Journal of Artificial Intelligence Research 81 (2024) 619-641 Submitted 07/2023; published 11/2024

Truth-tracking with Non-expert Information Sources

Joseph Singleton joesingo@gmail.com
Richard Booth boothr2@cardiff.ac.uk
Cardiff University, UK

Abstract

We study what can be learned when receiving propositional reports from multiple non-expert information sources. We suppose that sources report all that they consider possible, given their expertise. This may result in false and inconsistent reports when sources lack expertise on a topic. A learning method is truth-tracking, roughly speaking, if it eventually converges to correct beliefs about the "actual" world. This involves finding both the actual state of affairs in the domain described by the sources, and finding the extent of the expertise of the sources themselves. We investigate the extent to which truth-tracking is possible, and describe what information can be learned even if the actual world cannot be pinned down uniquely. We find that a broad spread of expertise among the sources allows the actual state of affairs to be found, even if no individual source is an expert on *all* topics. On the other hand, narrower expertise at the individual level allows the actual expertise to be found more easily. Finally, we turn to learning methods themselves: we provide a postulate-based characterisation of truth-tracking for general methods under mild assumptions, before looking at a couple of specific classes of methods from the belief change literature.

1. Introduction

In this paper we study truth-tracking in the logical framework of Singleton and Booth (2022) for reasoning about multiple non-expert information sources. Broadly speaking, the goal of truth-tracking is to find the true state of the world given some input which describes it. In our case this involves finding the true state of some propositional domain about which the sources give reports, and finding the extent of the expertise of the sources themselves.

The general problem of truth-tracking has been studied in various forms across many domains. Perhaps the oldest approach goes back to de Condorcet (1785), whose celebrated *Jury Theorem* states that a majority vote on a yes/no issue will yield the "correct" answer with probability approaching 1 as the number of voters tends to infinity, provided that each voter is more reliable than random choice. This result has since been generalised in many directions (e.g., by Grofman, Owen and Field (1983)). More widely, *epistemic social choice* (Elkind & Slinko, 2016) studies aggregation methods (e.g., voting rules) from the point of finding the "correct" result with high probability, where individual votes are seen as noisy approximations. Of particular relevance to our work is truth-tracking in *judgement aggregation* in social choice (Hartmann & Sprenger, 2012; Terzopoulou & Endriss, 2019), which also takes place in a logical framework. *Belief merging* has close links with judgement aggregation, and generalised jury theorems have been found here too (Everaere, Konieczny, & Marquis, 2010).

In crowdsourcing, the problem of *truth discovery* (Li, Gao, Meng, Li, Su, Zhao, Fan, & Han, 2016) looks at how information from unreliable sources can be aggregated to find the true value of a number of variables, and to find the true reliability level of the sources. This is close to our setting, since incoming information is not always assumed to be reliable, and information about the sources themselves is sought after. Work in this area combines empirical results (e.g., how well methods find the truth on test datasets for which true values are known) and theoretical guarantees, and is typically set in a probabilistic framework.

On the other hand, *formal learning theory* (Jain, Osherson, Royer, & Sharma, 1999) offers a non-probabilistic view on truth-tracking, stemming from the framework of identification in the limit (Gold, 1967). In this paradigm a learner receives an infinite sequence of information step-by-step, such that all true information eventually appears in the sequence. The learner outputs a hypothesis at each step, and aims to stabilise on the correct hypothesis after some finite number of steps. This framework has been combined with belief revision theory (Kelly, Schulte, & Hendricks, 1997; Baltag, Gierasimczuk, & Smets, 2019), ontology learning (Eschenbach & Özcep, 2010) and dynamic epistemic logic (Gierasimczuk, 2009a, 2009b, 2010; Baltag, Gierasimczuk, Özgün, Sandoval, & Smets, 2019). See also (Gierasimczuk, 2023) for a recent overview.

This is the approach we take, and in particular we adapt the truth-tracking setting of Baltag et al. (2019). We apply this to the logical framework of Singleton and Booth (2022). Briefly, this framework extends finite propositional logic with two new notions: that of a source having *expertise* on a formula, and a formula being *sound* for a source to report. Intuitively, expertise on φ means the source has the epistemic capability to distinguish between any pair of φ and $\neg\varphi$ states: they know whether or not φ holds in any state. A formula is sound for a source if it is true *up to their lack of expertise*. For example, if a source has expertise on φ but not ψ, then $\varphi \wedge \psi$ is sound whenever φ holds, since we can ignore the ψ part (on which the source has no expertise). The resulting logical language therefore addresses both the *ontic* facts of the world, through the propositional part, and the *epistemic* state of the sources, via expertise and soundness.

