0, U)$ to be a *fixed-point* of U, i.e. $U(T^*) = T^*$. This property greatly simplifies analysis of (some of) the axioms, especially for recursive operators for which no closed-form expression is available for T^*. In what follows, when a recursive scheme is understood we write T^n for $U^n(T_0)$, i.e. the n-th iteration of the scheme,

The strategy is as follows: we first define a distance metric d on the of operators, so that convergence in the sense of Definition 3.2.3 – i.e. $T_N^n(z) \to T_N^*(z)$ for all networks N and $z \in S_N \cup C_N$ – coincides with $T^n \to T^*$ in the sense of convergence in a metric space – i.e. $d(T^*, T^n) \to 0$. We may then speak of $U : \mathcal{D} \to \mathcal{D}$ being *continuous*, and a basic result from topology gives that $T^n \to T^*$ implies $U(T^n) \to U(T^*)$. Since $U(T^n) = T^{n+1}$ and limits are unique in a metric space, this implies $T^* = U(T^*)$ as desired.

For simplicity we work with *bounded operators*: T is *bounded* if there is $M \geq 0$ such that for all networks N and $z \in S_N \cup C_N$, $|T_N(z)| \leq M$. Note that this is equivalent to T bounded in the conventional sense when viewed as a function $\bigsqcup_N (S_N \cup C_N) \to \mathbb{R}$. A recursive scheme $(\mathcal{D}, T_0, U)$ is bounded if each operator in $\mathcal{D}$ is. Write $\mathcal{T}$ for the set of bounded operators. The limit of a bounded recursive scheme is again bounded.

Proposition 3.5.1. *If $(\mathcal{D}, T_0, U)$ is bounded and converges to T^*, then T^* is bounded.*

Proof. **[TODO:** Write.] □

Now, for sets X, Y, the *uniform norm* $\|f\|_\infty = \sup_{x \in X} |f(x)|$ is a norm on the set of bounded functions $f : X \to Y$, which in turn defines a metric $d(f, g) = \|f - g\|_\infty$. Applying this to bounded operators yields the following metric $d : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$:[6]

$$d(T, T') = \sup_N \max_{z \in S_N \cup C_N} |T_N(z) - T'_N(z)|.$$

Proposition 3.5.2. *A bounded recursive scheme $(\mathcal{D}, T_0, U)$ converges to T^* if and only if T^n converges to T^* in $(\mathcal{T}, d)$.*

Proof. **[TODO:** Write.] □

For T^* to be a fixed-point of U, clearly T^* needs to lie in the domain $\mathcal{D}$ of U. A sufficient condition is for $\mathcal{D}$ to be *closed* as a subset of $\mathcal{T}$, due to the following standard result.

Lemma 3.5.1. *Let (X, d) be a metric space. Then $A \subseteq X$ is closed[7] if and only if it contains all its limit points, i.e. $x_n \to x^*$ implies $x^* \in A$, for all convergent sequences $(x_n)_{n \in \mathbb{N}}$ in A.*

[TODO: Citation for proof, or just reproduce it here.]

The following definition of a continuous map between metric spaces is standard.

[6] Note that the boundedness assumption ensures the supremum is finite.

[7] A set $A \subseteq X$ is *open* if for all $x \in A$, there is $r > 0$ such that $\{x' \in X \mid d(x, x') < r\} \subseteq A$. A is *closed* if its complement $X \setminus A$ is open.

Definition 3.5.1. *A function $f : X \to Y$ between metric spaces (X, d_X), (Y, d_Y) is* continuous *if for all $x \in X$ and $\varepsilon > 0$ there is $\delta > 0$ such that for all $x' \in X$,*

$$d_X(x, x') < \delta \implies d_Y(f(x), f(x')) < \varepsilon.$$

Fixing a recursive scheme $(\mathcal{D}, T_0, U)$, restricting d to $\mathcal{D}$ gives a metric space structure on $\mathcal{D}$, and we may therefore speak of $U : \mathcal{D} \to \mathcal{D}$ being continuous. By another standard result, we may interchange the application of a continuous function with the limit.

Lemma 3.5.2. *Let $f : X \to Y$ be continuous and $(x_n)_{n \in \mathbb{N}}$ a convergent sequence in X. Then $\lim_{n\to\infty} f(x_n) = f(\lim_{n\to\infty} x_n)$.*

[TODO: Citation for proof, or just reproduce it here.]

Putting things together, we arrive at the following.

Proposition 3.5.3. *Let $(\mathcal{D}, T_0, U)$ be a bounded recursive scheme with limit T^*. If $\mathcal{D}$ is closed and U is continuous, then $U(T^*) = T^*$.*

Our aim is to now show that Proposition 3.5.3 can be applied to each of the example recursive operators from Section 3.2.2. We take each operator in turn.

Sums. Let $(\mathcal{D}, T_0, U)$ be as defined for Sums in Definition 3.2.4. It will prove more convenient to work with the following equivalent expression for U.

Lemma 3.5.3. *Define $U_1, U_2 : \mathcal{D} \to \mathcal{D}$ by*

$$\begin{aligned} U_1(T)_N(s) &= \sum_{c \in \mathrm{cl}_N(s)} T_N(c), \\ U_1(T)_N(c) &= \sum_{s \in \mathrm{src}_N(c)} U_1(T)_N(s) \end{aligned}$$

and

$$\begin{aligned} U_1(T)_N(s) &= \alpha T_N(s) \\ U_1(T)_N(c) &= \beta T_N(c), \end{aligned}$$

where

$$\alpha = \begin{cases} \frac{1}{\max_{t \in S_N} |T_N(t)|}, & \max_{t \in S_N} |T_N(t)| \neq 0 \\ 0, & \textit{otherwise} \end{cases},$$

$$\beta = \begin{cases} \frac{1}{\max_{d \in C_N} |T_N(d)|}, & \max_{d \in C_N} |T_N(d)| \neq 0 \\ 0, & \textit{otherwise} \end{cases}.$$

Then $U = U_2 \circ U_1$.

TruthFinder. **[TODO:** Write.]

CRH. **[TODO:** Write.]

3.6 Satisfaction of the Axioms

Go through example operators. Possibly give weakened coherence/conflict-coherence axioms satisfied by TruthFinder and CRH, upon showing regular **Claim-coherence** fails.

3.6.1 Modifying Sums

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

3.7 Related Work

3.8 Conclusion

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 4, 10, 34–36, 40, 46).

[2] Alon Altman and Moshe Tennenholtz. "Ranking systems: the PageRank axioms". In: *Proceedings of the 6th ACM conference on Electronic commerce*. ACM. 2005, pp. 1–8 (cited on pages 4, 36).

[3] Reid Andersen et al. "Trust-based Recommendation Systems: An Axiomatic Approach". In: *Proceedings of the 17th International Conference on World Wide Web*. WWW '08. New York, NY, USA: ACM, 2008, pp. 199–208. ISBN: 978-1-60558-085-2. DOI: 10.1145/1367497.1367525. URL: http://doi.acm.org/10.1145/1367497.1367525 (cited on page 4).

[4] Kenneth J. Arrow. "Social Choice and Individual Values". In: *Ethics* 62.3 (1952), pp. 220–222 (cited on pages 10, 54).

