Trustworthiness and Expertise: Social Choice and Logic-based Perspectives

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

Joseph Singleton

XXX 2022

Abstract

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

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

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

Contents

Acknowledgements

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

List of Publications

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

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

1 Introduction

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

1.1 Social Choice Perspectives

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

1.2 Logic-based Perspectives

1.3 Overview

Chapter-by-chapter breakdown of the thesis.

2 Truth Discovery

[**TODO:** Introduction]

2.1 Preliminaries

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

Input. Intuitively, a truth discovery problem consists of a number of *sources* and a number of *objects* of interest. Each source provides a number of *claims*, where a claim is comprised of an object and a *value*. Different sources may give conflicting claims by providing different values for the same object. For simplicity, we only consider categorical values in this work. Note that while this restriction is made in some approaches in the literature [14, 24, 19, 6, 26], in general truth discovery methods also handle continuous values [12, 21].

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

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

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

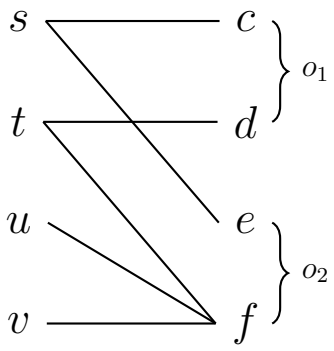


Figure 2.1: Illustrative example of a truth discovery problem, with sources s, t, u, v*, object* o_1 *with associated claims* c *and* d*, and* o_2 *with claims* e *and* f*.*

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

such that

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

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

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

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

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

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

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

Example 2.1.1 highlights a special case of our framework: the "binary" case in which the domain of each object consists of two values $D_o = \{\mathsf{true}, \mathsf{false}\}$. In this case we can think of each object as a propositional variable. This brings us close to the setting studied in *judgment aggregation* [8] and, specifically (since sources do not necessarily provide a claim for each object) to the setting of *binary aggregation with abstentions* [3, 5]. 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., [3]). 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 [12, 4, 23], whereas others give an assessment of the believability (or confidence, probability etc…) of *each claim* (o, v) [24, 14, 9, 27, 25, 26]. We opt for the latter, more general, approach.

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

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

We will ultimately take a hybrid approach: our methods and example will be defined in terms of numeric scores, but the axioms will only refer to ordinal properties. This approach is summarised succinctly by Altman and Tennenholtz [1], who

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

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

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

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

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

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

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

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

2.2 Example Operators

In this section we capture several example operators from the literature in our framework: a baseline *voting* method and its generalisation to *weighted* voting, *Sums* [14], *TruthFinder* [24] and *CRH* [12]. As is the case with many methods in the literature, the latter three methods operate iteratively: starting with an initial estimate, scores are repeatedly updated according to some procedure until convergence. Typically the update procedure is recursive, with source scores being updated on the basis of the current claims scores, and vice versa. To simplify the definition and analysis of such methods, we will introduce the class of *recursive operators*.

2.2.1 Voting

It is common in the literature to evaluate truth discovery methods against a non-trust-aware method, such as a simple voting procedure.[1] Here we consider each source to "vote" for their claims, and claims are ranked according to the number of votes received, i.e. by $|\operatorname{src}_N(c)|$. While this ignores the trust aspect of truth discovery entirely, this method will be useful for us as an axiomatic baseline. For example,

axioms which aim to address the trust aspect should not hold for voting, and an axiom referring to the ranking of claims may be too strong if it does hold for voting.

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

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

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

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

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

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

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

Example 2.2.1. *Set*

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

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

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

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

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

[1]This is often called *majority voting* in the truth discovery literature (e.g. [11, 20, 12]), but using the terminology of social choice theory it is better described as *plurality voting*.

2.2.2 Recursive Operators

To capture the mutual dependence between trust in sources and belief in claims, truth discovery methods generally involve recursive computation [14, 24, 22, 7, 26, 12, 9, 27]. Claim scores are updated on the basis of currently estimated source scores, before claim scores are updated on the basis of the new sources scores. If this process converges, the limiting scores should be a fixed-point of the update procedure, reflecting the desired mutual dependence. To formalise this idea, we define recursive operators.

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

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

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

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

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

Sums. Sums [14] is a simple and well-known operator adapted from the *Hubs and Authorities* [10] algorithm for ranking web pages. The premise is to extend the linear sum of weighted voting to both claim and source scores: we update the score of each source as the sum of the scores of its claims, and update the score of each claim as the sum of the scores of its sources. To prevent scores from growing without bound, they are normalised at each iteration by dividing by the maximum score (for sources and claims separately).