For the most part, formal learning theory supposes that all information received is true, and that all true information is eventually received.[1] This is not a tenable assumption with non-expert sources: some sources may simply lack the expertise to know whether φ is true or false. Instead we make a different (and strong) assumption: all and only *sound* reports are received. Thus, sources report everything consistent with their expertise, which necessitates inconsistent reports from non-experts, since both φ and $\neg\varphi$ will look consistent to a source that lacks expertise to determine whether φ holds. Consequently, the input to our learning methods should be distinguished from the inputs to belief revision and belief merging methods (Alchourrón, Gärdenfors, & Makinson, 1985; Konieczny & Pino Pérez, 2002) – also propositional formulas – which represent *beliefs* of the reporting sources. Indeed, we do not model beliefs of the sources at all.

The following example informally illustrates the core concepts of the logical framework and truth-tracking, and will be returned to throughout the paper.

Example 1. *Consider a medical scenario in which patient A is checked for conditions p and q. By examining A, a doctor* D *has expertise to determine whether A has at least one*

---

[1] But see Section 8.1 of Jain et al. (1999), who consider inaccurate data of various kinds, and Baltag et al. (2019), who consider erroneous reports provided that all errors are eventually corrected.

of p or q, but cannot tell which one(s) without a blood test. A test is only available for p, however, so that the technician T *performing the test has expertise on p but not q.*

Supposing A in fact suffers from q but not p, D *considers each of $p \wedge q$, $\neg p \wedge q$ and $p \wedge \neg q$ possible, whereas* T *considers both $\neg p \wedge q$ and $\neg p \wedge \neg q$ possible. Assuming both sources report* all *they consider possible, their combined expertise leaves $\neg p \wedge q$ as the only possibility. Intuitively, this means we* can *find the true values of p and q in this case.*

Now consider a patient B who suffers from both conditions. D *cannot distinguish A and B, so will provide the same reports, and* T *considers both $p \wedge q$ and $p \wedge \neg q$ possible. In this case* T *is more knowledgable than* D *– since* T *considers fewer situations possible – but we cannot narrow down the true value of q. Thus truth-tracking is only possible for p. The second patient still provides useful information, though, since together with the reports on A,* T*'s lack of expertise tells us all the (in)distinctions between states they are able to make. Namely,* T *cannot distinguish between $p \wedge q$ and $p \wedge \neg q$. Thus we can find the truth about* T*'s expertise.*

Contributions. This paper adapts learning-theoretic notions from formal learning theory - and in particular its intersection with belief revision (Baltag et al., 2019) - to handle non-expert information sources. We establish the limits of learning in this setting, and conditions under which one can learn the true facts of the world as well as the true extent of the expertise of the sources. We go on to characterise truth-tracking learning methods in terms of syntactic postulates, and look specifically at some methods previously introduced by Singleton and Booth (2022).

Paper outline. In Section 2 we outline the logical framework for reasoning about expertise. Section 3 introduces the key concepts of truth-tracking and solvable questions. We characterise solvable questions in Section 4, and explore what they can reveal about the actual world in Section 5. Section 6 looks at learning methods themselves, and characterises truth-tracking methods. We conclude in Section 7.

2. Preliminaries

In this section we recall the logical framework of Singleton and Booth (2022) for reasoning with non-expert sources.

Syntax. Let Prop be a finite set of propositional variables, and let $\mathcal{L}_0$ denote the propositional language generated from Prop. We use $\mathcal{L}_0$ to model the domain underlying the truth-tracking problem; it describes the "ontic" facts of the world, irrespective of the expertise of the sources. Formulas in $\mathcal{L}_0$ will be denoted by lower-case Greek letters (φ, ψ, etc).