[5] Sheldon Axler. *Linear Algebra Done Right*. Springer International Publishing, Nov. 2014. DOI: 10.1007/978-3-319-11080-6. URL: https://doi.org/10.1007/978-3-319-11080-6 (cited on page 26).

[6] Raju Balakrishnan and Subbarao Kambhampati. "Sourcerank: Relevance and trust assessment for deep web sources based on inter-source agreement". In: *Proceedings of the 20th international conference on World wide web*. 2011, pp. 227–236 (cited on page 12).

[7] Laure Berti-Equille and Javier Borge-Holthoefer. "Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics". In: *Synthesis Lectures on Data Management* 7.3 (2015), pp. 1–155 (cited on page 3).

[8] Felix Brandt et al. "Introduction to Computational Social Choice". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 1 (cited on pages 10, 11).

[9] Zoé Christoff and Davide Grossi. "Binary Voting with Delegable Proxy: An Analysis of Liquid Democracy". In: *Proc. TARK 2017*. 2017 (cited on pages 6, 13, 39, 55).

[10] 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 pages 7, 40).

[11] Elad Dokow and Ron Holzman. “Aggregation of binary evaluations with abstentions”. In: *Journal of Economic Theory* 145 (2010), pp. 544–561 (cited on pages 6, 39).

[12] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. “Integrating conflicting data: the role of source dependence”. In: *Proceedings of the VLDB Endowment* 2.1 (2009), pp. 550–561 (cited on page 12).

[13] 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 pages 12, 38).

[14] Yang Du et al. “Bayesian Co-Clustering Truth Discovery for Mobile Crowd Sensing Systems”. In: *IEEE Transactions on Industrial Informatics* (2019), pp. 1–1. ISSN: 1551-3203. DOI: 10.1109/TII.2019.2896287 (cited on pages 3, 9, 43).

[15] 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, 6, 10, 39, 54).

[16] Patricia Everaere, Sébastien Konieczny, Pierre Marquis, et al. “The Epistemic View of Belief Merging: Can We Track the Truth?.” In: *ECAI*. 2010, pp. 621–626 (cited on page 37).

[17] 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 4, 6, 9, 40, 43).

[18] Arpita Ghosh, Satyen Kale, and Preston McAfee. “Who moderates the moderators?” In: *Proceedings of the 12th ACM conference on Electronic commerce - EC 11*. ACM Press, 2011. DOI: 10.1145/1993574.1993599. URL: https://doi.org/10.1145%2F1993574.1993599 (cited on page 4).

[19] Manish Gupta and Jiawei Han. “Heterogeneous Network-based Trust Analysis: A Survey”. In: *SIGKDD Explor. Newsl.* 13.1 (Aug. 2011), pp. 54–71. ISSN: 1931-0145. DOI: 10.1145/2031331.2031341. URL: http://doi.acm.org/10.1145/2031331.2031341 (cited on pages 3, 4, 6).

[20] Stephan Hartmann and Jan Sprenger. “Judgment aggregation and the problem of tracking the truth”. In: *Synthese* 187.1 (July 2012), pp. 209–221. ISSN: 1573-0964. DOI: 10.1007/s11229-011-0031-5. URL: https://doi.org/10.1007/s11229-011-0031-5 (cited on page 37).

[21] 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, 36, 43).

[22] Sébastien Konieczny and Ramón Pino Pérez. “Improvement Operators”. In: *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*. Ed. by Gerhard Brewka and Jérôme Lang. AAAI Press, 2008, pp. 177–187. URL: http://www.aaai.org/Library/KR/2008/kr08-018.php (cited on page 31).

[23] Sébastien Konieczny and Ramón Pino Pérez. "Merging information under constraints: a logical framework". In: *Journal of Logic and computation* 12.5 (2002), pp. 773–808 (cited on page 31).

[24] Neema Kotonya and Francesca Toni. "Explainable Automated Fact-Checking for Public Health Claims". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 7740–7754 (cited on page 4).

[25] Justin Kruger et al. "Axiomatic Analysis of Aggregation Methods for Collective Annotation". In: *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*. AAMAS '14. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 1185–1192. ISBN: 978-1-4503-2738-1. URL: http://dl.acm.org/citation.cfm?id=2617388.2617437 (cited on page 12).

[26] Jean-François Laslier and M Remzi Sanver. *Handbook on approval voting*. Springer Science & Business Media, 2010 (cited on page 34).

[27] Yaliang Li et al. "A Survey on Truth Discovery". In: *SIGKDD Explor. Newsl.* 17.2 (2016), pp. 1–16. ISSN: 1931-0145. DOI: 10.1145/2897350.2897352. URL: http://doi.acm.org/10.1145/2897350.2897352 (cited on pages 3, 42).

[28] 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 4, 7, 9, 38, 40–43, 45).

[29] Fenglong Ma et al. "FaitCrowd: Fine Grained Truth Discovery for Crowdsourced Data Aggregation". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. event-place: Sydney, NSW, Australia. New York, NY, USA: ACM, 2015, pp. 745–754. ISBN: 978-1-4503-3664-2. DOI: 10.1145/2783258.2783314. URL: http://doi.acm.org/10.1145/2783258.2783314 (cited on page 3).

[30] Fenglong Ma et al. "Unsupervised Discovery of Drug Side-Effects from Heterogeneous Data Sources". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. Halifax, NS, Canada: ACM, 2017, pp. 967–976. ISBN: 978-1-4503-4887-4. DOI: 10.1145/3097983.3098129. URL: http://doi.acm.org/10.1145/3097983.3098129 (cited on page 3).

[31] Jermaine Marshall, Arturo Argueta, and Dong Wang. "A neural network approach for truth discovery in social sensing". In: *2017 IEEE 14th international conference on mobile Ad Hoc and sensor systems (MASS)*. IEEE. 2017, pp. 343–347 (cited on page 4).

[32] 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 52).

[33] Reshef Meir et al. "Truth Discovery via Proxy Voting". In: *CoRR* abs/1905.00629 (2019). arXiv: 1905.00629. URL: http://arxiv.org/abs/1905.00629 (cited on page 37).

[34] Lawrence Page et al. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66. Stanford InfoLab, Nov. 1999. URL: http://ilpubs.stanford.edu:8090/422/ (cited on page 36).

[35] 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 4, 6–9, 38, 40, 41, 43).

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

[37] 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).

[38] 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).

[39] 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).

[40] Moshe Tennenholtz. "Reputation Systems: An Axiomatic Approach". In: *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*. UAI '04. Banff, Canada: AUAI Press, 2004, pp. 544–551. ISBN: 0-9749039-0-6. URL: http://dl.acm.org/citation.cfm?id=1036843.1036909 (cited on pages 4, 10, 36).

[41] Dalia Attia Waguih and Laure Berti-Equille. "Truth discovery algorithms: An experimental evaluation". In: *arXiv preprint arXiv:1409.6428* (2014) (cited on page 4).

[42] Dong Wang et al. "On Truth Discovery in Social Sensing: A Maximum Likelihood Estimation Approach". In: *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*. IPSN '12. event-place: Beijing, China. New York, NY, USA: ACM, 2012, pp. 233–244. ISBN: 978-1-4503-1227-1. DOI: 10.1145/2185677.2185737. URL: http://doi.acm.org/10.1145/2185677.2185737 (cited on pages 3, 38).