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

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

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

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

TruthFinder. TruthFinder [24] 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 2.2.5. *Given parameters $\rho, \gamma \in (0, 1)$ and $\lambda \in [0, 1]$,* TruthFinder *is the recursive scheme $(\mathcal{D}, T^0, U)$, where $\mathcal{D}$ is the set of operators with $0 < T_N(s) < 1$ for all N and $s \in S$ with $\text{cl}_N(s) \neq \emptyset$, $T^0 \equiv 0.9$, and $U(T) = T'$, with*

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

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

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

We refer the reader to the original TruthFinder paper [24] for the interpretation of ρ and γ. As described above, λ controls the amount to which conflicting claims play a role in the evaluation of a given claim. Of special interest is the case $\lambda = 0$, in which the denominator in (2.1) is 1. Note that in (2.1) we have unfolded the definitions of [24] in order to obtain a single expression of $T'_N(c)$ in terms of the $T_N(s)$, at the expense of interpretability.

Let us return again to the network in Fig. 2.1. We take parameters $\rho = 0.5$ and $\gamma = 0.3$ (as per the experimental setup of Yin, Han, and Yu [24]) and $\lambda = 0.5$. Assuming that TruthFinder does indeed converge on this network – as it appears to do empirically – we have $T_N^{\text{tf}}(s) \approx 0.5067$, $T_n^{\text{tf}}(t) \approx 0.6590$ and $T_N^{\text{tf}}(u) = T_N^{\text{tf}}(v) = 0.7510$, which gives the ranking $s \sqsubset t \sqsubset u \simeq v$ on the sources. We have $T_N^{\text{tf}}(c) \approx 0.5328$, $T_N^{\text{tf}}(d) \approx 0.5670$, $T_N^{\text{tf}}(e) \approx 0.4807$ and $T_N^{\text{tf}}(f) \approx 0.7510$, which gives the ranking $e \prec c \prec d \prec f$ on the claims. Note that the source ranking coincides with that of Example 2.2.1, and the claim ranking refines that of Example 2.2.1 and Sums by ranking e *strictly* worse than c. Intuitively, this occurs because e has more sources reporting the conflicting claim (namely, f) than c does. If we instead take $\lambda = 0$,

so that sources for conflicting claims are not considered, then the ranking reverts to $c \approx e \prec d \prec f$ (and the source ranking remains the same).

CRH. Standing for "Conflict Resolution on Heterogeneous Data", CRH is an optimisation-based framework for truth discovery [12]. 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 [12, §2.4.1], where each categorical value is represented as a one-hot vector, and the source weight regularisation from [12, Eq. (4)]. We make a minor modification, however, by adding a small quantity ε to α_s and $T'_N(s)$ defined below; this ensures the logarithm in $T'_N(s)$ and the division in $T'_N(c)$ is well-defined and simplifies analysis of CRH later on.

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

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

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

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

with

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

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

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

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

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

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

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

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

2.3 The Axioms

Having laid out the formal framework, we now introduce axioms for truth discovery. Such axioms are formal properties an operator may satisfy, which encode intuitively desirable behaviour. Many of our axioms are adaptations of axioms for various problem in social choice theory (e.g. from voting [28] and ranking systems [1]), in which the axiomatic method has seen great success. We also consider standard social choice axioms which are *not* desirable for truth discovery, to highlight the differences with classical problems such as voting. We will later revisit the example operators of the previous section to see to what extent our axioms hold in practise.

2.3.1 Coherence

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

We aim to formulate this principle axiomatically as a *coherency* property relating the source ranking $\sqsubseteq$ and the claim ranking $\preceq$: sources making higher $\preceq$-ranked claims should rank highly in $\sqsubseteq$, and vice versa. To do so we adapt the idea behind the *Transitivity* axiom of Altman and Tennenholtz [1] for ranking systems.

Now, a difficulty arises when considering how to compare the claims of two sources. For a simple example, suppose sources have either *low*, *medium* or *high* trustworthiness. How should we rank a claim c with one *medium* sources versus a claim d with a *low* and a *high* source? In some situations we may want to prioritise the number of claims, so that d is preferred. In others we may want to avoid trusting *low* sources as much as possible, so that c is preferred. The third option of ranking c and d equally believable is also reasonable.