Let $\mathcal{S}$ be a finite set of sources. Here we make an important change to the setup of Singleton and Booth (2022): we do not include a special, completely reliable source. Indeed, having access to a completely reliable source of information would somewhat trivialise the truth-tracking problem, at least as far as learning ontic facts is concerned. The language $\mathcal{L}$ extends $\mathcal{L}_0$ with expertise and soundness formulas for each source $i \in \mathcal{S}$, and is defined by the following grammar:

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

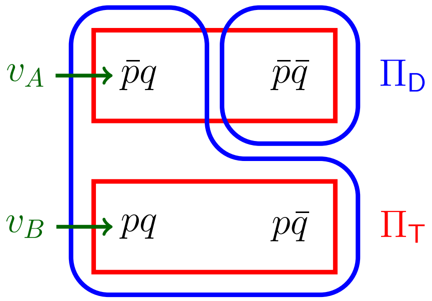

Figure 1: Example of a world W, which formalises Example 1. Here $\mathsf{Prop} = \{p, q\}$, $\mathcal{S} = \{\mathsf{D}, \mathsf{T}\}$ and $\mathcal{C} = \{A, B\}$.

for $\varphi \in \mathcal{L}_0$ and $i \in \mathcal{S}$. Formulas in $\mathcal{L}$ will be denoted by upper-case Greek letters (Φ, Ψ etc). Other logical connectives ($\vee$, $\rightarrow$, $\leftrightarrow$) are introduced as abbreviations. We read $\mathsf{E}_i\varphi$ as "i has expertise on φ", and $\mathsf{S}_i\varphi$ as "φ is sound for i". Note that we restrict the expertise and soundness formulas to propositional arguments, and do not consider nested formulas such as $\mathsf{E}_i\mathsf{S}_j\varphi$.

Semantics. Let $\mathcal{V}$ denote the set of propositional valuations over Prop. We represent the expertise of a source i with a *partition* Π_i of $\mathcal{V}$. Intuitively, this partition represents the distinctions between states the source is able to make: valuations in the same cell in Π_i are indistinguishable to i, whereas i is able to tell apart valuations in different cells. We say i has expertise on φ iff i can distinguish all φ states from $\neg\varphi$ states, and φ is sound for i if the "actual" state (as defined below) is indistinguishable from some φ state.

Let $\mathcal{C}$ be a finite set of *cases*, thought of as independent concrete instances of the domain of interest. For example, the cases in Example 1 are the patients A and B. We consider the expertise of sources to be fixed across all cases.

A *world* is a pair $W = \langle \{v_c\}_{c\in\mathcal{C}}, \{\Pi_i\}_{i\in\mathcal{S}} \rangle$, where

- $v_c \in \mathcal{V}$ is the "actual" valuation for case c;
- $\Pi_i \subseteq 2^{\mathcal{V}}$ is a partition representing the expertise of i.

Let $\mathcal{W}$ denote the set of worlds. Note that $\mathcal{W}$ is finite, since $\mathcal{V}$, $\mathcal{C}$ and $\mathcal{S}$ are. Note it is possible in a world W to have $v_{c_1} = v_{c_2}$ for $c_1 \neq c_2$. For $\varphi \in \mathcal{L}_0$, write $\|\varphi\| \subseteq \mathcal{V}$ for the models of φ, and write $v \Vdash \varphi$ iff $v \in \|\varphi\|$. The consequences of a set $\Gamma \subseteq \mathcal{L}_0$ is denoted by $\mathrm{Cn}_0(\Gamma)$, and we write $\Gamma \Vdash \varphi$ if $\varphi \in \mathrm{Cn}_0(\Gamma)$. For a partition Π, let $\Pi[v]$ denote the unique cell in Π containing v, and write $\Pi[U] = \bigcup_{v\in U} \Pi[v]$ for $U \subseteq \mathcal{V}$. For brevity, we write $\Pi[\varphi]$ instead of $\Pi[\|\varphi\|]$. We evaluate $\mathcal{L}$ formulas with respect to a world W and a case c as follows:

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

where the clauses for conjunction and negation are as standard. The semantics follows the intuition outlined above: $\mathsf{E}_i\varphi$ holds when Π_i separates the φ states from the $\neg\varphi$ states, and $\mathsf{S}_i\varphi$ holds when v_c is indistinguishable from some φ state. Thus, $\mathsf{S}_i\varphi$ means φ is true *up to the expertise of* i: if we weaken φ according to i's expertise, the resulting formula (with models $\Pi_i[\varphi]$) is true.

Note that expertise and soundness are closely related to *S5 knowledge* from epistemic logic. By taking the equivalence relations associated with each partition Π_i, we obtain a (multi-agent) S5 Kripke model, and have the correspondences $\mathsf{S}_i\varphi \equiv \neg\mathsf{K}_i\neg\varphi$ and $\mathsf{E}_i\varphi \equiv \mathsf{A}(\varphi \to \mathsf{K}_i\varphi)$, where K_i denotes knowledge of source i and A is the universal modality (Goranko & Passy, 1992). ($\mathsf{A}\varphi$ is true at a state iff φ is true at *all* states.) This gives expertise and soundness precise interpretations in terms of knowledge; we refer the reader to (Singleton & Booth, 2022; Singleton, 2021) for further discussion.