[43] Yaqing Wang et al. "Eann: Event adversarial neural networks for multi-modal fake news detection". In: *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*. 2018, pp. 849–857 (cited on page 4).

[44] Tomasz Wąs and Oskar Skibski. *Axiomatic Characterization of PageRank*. 2020. DOI: 10.48550/ARXIV.2010.08487. URL: https://arxiv.org/abs/2010.08487 (cited on page 36).

[45] Houping Xiao. "Multi-sourced Information Trustworthiness Analysis: Applications and Theory". PhD thesis. University at Buffalo, State University of New York, 2018 (cited on page 4).

[46] 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 42).

[47] 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 pages 4, 38).

[48] Yi Yang, Quan Bai, and Qing Liu. "A probabilistic model for truth discovery with object correlations". In: *Knowledge-Based Systems* 165 (2019), pp. 360–373. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2018.12.004. URL: http://www.sciencedirect.com/science/article/pii/S0950705118305914 (cited on pages 9, 37, 43).

[49] 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 pages 7, 40).

[50] 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 4, 6, 9, 38, 40, 41, 43, 44).

[51] Xiaoxin Yin and Wenzhao Tan. "Semi-supervised Truth Discovery". In: *Proceedings of the 20th International Conference on World Wide Web*. WWW '11. event-place: Hyderabad, India. New York, NY, USA: ACM, 2011, pp. 217–226. ISBN: 978-1-4503-0632-4. DOI: 10.1145/1963405.1963439. URL: http://doi.acm.org/10.1145/1963405.1963439 (cited on page 4).

[52] Daniel Yue Zhang et al. "On robust truth discovery in sparse social media sensing". In: *2016 IEEE International Conference on Big Data (Big Data)*. 2016-12, pp. 1076–1081. DOI: 10.1109/BigData.2016.7840710 (cited on pages 3, 6, 40).

[53] 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 6, 9, 38, 40, 43).

[54] 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 6, 9, 40, 43).

[55] 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 4, 11, 46, 50).

A Proofs for Chapter 2

A.1 Proof of Theorem 2.3.1

The following lemma is required before the proof.

Lemma A.1.1. *Suppose a network $N = (V, E)$ contains claims only for a single object $o \in \mathcal{O}$; that is, there exists $o \in \mathcal{O}$ such that $(s, f) \in E$ implies $obj_N(f) = o$ for all $s \in \mathcal{S}, f \in \mathcal{F}$. Then for any Symmetric operator T and $f_1, f_2 \in \mathcal{F}$, $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)| > 0$ implies $f_1 \approx_N^T f_2$.*

Proof. Suppose N has the stated property, T satisfies symmetry, and $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)| > 0$. Then there is a bijection $\varphi : \mathsf{src}_N(f_1) \to \mathsf{src}_N(f_2)$. Note that since f_1 and f_2 are for the same object no source can claim both facts, i.e. $\mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2) = \emptyset$.

Define a permutation π by

$$\pi(s) = \begin{cases} \varphi(s) & \text{if } s \in \mathsf{src}_N(f_1) \\ \varphi^{-1}(s) & \text{if } s \in \mathsf{src}_N(f_2) \\ s & \text{otherwise} \end{cases}$$

$$\pi(f) = \begin{cases} f_2 & \text{if } f = f_1 \\ f_2 & \text{if } f = f_2 \\ f & \text{otherwise} \end{cases}$$

and $\pi(o) = o$ for all $o \in \mathcal{O}$. That is, π swaps facts f_1 and f_2, and swaps the sources of f_1 with their counterparts in f_2. Note that $\pi = \pi^{-1}$.

Write $N' = \pi(N)$. We claim that $N' = N$. Write E, E' for the edges in N and N' respectively. First we will show $E \subseteq E'$. Suppose $(s, f) \in E$. There are three cases.

Case 1: $f = f_1$. Here we have $(s, f_1) \in E$, so $s \in \mathsf{src}_N(f_1)$. Consequently $\pi(s) = \varphi(s) \in \mathsf{src}_N(f_2)$, i.e. $(\pi(s), f_2) \in E$. By the definition of a graph isomorphism we get $(\pi(\pi(s)), \pi(f_2)) \in E'$. Noting that $\pi(f_2) = f_1 = f$ and $\pi(\pi(s)) = s$ (since $\pi = \pi^{-1}$), we have $(s, f) \in E'$ as desired.

Case 2: $f = f_2$. Similar to the above case, here we have $s \in \mathsf{src}_N(f_2)$ and so $\pi(s) = \varphi^{-1}(s) \in \mathsf{src}_N(f_1)$, i.e. $(\pi(s), f_1) \in E$. As before, applying the definition of a graph isomorphism and using $\pi = \pi^{-1}$, we get $(s, f) \in E'$.

Case 3: $f \notin \{f_1, f_2\}$. By hypothesis f relates to the same object as f_1 and f_2. This means $s \notin \mathsf{src}_N(f_1)$ and $s \notin \mathsf{src}_N(f_2)$, since otherwise s would make claims for multiple facts for a single object. Hence we have $\pi(s) = s$ and $\pi(f) = f$. This means $(s, f) = (\pi(s), \pi(f)) \in E'$ as required.

To complete the claim $E \subseteq E'$, suppose $(f, o) \in E$. There are again three cases: $f = f_1$, $f = f_2$, or $f \notin \{f_1, f_2\}$. In each case the definition of π and $\pi(N)$ easily yield $(f, o) \in E'$. Hence $E \subseteq E'$.

Now for the reverse direction: we must show $E' \subseteq E$. Let $(x, y) \in E'$. By definition of a graph isomorphism, we have $(\pi^{-1}(x), \pi^{-1}(y)) \in E$. Using $\pi^{-1} = \pi$ and the first part we get $(\pi(x), \pi(y)) = (\pi^{-1}(x), \pi^{-1}(y)) \in E \subseteq E'$. The definition of a graph isomorphism then yields $(x, y) \in E$ and so $E' \subseteq E$. Hence $E = E'$ and $N = N'$.

To conclude the proof, we apply Symmetry of T to get

$$\begin{aligned} f_1 \preceq^T_N f_2 &\iff \pi(f_1) \preceq^T_{N'} \pi(f_2) \\ &\iff f_2 \preceq^T_{N'} f_1 \\ &\iff f_2 \preceq^T_N f_1 \end{aligned}$$

and so $f_1 \approx^T_N f_2$ as required. □

Proof of Theorem 2.3.1. Suppose T is an operator satisfying Symmetry, Monotonicity and POI. Let $N \in \mathcal{N}$, $o \in \mathcal{O}$ and $f_1, f_2 \in \mathsf{obj}^{-1}_N(o)$. We need to show that $f_1 \preceq^T_N f_2$ iff $|\mathsf{src}_N(f_1)| \leq |\mathsf{src}_N(f_2)|$.