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

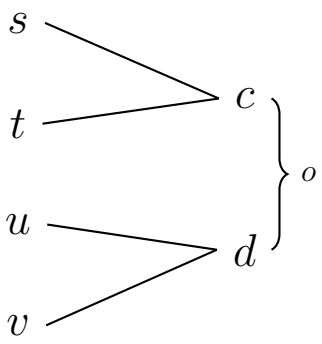


Figure 2.2: A network illustrating ***Claim-coherence****.*

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

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

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

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

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

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

We need a preliminary lemma.

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

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

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

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

We are now ready to state our first two axioms.

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

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

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

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

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

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

The reader may wonder why we only consider the *strict* pairwise relation in **Claim-coherence** (and **Source-coherence**). An alternative axiom might require that $c \preceq c'$ whenever $\mathrm{src}_N(s)$ precedes $\mathrm{src}_N(s')$ with respect to $\sqsubseteq$ (not necessarily strictly). However, this property implies that $c \approx c'$ whenever $\mathrm{src}_N(c) = \mathrm{src}_N(c')$. We have already seen an example operator where this does not hold: TruthFinder ranks $e \prec c$ in the network N from Fig. 2.1, but $\mathrm{src}_N(c) = \mathrm{src}_N(e) = \{s\}$. Intuitively, c and e are "tied" when it come to the quality of their own sources, but there are fewer sources *disagreeing* with c (the "antisources") than e. Stating our coherence properties in the strict form permits an operator to consider antisources in cases where there is no clear comparison on the basis of sources alone.

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

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

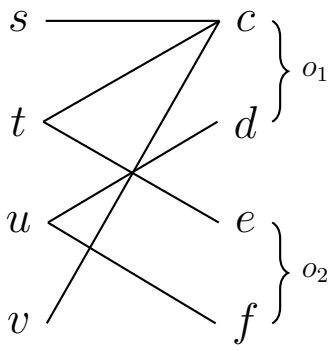


Figure 2.3: A network isomorphic to the one shown in Fig. 2.1.

2.3.2 Symmetry

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

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

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

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

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

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

We illustrate **Symmetry** with an example.

Example 2.3.2. *Consider the network N from Fig. 2.1 and N' from Fig. 2.3, where we take the sources, objects and domains to be the same in both networks. Then N and N' are isomorphic via the mapping F expressed in cycle notation as $(suv)(cf)(de)(o_1o_2)$. For example, s plays the same role in N as u in N', c plays the same role in N as f in N', the role of objects o_1 and o_2 are swapped, etc.* ***Symmetry*** *requires that the source and claim rankings*

in N' are already determined by the rankings of N. For example, if the source ranking in N is $s \sqsubset_N u \simeq_N v \sqsubset_N t$, we must have $u \sqsubset_{N'} v \simeq_{N'} s \sqsubset_{N'} t$.

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

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

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

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

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

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

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

2.3.3 Monotonicity

Given that voting is not a viable truth discovery method, the believability of a claim c should not increase monotonically with $|\mathrm{src}_N(c)|$. Moreover, it should not increase with the *set* of sources $\mathrm{src}_N(c)$, ordered by set inclusion: $\mathrm{src}_N(c) \subseteq \mathrm{src}_N(d)$ should not in general imply $c \preceq d$. Indeed, consider an adversarial source t deliberately making false claims, and suppose $\mathrm{src}_N(c) = \{s\}$ and $\mathrm{src}_N(d) = \{s, t\}$. Then $\mathrm{src}_N(c) \subseteq \mathrm{src}_N(d)$, but the extra support from t should actually *decrease* the believability of d – since t only provides false claims – not increase it.

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

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

Responsiveness. In the context of voting, positive responsiveness requires that if a voter switches their vote from candidate B to a winning candidate A, then A becomes the unique winner [28]. A naive version of positive responsiveness for

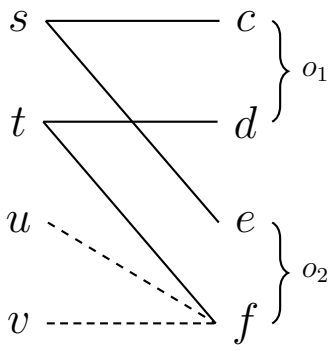


*Figure 2.4: Networks N_0 (solid edges only), $N_1 = N_0 + (u, f)$ and $N_2 = N_1 + (v, f)$ illustrating **Fresh-pos-resp** and **Source-pos-resp**.*