Example 2. *Take* W *from Fig. 1, which formalises Example 1. Then* $W, c \models \mathsf{E}_{\mathsf{D}}(p \vee q)$ *for all* $c \in \mathcal{C}$, *since* $\|p \vee q\|$ *is a cell in* Π_{D}. *We also have* $W, A \models \neg p \wedge \mathsf{S}_{\mathsf{D}}p$, *i.e., patient* A *does not suffer from condition* p, *but it is consistent with* D*'s expertise that they do.*

We write $W, c \models \Gamma$, for a set of formulas $\Gamma \subseteq \mathcal{L}$, if $W, c \models \Phi$ for all $\Phi \in \Gamma$. For a set $S \subseteq \mathcal{W}$, we write $S, c \models \Phi$ iff $W, c \models \Phi$ for all $W \in S$.

Reports. A *report* is a triple $\langle i, c, \varphi\rangle$, where $i \in \mathcal{S}$, $c \in \mathcal{C}$ and $\varphi \in \mathcal{L}_0$ with $\varphi \not\equiv \bot$. In this paper, we interpret such triples as source i reporting that φ is possible in case c. An *input sequence* σ is a finite sequence of reports.

A *method* L maps each input sequence σ to a set of worlds $L(\sigma) \subseteq \mathcal{W}$, called the *conjecture* of L on σ.[2] We say L *implies* $S \subseteq \mathcal{W}$ on the basis of σ if $L(\sigma) \subseteq S$. L is *consistent* if $L(\sigma) \neq \emptyset$ for all input sequences σ.

3. Truth-Tracking

We adapt the framework for truth-tracking from (Gierasimczuk, 2010; Baltag, Gierasimczuk, & Smets, 2015; Baltag et al., 2019), which finds its roots in formal learning theory. In this framework, a learning method receives increasing initial segments of an infinite sequence – called a *stream* – which enumerates all (and only) the true propositions observable at the "actual" world. Truth-tracking requires the method to eventually find the actual world (or some property thereof), given *any* stream.

As mentioned in the introduction, in our setting we cannot assume the sources themselves report only true propositions. Instead, our streams will enumerate all the *sound* reports. Thus, a stream may include false reports, but such false reports only arise due to lack of expertise of the corresponding source.[3] Moreover, *all* sound reports will eventually arise. Since $\mathsf{S}_i\varphi$ means φ is possible from the point of view of i's expertise, we can view a stream as each source sharing *all that they consider possible* for each case $c \in \mathcal{C}$. In particular, a non-expert source may report both φ and $\neg\varphi$ for the same case.

---

[2] We depart from the original framework of Singleton and Booth (2022) here by taking a semantic view of belief change operators, with the output a set of worlds instead of formulas.

[3] Alternatively, we can consider statements of the form "φ is sound for i in case c" as a higher-order "proposition"; a stream then enumerates all true propositions of this kind.

Definition 1. *An infinite sequence of reports ρ is a* stream *for W iff for all i, c, φ*[4]*:*

$$\langle i, c, \varphi \rangle \in \rho \iff W, c \models \mathsf{S}_i \varphi.$$

We refer to the left-to-right implication as *soundness* of ρ for W, and the right-to-left direction as *completeness*. Note that every world W has some stream: the set $\{\langle i, c, \varphi \rangle \mid W, c \models \mathsf{S}_i \varphi\}$ is countable, so can be indexed by $\mathbb{N}$ to form a stream. For $n \in \mathbb{N}$ we let ρ_n denote the n-th report in ρ, and write $\rho[n]$ for the finite initial segment of ρ of length n.

Example 3. *Consider W from Fig. 1 and case A. From the point of view of D's expertise, the "actual" valuation could be pq, $\bar{p}q$, $p\bar{q}$. Consequently, in a stream for W, D will report p, $\neg p$, q, $\neg q$, $p \vee q$, and so on. A report that D will* not *give is $\neg(p \vee q)$, since D has expertise to know this is false.*

Note that v_A and v_B are indistinguishable to D, so the reports of D in any stream will be the same for both cases. In contrast, T can *distinguish the two cases, and will report $\neg p$ in case A but not in B, and p in case B but not in A.*

In line with the notion originally given by Baltag et al. (2015) we define a *question Q* to be a partition of $\mathcal{W}$. That is, a question is a set of disjoint *answers* $A \in Q$, with each world W appearing in a unique cell $Q[W]$ – the correct answer at W.