Let N' be the network obtained from N by removing all claims for facts other than those for object o; that is, $N' = (V, E')$ where E is the set of edges in N and

$$E' = (E \cap (\mathcal{S} \times \mathsf{obj}^{-1}_N(o))) \cup (E \cap (\mathcal{F} \times \mathcal{O}))$$

Note that the fact-object affiliations are the same in N' as in N, and the set of sources for each fact in $\mathsf{obj}^{-1}_N(o)$ is the same. Therefore POI applies, and it is sufficient to show that $f_1 \preceq^T_{N'} f_2$ iff $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$.

First suppose $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. If $|\mathsf{src}_{N'}(f_1)| = |\mathsf{src}_{N'}(f_2)|$, then we have $f_1 \approx^T_{N'} f_2$ by Symmetry and Lemma A.1.1; in particular $f_1 \preceq^T_{N'} f_2$. Otherwise $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Removing k sources from f_2 to obtain a new network N'', we can apply the lemma to get $f_1 \approx^T_{N''} f_2$. We may then add these sources *back* to obtain N' again; k applications of Monotonicity then give $f_1 \prec^T_{N'} f_2$ as required.

To complete the proof we show that $f_1 \preceq^T_{N'} f_2$ implies $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. Indeed, suppose $f_1 \preceq^T_{N'} f_2$ but $|\mathsf{src}_{N'}(f_1) > |\mathsf{src}_{N'}(f_2)|$. Then the argument above gives $f_1 \succ^T_{N'} f_2$, which is clearly a contradiction. Hence the proof is complete. □

A.2 Proof of Theorem 2.3.3

The proof of this theorem is similar in spirit to that of Theorem 2.3.1. Like in that case, a preliminary result is required first.

Lemma A.2.1. *Let N be a network and $f_1, f_2 \in \mathcal{F}$. Write $o_1 = \mathsf{obj}_N(f_1)$, $o_2 = \mathsf{obj}_N(f_2)$. Suppose N has the following properties:*

1. *There is $o^* \in \mathcal{O} \setminus \{o_1, o_2\}$ such that $f \in \mathcal{F} \setminus \{f_1, f_2\} \implies \mathsf{obj}_N(f) = o^*$; and*

2. $\mathsf{src}_N(f) = \emptyset$ *for all* $f \in \mathcal{F} \setminus \{f_1, f_2\}$.

Then for any operator T *satisfying Symmetry,* $|\mathsf{src}_N(f_1)| = |\mathsf{src}_N(f_2)|$ *implies* $f_1 \approx_N^T f_2$.

Proof. The proof is similar to that of Lemma A.1.1. Suppose $|src_N(f_1)| = |\mathsf{src}_N(f_2)|$. Write

$$Q_1 = \mathsf{src}_N(f_1) \setminus \mathsf{src}_N(f_2)$$
$$Q_2 = \mathsf{src}_N(f_2) \setminus \mathsf{src}_N(f_1)$$

Then $|Q_1| = |Q_2|$, so there exists a bijection $\varphi : Q_1 \to Q_2$. Define a permutation π as follows:

$$\pi(s) = \begin{cases} \varphi(s) & \text{if } s \in Q_1 \\ \varphi^{-1}(s) & \text{if } s \in Q_2 \\ s & \text{otherwise} \end{cases}$$

$$\pi(f) = \begin{cases} f_2 & \text{if } f = f_1 \\ f_1 & \text{if } f = f_2 \\ f & \text{otherwise} \end{cases}$$

$$\pi(o) = \begin{cases} o_2 & \text{if } o = o_1 \\ o_1 & \text{if } o = o_2 \\ o & \text{otherwise} \end{cases}$$

That is, π swaps f_1 and f_2, swaps o_1 and o_2, and swaps sources in Q_1 with their counterparts in Q_2. Note that $\pi = \pi^{-1}$. Write $N' = \pi(N)$. We claim that $N' = N$. Write E, E' for the edges in N and N' respectively. First we show that $E \subseteq E'$. For this, first suppose $(s, f) \in E$ for some $s \in \mathcal{S}$, $f \in \mathcal{F}$. By definition of E, either $f = f_1$ or $f = f_2$.

Case 1: $f = f_1$. Here $\pi(f) = f_2$, and we have either $s \in Q_1$ or $s \in \mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2)$. In the first case, $\pi(s) = \varphi(s) \in Q_2 \subseteq \mathsf{src}_N(f_2) = \mathsf{src}_N(\pi(f))$. In the second case $\pi(s) = s \in \mathsf{src}_N(f_2) = \mathsf{src}_N(\pi(f))$. In either case, $(\pi(s), \pi(f)) \in E$.

Applying the definition of a graph isomorphism we get $(\pi(\pi(s)), \pi(\pi(f))) \in E'$. But $\pi = \pi^{-1}$, so this means $(s, f) \in E'$ as desired.

Case 2: $f = f_2$. This case is similar. Here $\pi(f) = f_1$. If $s \in Q_2$, then $\pi(s) = \varphi^{-1}(s) \in Q_1 \subseteq \mathsf{src}_N(f_1) = \mathsf{src}_N(\pi(f))$. Otherwise $s \in \mathsf{src}_N(f_1) \cap \mathsf{src}_N(f_2)$ and $\pi(s) = s \in \mathsf{src}_N(f_1) = \mathsf{src}_N(\pi(f))$. Again, we have $(\pi(s), \pi(f)) \in E$ in either case, so $(s, f) \in E'$.

Note that these two cases cover all possibilities since by hypothesis $\mathsf{src}_N(f) = \emptyset$ if $f \notin \{f_1, f_2\}$.

Next, suppose $(f, o) \in E$. If $f = f_1$ then $o = o_1$, so $(\pi(f), \pi(o)) = (f_2, o_2) \in E$. Similarly if $f = f_2$ then $o = o_2$ and $(\pi(f), \pi(o)) = (f_1, o_1) \in E$. If $f \notin \{f_1, f_2\}$ then $\pi(f) = f$ and $o = o^*$, so $\pi(o) = o$. We see that in all cases, $(\pi(f), \pi(f)) \in E$. Applying the same argument as in case 1 above, we see that $(f, o) \in E'$. This shows $E \subseteq E'$.

To complete the claim that $N = N'$ we must show $E' \subseteq E$. This can be shown using the same argument used in Lemma A.1.1. Indeed, suppose $(x, y) \in E'$. Then by definition of a graph isomorphism, $(\pi^{-1}(x), \pi^{-1}(y)) \in E$. Using the fact that $\pi = \pi^{-1}$ and $E \subseteq E'$ we get $(\pi(x), \pi(y)) \in E'$, which then yields $(x, y) \in E$ as required. Hence $E = E'$ and $N = N'$.