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

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

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

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

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

Example 2.3.3. *Fig. 2.4 illustrates **Fresh-pos-resp** and **Source-pos-resp**. Let N_0 denote the network including only the solid edges, $N_1 = N_0 + (u, f)$, and $N_2 = N_1 + (v, f)$. Note that N_2 is our running example network from Fig. 2.1. Assuming **Symmetry**, everything is tied in N_0: we have $s \simeq_{N_0} t$ and $c \approx_{N_0} d \approx_{N_0} \approx_{N_0} e \approx_{N_0} f$. Since N_1 is the result of adding the report (u, f) and u makes no claims in N_0, **Fresh-pos-resp** gives $e \prec_{N_1} f$. Since*

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

*$s \in \mathrm{src}_{N_0}(e) \subseteq \mathrm{antisrc}_{N_0}(f)$ and $t \in \mathrm{src}_{N_0}(f)$, **Source-pos-resp** gives $s \sqsubset_{N_1} t$. Going from N_1 to N_2 we can repeat exactly the same arguments to find $e \prec_{N_2} f$ and $s \sqsubset_{N_2} t$.*

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

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

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

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

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

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

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

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

While Proposition 2.3.3 only addresses a somewhat trivial case, it will turn out to be useful in characterising voting behaviour more generally in Sections 2.3.4 and 2.3.6. It can be seen as one of the many generalisations of *May's Theorem* [13], which characterises the majority voting rule in two-candidate elections. To prove it, we need a preliminary result.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Trust-based monotonicity. Suppose $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$. 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 2.3.3. *Given a network N, a source $s \in S$ is* marginally trustworthy *with respect to an operator T if there exist claims $c, d \in C$ such that $s \notin \text{src}_N(c)$, $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$ and $c \preceq_N^T d$. Similarly, s is* marginally untrustworthy *if there are $c, d \in C$ such that $s \notin \text{src}_N(c)$, $\text{src}_N(d) = \text{src}_N(c) \cup \{s\}$ 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 $s \sqsubset_N^T t$.

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

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

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

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

2.3.4 Independence

Another common class of axioms in social choice theory are *independence* axioms, which require that some aspect of the output is independent of "irrelevant" parts of the input. The original example is Arrow's *Independence of Irrelevant Alternatives* (IIA) in voting theory [2], which says, roughly speaking, that the ranking of candidates A and B should depend only on the individual rankings of A and B, not on any "irrelevant" alternative C. It has been adapted to several settings in which the axiomatic method has been applied. Perhaps closest to our setting is judgment aggregation, where independence requires the collective acceptance of a report φ does not depend on how the individuals accept or reject some other report ψ [8].

A version of IIA can be easily stated in our framework: the ranking of claims c and d should depend only on the sources reporting c and d, not on the sources for

other claims. However, this axiom is clearly *undesirable* for truth discovery. Indeed, consider again the network N from Fig. 2.1. As we have argued informally, claim c is intuitively weaker than d because how of their respective sources interact with other claims in the network. Nevertheless, we state this axiom as a point of comparison with classical social choice problems such as voting.

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

That is, if c and d have the same sources in N and N', they have the same relative ranking in both networks. The undesirability of **Classical-independence** can be formalised axiomatically: together with our earlier axioms, it implies voting-like behaviour within the claims for each object.[4] Note that for the special case of binary networks, similar results have been shown in the literature on binary aggregation with abstentions [3].

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

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

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

Now, note that since $c, d \in \mathrm{cl}_N(o)$, for $o' \neq o$ and $e \in \mathrm{cl}_N(o')$ we have $e \notin \{c, d\}$, so $\mathrm{src}_N(e) = \emptyset$. Hence $\mathrm{src}_N(o') = \emptyset$ for such o'. Since T satisfies **Symmetry** and **Fresh-pos-resp**, we may apply Proposition 2.3.3 (1) to find $c \preceq^T_{N'} d$ iff $|\mathrm{src}_{N'}(c)| \leq |\mathrm{src}_{N'}(d)|$. But $|\mathrm{src}_{N'}(c)| = |\mathrm{src}_N(c)|$, and likewise for d. Consequently

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

as desired. □

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

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

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

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

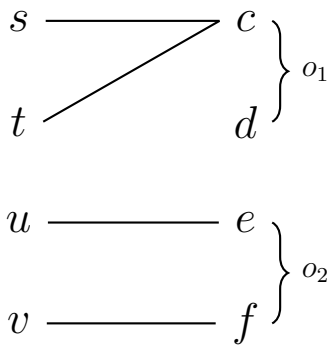


Figure 2.5: A network illustrating ***Disjoint-independence****.*

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

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

Proof.