Example 4. *We consider some example questions.*

1. *Any formula $\Phi \in \mathcal{L}$ and case c defines a question $Q_{\Phi,c}$, whose two cells consist of the worlds satisfying Φ, respectively $\neg\Phi$, in case c. Intuitively, this question asks whether Φ is true or false in case c.*
2. *The finest question $Q_\perp = \{\{W\} \mid W \in \mathcal{W}\}$ asks: what is the "actual" world?*
3. *More generally, for any set X and function $f : \mathcal{W} \to X$, the equivalence relation given by $W \simeq_f W'$ iff $f(W) = f(W')$ defines a question Q_f.*

   In this way any data associated with a world gives rise to a question. For example, if $f(W) = \{i \in \mathcal{S} \mid \Pi_i^W[p] = \|p\|\}$ we ask for the set of sources with expertise on p; if $f(W) = |\{c \in \mathcal{C} \mid W, c \models p\}|$ we ask for the number of cases where p holds, etc.

   In fact, all questions are of this form: given Q we may define $f : \mathcal{W} \to Q$ by $f(W) = Q[W]$; then $Q_f = Q$.

A method solves Q if it eventually implies the correct answer when given any stream.

Definition 2. *A method L* solves *a question Q if for all worlds W and all streams ρ for W, there is $n \in \mathbb{N}$ such that $L(\rho[m]) \subseteq Q[W]$ for all $m \geq n$. A question Q is* solvable *if there is some consistent method L which solves Q.*

Note that we do not require $W \in L(\rho[m])$. Solvability can be also expressed in terms of eliminating incorrect worlds.

---

[4] By abuse of notation, we use set membership for a sequence ρ to mean that the element appears somewhere in the sequence.

Proposition 1. *A method L solves Q if and only if for all W, all streams ρ for W, and all $W' \notin Q[W]$, there is $n_{W'} \in \mathbb{N}$ such that $W' \notin L(\rho[m])$ for all $m \geq n_{W'}$.*

Proof. "if": Taking $n = \max\{n_{W'} \mid W' \notin Q[W]\}$, which exists since $\mathcal{W}$ is finite, $L(\rho[m]) \subseteq Q[W]$ for $m \geq n$.

"only if": Taking n from the definition of L solving Q, we may simply take $n_{W'} = n$ for all $W' \notin Q[W]$. □

4. Characterising Solvable Questions

In this section we explore solvability of questions, finding that there is a unique "hardest" question which subsumes all solvable questions. We show this is itself solvable, and thus obtain a precise characterisation of solvability.

Following Baltag et al. (2015), questions are partially ordered by partition refinement: $Q \preceq Q'$ iff each $A' \in Q'$ can be written as a union of answers from Q. Equivalently, $Q[W] \subseteq Q'[W]$ for all W. This can be interpreted as a *difficulty ordering*: if $Q \preceq Q'$ then each answer of Q' is just a disjunction of answers of Q, and thus Q' is *easier* than Q. Naturally, if Q is solvable then so too is any easier question.

Proposition 2. *If Q is solvable and $Q \preceq Q'$, then Q' is solvable.*

Proof. The method which solves Q also solves Q'. □

Since question solving is based on streams of sound reports, worlds satisfying the same soundness statements cannot be distinguished by any solvable question. To formalise this, define a preorder $\sqsubseteq$ on $\mathcal{W}$ by

$$W \sqsubseteq W' \iff \forall i, c, \varphi : \ (W, c \models \mathsf{S}_i\varphi \implies W', c \models \mathsf{S}_i\varphi).$$

Thus, $W \sqsubseteq W'$ iff any report sound for W is also sound for W'. We denote by $\sqsubset$ and $\approx$ the strict and symmetric parts of $\sqsubseteq$, respectively.

Lemma 1. *$W \sqsubseteq W'$ if and only if for all $i \in \mathcal{S}$ and $c \in \mathcal{C}$, $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$.*

Proof. "if": Suppose $W, c \models \mathsf{S}_i\varphi$. Then $v_c^W \in \Pi_i^W[\varphi]$, so there is $u \in \|\varphi\|$ such that $v_c^W \in \Pi_i^W[u]$. Consequently $u \in \Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$, which means $v_c^{W'} \in \Pi_i^{W'}[u] \subseteq \Pi_i^{W'}[\varphi]$. Hence $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$.