Finally, using Symmetry of T we have

$$\begin{aligned} f_1 \preceq^T_N f_2 &\iff \pi(f_1) \preceq^T_{\pi(N)} \pi(f_2) \\ &\iff f_2 \preceq^T_{N'} f_1 \\ &\iff f_2 \preceq^T_N f_1 \end{aligned}$$

Consequently $f_1 \approx^T_N f_2$. □

Proof of Theorem 2.3.3. The 'if' direction is clear since *Voting* satisfies Strong Independence, Monotonicity and Symmetry (see Theorem 2.4.1). For the other direction, suppose T satisfies the stated axioms. Let N be a network and $f_1, f_2 \in \mathcal{F}$. We will construct a network N' where all claims for facts other than f_1, f_2 are removed, and these facts are grouped under a single object. To do so, let $o_1 = \mathsf{obj}_N(f_1)$, $o_2 = \mathsf{obj}_N(f_2)$ and take $o^* \in \mathcal{O} \setminus \{o_1, o_2\}$. Define an edge set E' by

$$(s, f) \in E' \iff f \in \{f_1, f_2\} \text{ and } s \in \mathsf{src}_N(f)$$

$$(f, o) \in E' \iff (f \in \{f_1, f_2\} \text{ and } o = \mathsf{obj}_N(f)) \text{ or } (f \notin \{f_1, f_2\} \text{ and } o = o^*)$$

Then let N' be the network with edge set E'. Note that $\mathsf{src}_{N'}(f_j) = \mathsf{src}_N(f_j)$. By Strong Independence it is therefore sufficient to show that $f_1 \preceq^T_{N'} f_2$ iff $|src_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. Note also that N' satisfies the hypothesis of Lemma A.2.1.

Now, suppose $|\mathsf{src}_{N'}(f_1)| \leq |\mathsf{src}_{N'}(f_2)|$. If $|\mathsf{src}_{N'}(f_1)| = |\mathsf{src}_{N'}(f_2)|$ then by Lemma A.2.1 $f_1 \approx^T_{N'} f_2$, and in particular $f_1 \preceq^T_{N'} f_2$.

Otherwise, $|\mathsf{src}_{N'}(f_2)| - |\mathsf{src}_{N'}(f_1)| = k > 0$. Consider N'' where k sources from $\mathsf{src}_{N'}(f_2)$ are removed, and all other claims remain. By the lemma, $f_1 \approx^T_{N''} f_2$. Applying Monotonicity k times we can produce N' from N'' and get $f_1 \prec^T_{N'} f_2$ as desired.

For the other implication, suppose $f_1 \preceq^T_{N'} f_2$ and, for contradiction, $|\mathsf{src}_{N'}(f_1)| > |\mathsf{src}_{N'}(f_2)|$. Applying Monotonicity again as above gives $f_1 \succ^T_{N'} f_2$ and the required contradiction. □

A.3 Proof of Theorem 2.4.1

Proof. We will show that *Voting* satisfies Symmetry, Unanimity, Groundedness, Monotonicity, POI, Strong Independence and PCI, and that Coherence is *not* satisfied. For Symmetry and PCI we use the (stronger) numerical variants *numerical Symmetry* and *numerical PCI*, introduced in Section 2.4.2. T will denote the (numerical) *Voting* operator in what follows.

Symmetry. Suppose N and $\pi(N)$ are equivalent networks. Let $f \in \mathcal{F}$. By definition of equivalent networks we have $s \in \mathsf{src}_N(f)$ iff $\pi(s) \in \mathsf{src}_{\pi(N)}(\pi(f))$ for all $s \in \mathcal{S}$. Consequently π restricted to $\mathsf{src}_N(f)$ is a bijection into $\mathsf{src}_{\pi(N)}(\pi(f))$, and hence

$$T_N(f) = |\mathsf{src}_N(f)| = |\mathsf{src}_{\pi(N)}(\pi(f))| = T_{\pi(N)}(\pi(f))$$

Now let $s \in \mathcal{S}$. Clearly we have $T_N(s) = 1 = T_{\pi(N)}(\pi(s))$. Hence T satisfies numerical Symmetry and therefore Symmetry.

Unanimity and Groundedness. Suppose $N \in \mathcal{N}$ and $f \in \mathcal{F}$. If $\mathsf{src}_N(f) = \mathcal{S}$ then for any $g \in \mathcal{F}$,

$$T_N(g) = |\mathsf{src}_N(g)| \leq |\mathcal{S}| = |\mathsf{src}_N(f)| = T_N(f)$$

so $g \preceq_N^T f$ and Unanimity is satisfied. If instead $\mathsf{src}_N(f) = \emptyset$, we have

$$T_N(g) = |\mathsf{src}_N(g)| \geq 0 = |\mathsf{src}_N(f)| = T_N(f)$$

so $f \preceq_N^T g$ and Groundedness is satisfied.

Monotonicity. Let N, N', s and f be as given in the statement of Monotonicity. It is clear that $|\mathsf{src}_{N'}(f)| = |\mathsf{src}_N(f)| + 1$. Also, for any $g \in \mathcal{F}$, $g \neq f$, the set of sources in N' is the same as in N but with s possibly removed. Hence $|\mathsf{src}_{N'}(g)| \leq |\mathsf{src}_N(g)$. Therefore $g \preceq_N^T f$ implies

$$|\mathsf{src}_{N'}(g)| \leq |\mathsf{src}_N(g)| \leq |\mathsf{src}_N(f)| < |\mathsf{src}_{N'}(f)|$$

and so $g \prec_{N'}^T f$ as required.

Independence axioms. Next we show Strong Independence, which implies POI. Suppose $N_1, N_2 \in \mathcal{N}$, $f_1, f_2 \in \mathcal{F}$ and $\mathsf{src}_{N_1}(f_j) = \mathsf{src}_{N_2}(f_j)$ for each $j \in \{1, 2\}$. Clearly we have

$$T_{N_1}(f_j) = |\mathsf{src}_{N_1}(f_j)| = |\mathsf{src}_{N_2}(f_j)| = T_{N_2}(f_j)$$

Consequently

$$\begin{aligned} f_1 \preceq_{N_1}^T f_2 &\iff T_{N_1}(f_1) \leq T_{N_1}(f_2) \\ &\iff T_{N_2}(f_1) \leq T_{N_2}(f_2) \\ &\iff f_1 \preceq_{N_2}^T f_2 \end{aligned}$$

as required for Strong Independence.

For PCI we proceed as with Symmetry by showing numerical PCI. Let N_1, N_2 have a common connected component G. Let $f \in G \cap \mathcal{F}$. By definition of a connected component, $s \in \mathsf{src}_{N_1}(f)$ iff $s \in \mathsf{src}_{N_2}(f)$, so $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$. Hence

$$T_{N_1}(f) = |\mathsf{src}_{N_1}(f)| = |\mathsf{src}_{N_2}(f)| = T_{N_2}(f)$$

For $s \in G \cap \mathcal{S}$, we trivially have $T_{N_1}(s) = 1 = T_{N_1}(s)$. Hence numerical PCI is satisfied.

Coherence. The violation of Coherence follows from Theorem 2.3.2, since we have already shown that Symmetry, Monotonicity and POI are satisfied. □

A.4 Proof of Lemma 2.4.2

Proof. The first statement follows easily from the definition of the limit. We shall prove only the second one.

First we prove the 'if' direction. Write $D = T_N^*(f_1) - T_n^*(f_2)$. We need to show that $D < 0$. Write $d_n = T_N^n(f_1) - T_N^n(f_2)$ so that $D = \lim_{n \to \infty} d_n$. Take $\varepsilon = \rho/2 > 0$.

Then for sufficiently large n we have $d_n \leq -\rho/2 < 0$. Taking $n \to \infty$, we have $D = \lim_{n\to\infty} d_n \leq -\rho/2 < 0$ as required.