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

□

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

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

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

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

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

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

2.3.5 Conflicting claims

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

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

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

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

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

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

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

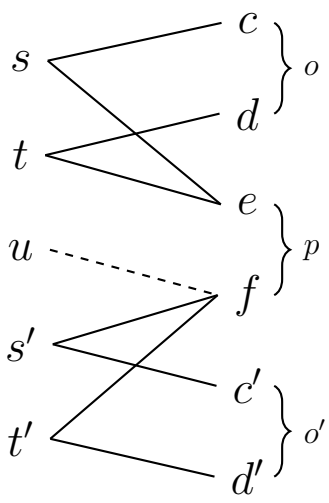


Figure 2.6: Network used to illustrate the impossibility results of Theorem 2.3.2.

to get $d \prec^T_{N_1} d'$. But by **Symmetry** as above, we have $c \approx^T_{N_1} d$ and $c' \approx^T_{N_1} d'$. Consequently $c \approx^T_{N_1} d \prec^T_{N_1} d' \approx^T_{N_1} c'$, i.e. $c \prec^T_{N_1} c'$. Clearly this contradicts $c' \prec^T_{N_1} c$. If $c \prec^T_{N_1} c'$ we obtain a contradiction by an identical argument. Hence $c \approx^T_{N_1} c'$.

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

2. Suppose T additionally satisfies **Source-pos-resp** and **Conflict-coherence**. By the same argument as above, **Conflict-coherence** and **Symmetry** together dictate that $c \approx^T_{N_1} c'$. But by **Symmetry** and **Source-pos-resp**, we have $s \sqsubset^T_{N_1} s'$. **Claim-coherence** then implies $c \prec^T_{N_1} c'$: contradiction.

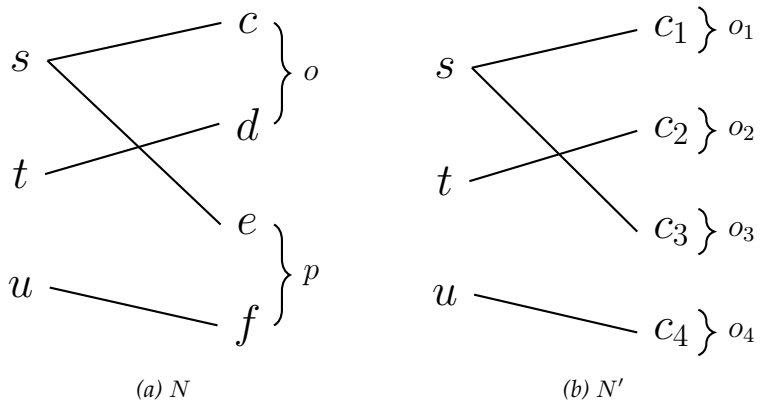
3. Suppose T additionally satisfies **Source-pos-resp** and **Anti-coherence**. Again, $s \sqsubset^T_{N_1} s'$. **Claim-coherence** implies $c \prec^T_{N_1} c'$. Since $\text{antisrc}_{N_1}(d) = \{s\}$ and $\text{antisrc}_{N_1}(d') = \{s'\}$, **Anti-coherence** gives $d' \prec^T_{N_1} d$. But recall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

□

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

2.3.6 Axiomatic Characterisation of Voting

Recall from Proposition 2.3.4 that **Symmetry**, **Fresh-pos-resp** and **Classical-independence** force an operator to rank claims for the object simply by their number of sources, as in voting from Section 2.2.1. In this section we give two further axioms which



(a) N *(b) N'*

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

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

Flat-sources. For all $s, s' \in S$, $s \simeq_N^T s'$.