For the 'only if' direction, suppose $D < 0$. Let $\rho = -D$. Then for any $\varepsilon > 0$, by the definition of the limit there is $K \in \mathbb{N}$ such that $|d_n - D| < \varepsilon$ for $n \geq K$; in particular, $d_n < \varepsilon + D = \varepsilon - \rho$ as required. □

A.5 Proof of Theorem 2.4.2

The following results will be helpful to simplify the proof of Theorem 2.4.2.

Lemma A.5.1. norm *has the following properties.*

1. norm *preserves numerical Symmetry, in the sense that* $\mathsf{norm}(T)$ *satisfies numerical Symmetry whenever* T *does.*
2. norm *leaves rankings unchanged, in the following sense. For* $T \in \mathcal{T}_{Num}$, $N \in \mathcal{N}$, $s_1, s_2 \in \mathcal{S}$, $f_1, f_2 \in \mathcal{F}$,

$$s_1 \sqsubseteq_N^T s_2 \iff s_1 \sqsubseteq_N^{\mathsf{norm}(T)} s_2$$
$$f_1 \preceq_N^T f_2 \iff f_1 \preceq_N^{\mathsf{norm}(T)} f_2$$

Proof. For part (i), suppose T satisfies numerical Symmetry, and write $T' = U(T)$. Let N and $\pi(N)$ be equivalent networks. First note that

$$\max_{x\in\mathcal{S}} |T_N(x)| = \max_{x\in\mathcal{S}} |T_{\pi(N)}(\pi(x))| = \max_{x\in\mathcal{S}} |T_{\pi(N)}(x)|$$

where the second equality follows since π restricted to $\mathcal{S}$ is a surjection into $\mathcal{S}$ by the definition of equivalent networks. If this maximum is 0, then $T'_N(s) = 0 = T'_{\pi(N)}(s)$ for all $s \in \mathcal{S}$. Otherwise,

$$T'_N(s) = \frac{T_N(s)}{\max_{x\in\mathcal{S}} |T_N(x)|} = \frac{T_{\pi(N)}(\pi(s))}{\max_{x\in\mathcal{S}} |T_{\pi(N)}(x)|} = T'_{\pi(N)}(\pi(s))$$

One can show that $T'_N(f) = T'_{\pi(N)}(\pi(f))$ by an identical argument. Hence $T' = U(T)$ satisfies numerical Symmetry also.

Now we prove part (ii). First suppose $s_1 \sqsubseteq_N^T s_2$. Write $T' = \mathsf{norm}(T)$. We have $T'_N(x) = \alpha T_N(x)$ for some $\alpha \geq 0$ and all $x \in \mathcal{S}$ (either $\alpha = 1/\max_{x\in\mathcal{S}} |T_N(x)|$ or $\alpha = 0$). We therefore have

$$\begin{aligned} s_1 \sqsubseteq_N^T s_2 &\implies T_N(s_1) \leq T_N(s_2) \\ &\implies \alpha T_N(s_1) \leq \alpha T_N(s_2) \\ &\implies T'_N(s_1) \leq T'_N(s_2) \\ &\implies s_1 \sqsubseteq_N^{T'} s_2 \end{aligned}$$

as desired.

Now suppose $s_1 \sqsubseteq_N^{T'} s_2$, i.e. $\alpha T_N(s_1) \leq \alpha T_N(s_2)$. If $\alpha > 0$ then dividing by α readily gives $s_1 \sqsubseteq_N^T s_2$. Otherwise, $\alpha = 0$. This means $\max_{x\in\mathcal{S}} |T_N(x)| = 0$, and thus $T_N(x) = 0$ for all $x \in \mathcal{S}$. In particular $T_N(s_1) = 0 \leq 0 = T_N(s_2)$ so $s_1 \sqsubseteq_N^T s_2$.

The second statement regarding fact ranking may be shown using an identical argument. □

Corollary A.5.1. norm *preserves Coherence, Unanimity, Groundedness and PCI.*

Proof of Theorem 2.4.2. Throughout this proof, $(T^n)_{n\in\mathbb{N}}$ will denote the iterative operator *Sums*, T^* will denote the limit operator, and $U = \mathsf{norm} \circ U^{\mathsf{Sums}}$ will denote the update function for *Sums*.

Coherence. Source-Coherence was shown in the main text. The proof that Fact-Coherence is satisfied is similar, and uses Lemma 2.4.3. Suppose $N \in \mathcal{N}$, $T = T^n$ for some $n \in \mathbb{N}$, $\varepsilon, \rho > 0$, and $\mathsf{src}_N(f_1)$ is (ε, ρ)-less trustworthy than $\mathsf{src}_N(f_2)$ with respect to N and $\tilde{T}$ under a bijection φ, where $\tilde{T} = U(T)$. Let $\hat{s} \in \mathsf{src}_N(f_1)$ be such that $\tilde{T}_N(s) - \tilde{T}_N(\varphi(s)) \leq \varepsilon - \rho$.

Write $T' = U^{\mathsf{Sums}}(T)$ so that $\tilde{T} = \mathsf{norm}(T')$, and set

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

We may assume without loss of generality that $\varepsilon < \frac{1}{|\mathcal{S}|}\rho$. Note that for $s \in \mathcal{S}$, $\tilde{T}_N(s) = \alpha T'_N(s)$ and therefore $T'_N(s) = \frac{1}{\alpha}\tilde{T}_N(s)$. Writing

$$\beta = \frac{1}{\max\limits_{y\in\mathcal{F}} |T'_N(y)|}$$

and applying a similar argument as for showing Source-Coherence, we find

$$\begin{aligned}
\tilde{T}_N(f_1) - \tilde{T}_N(f_2) &= \beta \sum_{s\in\mathsf{src}_N(f_1)} \Big(T'_N(s) - T'_N(\varphi(s))\Big) \\
&= \frac{\beta}{\alpha} \sum_{s\in\mathsf{src}_N(f_1)} \Big(\tilde{T}_N(s) - \tilde{T}_N(\varphi(s))\Big) \\
&= \frac{\beta}{\alpha}\left[\underbrace{\Big(\tilde{T}_N(\hat{s}) - \tilde{T}_N(\varphi(\hat{s}))\Big)}_{\leq \varepsilon-\rho} + \sum_{s\in\mathsf{src}_N(f_1)\setminus\{\hat{s}\}} \underbrace{\Big(\tilde{T}_N(s) - \tilde{T}_N(\varphi(s))\Big)}_{\leq\varepsilon}\right] \\
&\leq \frac{\beta}{\alpha} \cdot \underbrace{\Big(|\mathcal{S}|\varepsilon - \rho\Big)}_{<0}
\end{aligned}$$

Now we need to bound β/α from below. Since we assume $T = T^n$ for some $n \in \mathbb{N}$, for any $y \in \mathcal{F}$ we have

$$|T'_N(y)| = \sum_{s\in\mathsf{src}_N(y)} \underbrace{T'_N(s)}_{\leq|\mathcal{F}|} \leq |\mathsf{src}_N(y)| \cdot |\mathcal{F}| \leq |\mathcal{S}| \cdot |\mathcal{F}|$$

Therefore

$$\beta \geq \frac{1}{|\mathcal{S}| \cdot |\mathcal{F}|}$$

Next, we claim there is some fact $\bar{f} \in \mathcal{F}$ with $T_N(\bar{f}) \geq 1/2$ and $\mathsf{src}_N(\bar{f}) \neq \emptyset$. Indeed, if $T = T^1 = T^{\mathsf{fixed}}$ then take any fact with at least one associated source.[1]

Otherwise, since we assume not all scores are 0 in the limit, there is some $\bar{f}$ with $T_N(\bar{f}) = 1$ due to the application of norm. Clearly $\mathsf{src}_N(\bar{f}) \neq \emptyset$, since we would have $T_N(\bar{f}) = 0$ otherwise.