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

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

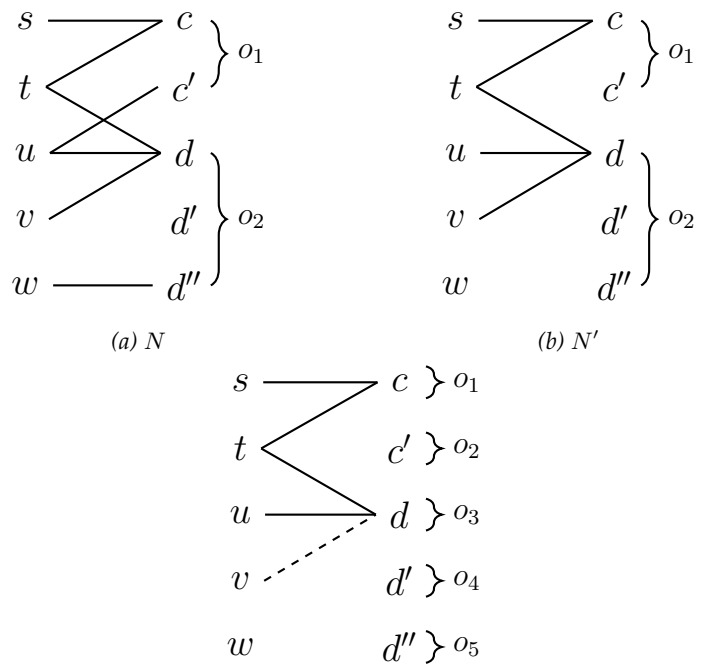
1. *$S' = S$.*
2. *f is a bijection $C_N \to C_{N'}$ such that $(s, c) \in R$ iff $(s, f(c)) \in R'$.*
3. *For al $o \in O'$, $|D'_o| = 1$.*

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

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

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

We can now characterise voting, up to ranking equivalence.



(a) N *(b) N'*

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

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

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

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

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

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

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

Table 2.2: Axiom satisfaction for the example operators.

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

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

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

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

2.4 Fixed-points for Recursive Operators

2.5 Satisfaction of the Axioms

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof. First, note that for any claim d,

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

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

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

□

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

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

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

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

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

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

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

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

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

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

[7] While Pasternack and Roth [14] do not consider convergence, Sums is an adaptation of the *Hubs and Authorities* algorithm, for which Kleinberg [10] proves convergence: phrased in our terminology,

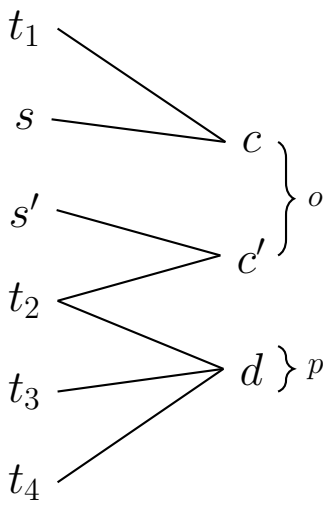


*Figure 2.9: Counterexample for **Source-coherence** for Weighted Agreement.*

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

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

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

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

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

Thus

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

Since the maximum is achieved for some i, we get

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

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

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

Combining (2.7) and (2.8), we get

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

as required. □

he shows that the vector of source scores converge to a unit eigenvector of the matrix MM^{T} corresponding to the largest eigenvector (in absolute value), where M is the $|S| \times |C|$ matrix defined by $M_{sc} = \mathbb{1}[s \in \mathrm{src}_N(c)]$. Similarly, claim scores converge to a unit eigenvector of $M^{\mathrm{T}}M$. [**TODO:** possibly signpost that we also take linear algebra approach for unbounded sums?]

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

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

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

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

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

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

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

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

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

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

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

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

We need to show $T^{\text{sums}} \sim T^w$, i.e. that T^{sums} and T^w give the same rankings on all networks N. If N contains no reports then both T_N^{sums} and T^w are zero, and

therefore output the same rankings. Suppose N contains at least one report. Since we just divide by a constant in (2.10), $s \sqsubseteq_N^{T^{\text{sums}}} s'$ iff $s \sqsubseteq_N^{T^w} s'$ for all sources s and s'. Using the fact that $T^{\text{sums}} = U(T^{\text{sums}})$ from Lemma 2.5.3, it is easily seen that $T_N^{\text{sums}}(c) = \sum_{s \in \text{src}_N(c)} w_N(s) = T_N^w(c)$. Hence T_N^{sums} and T_N^w give exactly the same scores for claims, and in particular the rankings also coincide. □

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

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

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

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

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

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

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

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

For the negative axioms, we refer to networks shown in Fig. 2.10. For **Fresh-pos-resp** and **Source-pos-resp**, let N_0 denote the network without the dashed report (v, d), so that $N_0 + (v, d)$ is the full network. It can be shown that the rankings are the same under Sums in both networks, with $s \simeq t \simeq u \simeq v \sqsubset x_1 \simeq x_2 \simeq x_3 \simeq x_4$

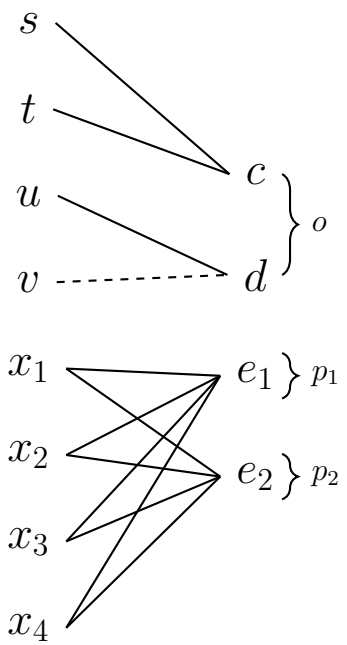


Figure 2.10: Networks used as counterexamples for Sums axiom failures.

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

For **Classical-independence** and **Disjoint-independence**, let N_1 and N_2 denote the upper and lower components of the network in Fig. 2.10, excluding the dashed report (v, d). Then $N_0 = N_1 \sqcup N_2$. Hence $c \approx_{N_1 \sqcup N_2}^{T^{\text{sums}}} d$. However, it is straightforward to check that in the network N_1 alone we have $d \prec_{N_1}^{T^{\text{sums}}} c$; this violates **Disjoint-independence**. Taking N_0' to be the network obtained from N_0 by removing all reports from $x_1, \ldots, x_4$, we have $d \prec_{N_0'}^{T^{\text{sums}}} c$ and $c \approx_{N_0}^{T^{\text{sums}}} d$. Since c and d have the same sources in both networks, this violates **Classical-independence**.

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

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

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

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

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

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

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

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

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

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

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

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

Hence

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

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

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

as required.

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

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

This completes the proof. □

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

Proof. From Lemma 2.5.4 we have

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

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

2.5.1 Modifying Sums

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

2.6 Related Work

2.7 Conclusion

Bibliography

[1] Alon Altman and Moshe Tennenholtz. "Axiomatic Foundations for Ranking Systems". In: *J. Artif. Int. Res.* 31.1 (Mar. 2008), pp. 473–495. ISSN: 1076-9757. URL: http://dl.acm.org/citation.cfm?id=1622655.1622669 (cited on pages 5, 11).

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

[3] Zoé Christoff and Davide Grossi. "Binary Voting with Delegable Proxy: An Analysis of Liquid Democracy". In: *Proc. TARK 2017*. 2017 (cited on pages 4, 20).

[4] Hu Ding, Jing Gao, and Jinhui Xu. "Finding global optimum for truth discovery: Entropy based geometric variance". In: *Proc. 32nd International Symposium on Computational Geometry (SoCG 2016)*. 2016 (cited on page 5).

[5] Elad Dokow and Ron Holzman. "Aggregation of binary evaluations with abstentions". In: *Journal of Economic Theory* 145 (2010), pp. 544–561 (cited on page 4).

[6] Xin Luna Dong, Laure Berti-Equille, and Divesh Srivastava. "Truth Discovery and Copying Detection in a Dynamic World". In: *Proc. VLDB Endow.* 2.1 (Aug. 2009), pp. 562–573. ISSN: 2150-8097. DOI: 10.14778/1687627.1687691. URL: https://doi.org/10.14778/1687627.1687691 (cited on page 3).

[7] Yang Du et al. "Bayesian Co-Clustering Truth Discovery for Mobile Crowd Sensing Systems". In: *IEEE Transactions on Industrial Informatics* (2019), pp. 1–1. ISSN: 1551-3203. DOI: 10.1109/TII.2019.2896287 (cited on page 8).

[8] Ulle Endriss. "Judgment Aggregation". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 17 (cited on pages 4, 19).

[9] Alban Galland et al. "Corroborating Information from Disagreeing Views". In: *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. WSDM '10. New York, New York, USA: ACM, 2010, pp. 131–140. ISBN: 978-1-60558-889-6. DOI: 10.1145/1718487.1718504. URL: http://doi.acm.org/10.1145/1718487.1718504 (cited on pages 5, 8).