Let $\bar{x} \in \mathsf{src}_N(\bar{f})$. Then

$$|T'_N(\bar{x})| = T'_N(\bar{x}) = \underbrace{T_N(\bar{f})}_{\geq 1/2} + \underbrace{\sum_{f \in \mathsf{facts}_N(\bar{x}) \setminus \{\bar{f}\}} T_N(f)}_{\geq 0} \geq \frac{1}{2}$$

This means

$$\frac{1}{\alpha} = \max_{x \in \mathcal{S}} |T'_N(x)| \geq |T'_N(\bar{x})| \geq \frac{1}{2}$$

and so, finally,

$$\frac{\beta}{\alpha} \geq \frac{1}{|\mathcal{S}| \cdot |\mathcal{F}|} \cdot \frac{1}{2}$$

Combined with what was shown before, this means

$$\tilde{T}_N(f_1) - \tilde{T}_N(f_2) \leq \frac{1}{2 \cdot |\mathcal{S}| \cdot |\mathcal{F}|}\Big(|\mathcal{S}|\varepsilon - \rho\Big)$$

and Fact-Coherence follows from Lemma 2.4.3.

Symmetry. As a consequence of Lemma 2.4.4, to show Symmetry it is sufficient to show that T^{fixed} satisfies numerical Symmetry, and that $U = \mathsf{norm} \circ U^{\text{Sums}}$ preserves numerical Symmetry. Since T^{fixed} is constant with value $1/2$, it is clear that numerical Symmetry is satisfied. Moreover, Lemma A.5.1 part (i) already shows that norm preserves numerical Symmetry, so we only need to show that U^{Sums} does.

To that end, suppose $T \in \mathcal{T}_{Num}$ satisfies numerical symmetry, and write $T' = U^{\text{Sums}}(T)$. Let N and $\pi(N)$ be equivalent networks and $s \in \mathcal{S}$. Then

$$T'_{\pi(N)}(\pi(s)) = \sum_{y \in \mathsf{facts}_{\pi(N)}(\pi(s))} T_{\pi(N)}(y)$$

Note that $f \in \mathsf{facts}_N(s)$ iff $\pi(f) \in \mathsf{facts}_{\pi(N)}(\pi(s))$. Rephrased slightly, we have $y \in \mathsf{facts}_{\pi(N)}(\pi(s))$ iff $\pi^{-1}(y) \in \mathsf{facts}_N(s)$. Hence we may make a 'substitution' $f = \pi^{-1}(y)$ and sum over $\mathsf{facts}_N(s)$, i.e.

$$T'_{\pi(N)}(\pi(s)) = \sum_{f \in \mathsf{facts}_N(s)} T_{\pi(N)}(\pi(f))$$

Applying numerical symmetry for T, we get

$$\begin{aligned} T'_{\pi(N)}(\pi(s)) &= \sum_{f \in \mathsf{facts}_N(s)} T_N(f) \\ &= T'_N(s) \end{aligned}$$

Following the same tactic, one may also show that $T'_{\pi(N)}(\pi(f)) = T'_N(f)$ for all $f \in \mathcal{F}$. Hence U^{Sums} preserves numerical Symmetry, and we are done.

[1]Note that this is always possible since a truth discovery network contains at least one claim by definition.

Unanimity and Groundedness. Unanimity and Groundedness can be proved together using Lemma 2.4.5 and Corollary A.5.1. By these results it is sufficient that T^{fixed} satisfies Unanimity and Groundedness – this is trivial – and that U^{Sums} preserves them.

Suppose T satisfies Unanimity and Groundedness and write $T' = U^{\text{Sums}}(T)$. Assume without loss of generality that $T = T^n$ for some $n \in \mathbb{N}$ so that $T'_N \geq 0$. Suppose $N \in \mathcal{N}$, $f \in \mathcal{F}$ and that $\mathsf{src}_N(f) = \mathcal{S}$. Let $g \in \mathcal{F}$. We must show that $g \preceq_N^{T'} f$. We have

$$T'_N(g) = \sum_{s \in \mathsf{src}_N(g)} T'_N(s) \leq \sum_{s \in \mathcal{S}} T'_N(s) = T'_N(f)$$

i.e. $g \preceq_N^{T'} f$ as required for Unanimity. For Groundedness, suppose $\mathsf{src}_N(f) = \emptyset$. We must show $f \preceq_N^{T'} g$. Indeed, the sum in the expression for $T'_N(f)$ is taken over the empty set, which by convention is 0. Since $T'_N(g) \geq 0$, we are done. $\square$

A.6 Proof of Theorem 2.4.3

Proof. Here we give only the technical details for the argument showing *SC-Voting* satisfies Symmetry, since the results for the other axioms were given in the main text.

Symmetry. Since *Voting* satisfies Symmetry, it is clear that $f_1 \preceq_N^{T^{SCV}} f_2$ iff $\pi(f_1) \preceq_{\pi(N)}^{T^{SCV}} \pi(f_2)$ for any equivalent networks N and $\pi(N)$. We need to show that $s_1 \sqsubseteq_N^{T^{SCV}} s_2$ iff $\pi(s_1) \sqsubseteq_{\pi(N)}^{T^{SCV}} \pi(s_2)$.

First we will show that $\lhd_N$ and $\lhd_{\pi(N)}$ have a similar symmetry property: $s_1 \lhd_N s_2$ iff $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$. Indeed, suppose $s_1 \lhd_N s_2$. Then there is a bijection $\varphi : \mathsf{facts}_N(s_1) \to \mathsf{facts}_N(s_2)$ with $f \preceq_N^{T^{SCV}} \varphi(f)$, and there is some $\hat{f}$ with $\hat{f} \prec_N^{T^{SCV}} \varphi(\hat{f})$.

It can be seen that π restricted to $\mathsf{facts}_N(s_i)$ is a bijection into $\mathsf{facts}_{\pi(N)}(\pi(s_i))$. Let π_1 and π_2 denote these restrictions for $i = 1, 2$ respectively. Set $\theta = \pi_2 \circ \varphi \circ \pi_1^{-1}$, so that θ maps $\mathsf{facts}_{\pi(N)}(\pi(s_1))$ into $\mathsf{facts}_{\pi(N)}(\pi(s_2))$. As a composition of bijections, θ is itself bijective.

Let $g \in \mathsf{facts}_{\pi(N)}(\pi(s_1))$. Write $f = \pi_1^{-1}(g) \in \mathsf{facts}_N(s_1)$. By the property of φ, we have

$$f \preceq_N^{T^{SCV}} \varphi(f)$$

By the symmetry property of the fact-ranking (which follows from symmetry of *Voting*), we can apply π to the above to get

$$\pi(f) \preceq_{\pi(N)}^{T^{SCV}} \pi(\varphi(f))$$

Since $f \in \mathsf{facts}_N(s_1)$ and $\varphi(f) \in \mathsf{facts}_N(s_2)$, we have $\pi(f) = \pi_1(f)$ and $\pi(\varphi(f)) = \pi_2(\varphi(f))$. Using this fact in the above inequality and recalling $f = \pi^{-1}(g)$ we get

$$g = \pi_1(f) = \pi(f) \preceq_{\pi(N)}^{T^{SCV}} \pi(\varphi(f)) = \pi_2(\varphi(f)) = \pi_2(\varphi(\pi_1^{-1}(g))) = \theta(g)$$

i.e. $g \preceq_{\pi(N)}^{T^{SCV}} \theta(g)$. Applying the same argument with $\hat{g} = \pi_1^{-1}(\hat{f})$ we get $\hat{g} \prec_{\pi(N)}^{T^{SCV}} \theta(\hat{g})$.

This shows that $\text{facts}_{\pi(N)}(\pi(s_1))$ is less believable than $\text{facts}_{\pi(N)}(\pi(s_2))$ with respect to *SC-Voting* (whose fact-ranking coincides with *Voting*) in $\pi(N)$ under θ. Hence $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$.

We have shown $s_1 \lhd_N s_2 \implies \pi(s_1) \lhd_{\pi(N)} \pi(s_2)$. For the converse implication, apply the same argument starting from $\pi(s_1) \lhd_{\pi(N)} \pi(s_2)$ with the π^{-1}.

Next, we note that for $i = 1, 2$ and any $t \in \mathcal{S}$,

$$\begin{aligned} t \in W_N(s_i) &\iff t \lhd_N s_i \\ &\iff \pi(t) \lhd_{\pi(N)} \pi(s_i) \\ &\iff \pi(t) \in W_{\pi(N)}(\pi(s_i)) \end{aligned}$$

Consequently π restricted to $W_N(s_i)$ is a bijection into $W_{\pi(N)}(\pi(s_i))$, which means $|W_N(s_i)| = |W_{\pi(N)}(\pi(s_i))|$. Finally, this means

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

as required for Symmetry. □

A.7 Proof of Theorem 2.4.5

Proof. Here we show that *UnboundedSums* satisfies Symmetry, PCI, Unanimity and Groundedness, since the other axioms were dealt with in the main text.

Throughout the proof, let $(T^n)_{n\in\mathbb{N}}$ denote *UnboundedSums*, T^* denote the ordinal limit of *UnboundedSums*, and for a network N let J_N be as in Theorem 2.4.4. Then the rankings in N induced by T^n for $n \geq J_N$ are the same as T^*.

Symmetry. In the proof of Theorem 2.4.2, we saw that the update function U^{Sums} preserves numerical Symmetry, in the sense that if T satisfies numerical Symmetry then $U^{\text{Sums}}(T)$ does also. Since it is clear that the prior operator for *UnboundedSums* satisfies numerical Symmetry, T^n satisfies numerical Symmetry and consequently Symmetry for all $n \in \mathbb{N}$.

Now, let N and $\pi(N)$ be equivalent networks. Let $J, J' \in \mathbb{N}$ be such that $T^*(N)$ and $T^*(\pi(N))$ are given by T_N^J and $T_{\pi(N)}^{J'}$ respectively and take $n \geq \max\{J, J'\}$. For $s_1, s_2 \in \mathcal{S}$ we have by Symmetry of T^n,

$$\begin{aligned} s_1 \sqsubseteq_N^{T^*} s_2 &\iff s_1 \sqsubseteq_N^{T^n} s_2 \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^n} \pi(s_2) \\ &\iff \pi(s_1) \sqsubseteq_{\pi(N)}^{T^*} \pi(s_2) \end{aligned}$$

as required for Symmetry. Using an identical argument, one can show that $f_1 \preceq_N^{T^*} f_2$ iff $\pi(f) \preceq_{\pi(N)}^{T^*} \pi(f_2)$. Hence T^* satisfies Symmetry.

PCI. As with Symmetry, we will show that T^n satisfies numerical PCI, and consequently PCI, for all $n \in \mathbb{N}$. Let N_1, N_2 be networks with a common connected component G. Let $s \in G \cap \mathcal{S}$ and $f \in G \cap \mathcal{F}$. Note that $\mathsf{facts}_{N_1}(s) = \mathsf{facts}_{N_2}(s)$ and $\mathsf{src}_{N_1}(f) = \mathsf{src}_{N_2}(f)$ since by definition a source is connected to its facts and vice versa. For $n = 1$ we have

$$T^1_{N_1}(s) = 1 = T^1_{N_2}(s)$$

$$T^1_{N_1}(f) = |\mathsf{src}_{N_1}(f)| = |\mathsf{src}_{N_2}(f)| = T^1_{N_2}(f)$$

so T^1 has numerical PCI. Supposing T^n has numerical PCI for some $n \in \mathbb{N}$, we have

$$T^{n+1}_{N_1}(s) = \sum_{g \in \mathsf{facts}_{N_1}(s)} \underbrace{T^n_{N_1}(g)}_{=T^n_{N_2}(g)} = \sum_{g \in \mathsf{facts}_{N_2}(s)} T^n_{N_2}(g) = T^{n+1}_{N_2}(s)$$

and similarly

$$T^{n+1}_{N+1}(f) = T^{n+1}_{N_2}(f)$$

Hence, by induction, T^n has numerical PCI for all $n \in \mathbb{N}$, and we are done.

Unanimity and Groundedness. For Unanimity, suppose $\mathsf{src}_N(f) = \mathcal{S}$. For any $g \in \mathcal{F}$ and $n \in \mathbb{N}$ we have

$$\begin{aligned} T^n_N(g) &= \sum_{s \in \mathsf{src}_N(g)} T^n_N(s) \\ &\leq \sum_{s \in \mathsf{src}_N(g)} T^n_N(s) + \sum_{s \in \mathcal{S} \setminus \mathsf{src}_N(g)} T^n_N(s) \\ &= \sum_{s \in \mathcal{S}} T^n_N(s) \\ &= \sum_{s \in \mathsf{src}_N(f)} T^n_N(s) \\ &= T^n_N(f) \end{aligned}$$

so $g \preceq^{T^n}_N f$ for all $n \in \mathbb{N}$. Since the ranking of T^* corresponds to T^n for large n, we have $g \preceq^{T^*}_N f$ as required

For Groundedness, suppose $\mathsf{src}_N(f) = \emptyset$. Then $T^n_N(f) = 0$ for all $n \in \mathbb{N}$. For any $g \in \mathcal{F}$, we have $T^n_N(g) \geq 0 = T^n_N(f)$. Consequently $f \preceq^{T^n}_N g$ for all $n \in \mathbb{N}$. As above, this gives $f \preceq^{T^*}_N g$ as required. □