[10] Jon M. Kleinberg. "Authoritative Sources in a Hyperlinked Environment". In: *J. ACM* 46.5 (Sept. 1999), pp. 604–632. ISSN: 0004-5411. DOI: 10.1145/324133.324140. URL: http://doi.acm.org/10.1145/324133.324140 (cited on pages 8, 28).

[11] Yaliang Li et al. "A Survey on Truth Discovery". In: *SIGKDD Explor. Newsl.* 17.2 (2016), pp. 1–16. ISSN: 1931-0145. DOI: 10.1145/2897350.2897352. URL: http://doi.acm.org/10.1145/2897350.2897352 (cited on page 7).

[12] Yaliang Li et al. "Conflicts to Harmony: A Framework for Resolving Conflicts in Heterogeneous Data by Truth Discovery". In: *IEEE Transactions on Knowledge and Data Engineering* 28.8 (Aug. 2016), pp. 1986–1999. ISSN: 1041-4347. DOI: 10.1109/TKDE.2016.2559481 (cited on pages 3, 5–8, 10).

[13] Kenneth O May. "A set of independent necessary and sufficient conditions for simple majority decision". In: *Econometrica: Journal of the Econometric Society* (1952), pp. 680–684 (cited on page 17).

[14] Jeff Pasternack and Dan Roth. "Knowing What to Believe (when You Already Know Something)". In: *Proceedings of the 23rd International Conference on Computational Linguistics*. COLING '10. Beijing, China: Association for Computational Linguistics, 2010, pp. 877–885. URL: http://dl.acm.org/citation.cfm?id=1873781.1873880 (cited on pages 3, 5, 6, 8, 28).

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

[16] 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 iii).

[17] 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 iii).

[18] 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 iii).

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

[20] Houping Xiao and Shiyu Wang. "A Joint Maximum Likelihood Estimation Framework for Truth Discovery: A Unified Perspective". In: *IEEE Transactions on Knowledge and Data Engineering* (2015), pp. 1–1. DOI: 10.1109/TKDE.2022.3173911 (cited on page 7).

[21] Houping Xiao et al. "A Truth Discovery Approach with Theoretical Guarantee". In: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 1925–1934. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939816. URL: http://doi.acm.org/10.1145/2939672.2939816 (cited on page 3).

[22] Yi Yang, Quan Bai, and Qing Liu. "A probabilistic model for truth discovery with object correlations". In: *Knowledge-Based Systems* 165 (2019), pp. 360–373. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2018.12.004. URL: http://www.sciencedirect.com/science/article/pii/S0950705118305914 (cited on page 8).

[23] Yi Yang, Quan Bai, and Qing Liu. "On the Discovery of Continuous Truth: A Semi-supervised Approach with Partial Ground Truths". In: *Web Information Systems Engineering – WISE 2018*. Springer International Publishing, 2018, pp. 424–438. DOI: 10.1007/978-3-030-02922-7_29. URL: https://doi.org/10.1007%2F978-3-030-02922-7_29 (cited on page 5).

[24] Xiaoxin Yin, Jiawei Han, and Philip S. Yu. "Truth Discovery with Multiple Conflicting Information Providers on the Web". In: *IEEE Transactions on Knowledge and Data Engineering* 20.6 (June 2008), pp. 796–808. ISSN: 1041-4347. DOI: 10.1109/TKDE.2007.190745 (cited on pages 3, 5, 6, 8, 9).

[25] Daniel Yue Zhang et al. "On robust truth discovery in sparse social media sensing". In: *2016 IEEE International Conference on Big Data (Big Data)*. 2016-12, pp. 1076–1081. DOI: 10.1109/BigData.2016.7840710 (cited on page 5).

[26] Liyan Zhang et al. "Latent Dirichlet Truth Discovery: Separating Trustworthy and Untrustworthy Components in Data Sources". In: *IEEE Access* 6 (2018), pp. 1741–1752. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2780182 (cited on pages 3, 5, 8).

[27] Shi Zhi et al. "Modeling Truth Existence in Truth Discovery". In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. Sydney, NSW, Australia: ACM, 2015, pp. 1543–1552. ISBN: 978-1-4503-3664-2. DOI: 10.1145/2783258.2783339. URL: http://doi.acm.org/10.1145/2783258.2783339 (cited on pages 5, 8).

[28] William S. Zwicker. "Introduction to the Theory of Voting". In: *Handbook of Computational Social Choice*. Ed. by Felix Brandt et al. 1st. New York, NY, USA: Cambridge University Press, 2016. Chap. 2 (cited on pages 11, 15).