"only if": Let $u \in \Pi_i^W[v_c^W]$. Let φ be any formula with $\|\varphi\| = \{u\}$. Then $W, c \models \mathsf{S}_i\varphi$, so $W \sqsubseteq W'$ gives $W', c \models \mathsf{S}_i\varphi$, i.e., $v_c^{W'} \in \Pi_i^{W'}[u]$, so $u \in \Pi_i^{W'}[v_c^{W'}]$. Hence $\Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$. □

Note that $\Pi_i[v_c]$ is the set of valuations indistinguishable from the "actual" valuation in case c, for source i. In light of Lemma 1, we can interpret $W \sqsubseteq W'$ as saying that all sources are *more knowledgeable* in each case c in world W than in W'. However, $W \sqsubseteq W'$ does not say anything about the partition cells not containing some v_c.

Proposition 3. *The following are equivalent.*

1. *W and W' have exactly the same streams.*

2. $W \approx W'$.

3. *For all* $i \in \mathcal{S}$ *and* $c \in \mathcal{C}$, $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$.

Proof. (2) and (3) are easily seen to be equivalent in light of Lemma 1. To show (1) is equivalent to (2), first suppose W and W' have the same streams, and suppose $W, c \models \mathsf{S}_i\varphi$. Taking an arbitrary stream ρ for W, completeness gives $\langle i, c, \varphi\rangle \in \rho$. But ρ is a stream for W' too, and soundness gives $W', c \models \mathsf{S}_i\varphi$. Hence $W \sqsubseteq W'$. A symmetrical argument shows $W' \sqsubseteq W$.

On the other hand, if $W \approx W'$ then W and W' satisfy exactly the same soundness statements, so it is clear that any sequence ρ is a stream for W iff it is a stream for W'. □

Since it will play a special role throughout, we denote by Q^* the question formed by the equivalence relation $\approx$. Then $Q^*[W]$ is the set of W' with $W \approx W'$. Since no solvable question can distinguish $\approx$-equivalent worlds, we have the following.

Lemma 2. *If* Q *is solvable then* $Q^* \preceq Q$.

Proof. Suppose L is a consistent method solving Q. We show $Q^*[W] \subseteq Q[W]$ for all W. Indeed, let $W' \in Q^*[W]$. Then $W' \approx W$. Taking any stream ρ for W, there is n such that $L(\rho[m]) \subseteq Q[W]$ for $m \geq n$. On the other hand ρ is also a stream for W' by Proposition 3, so there is n' such that $L(\rho[m]) \subseteq Q[W']$ for $m \geq n'$. Setting $m = \max\{n, n'\}$ and using the fact that L is consistent, we find $\emptyset \subset L(\rho[m]) \subseteq Q[W] \cap Q[W']$. Since Q is a partition, this means $Q[W] = Q[W']$, i.e., $W' \in Q[W]$. □

So, any solvable question is coarser than Q^*. Fortunately, Q^* itself is solvable since we work in a finite framework. For a sequence σ, write $\mathcal{X}_\sigma^{\mathsf{snd}}$ for the set of worlds W such that $W, c \models \mathsf{S}_i\varphi$ for all $\langle i, c, \varphi\rangle \in \sigma$. To solve Q^* it suffices to conjecture the $\sqsubseteq$-minimal worlds in $\mathcal{X}_\sigma^{\mathsf{snd}}$.

Proposition 4. Q^* *is solvable.*

Proof. Set $L(\sigma) = \min_\sqsubseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ if $\mathcal{X}_\sigma^{\mathsf{snd}} \neq \emptyset$, and $L(\sigma) = \mathcal{W}$ otherwise (where $W \in \min_\sqsubseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ iff $W \in \mathcal{X}_\sigma^{\mathsf{snd}}$ and there is no $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ with $W' \sqsubset W$). Note that L is consistent since $\mathcal{W}$ is finite and non-empty. We show that L solves Q^* by Proposition 1. Take any world W and a stream ρ. First note that, by soundness of ρ, $W \in \mathcal{X}_{\rho[n]}^{\mathsf{snd}}$ for all $n \in \mathbb{N}$, so we are always in the first case in the definition of L.

Take $W' \notin Q^*[W]$. Then $W \not\approx W'$. Consider two cases:

- **Case 1:** $W \not\sqsubseteq W'$. By definition, there are i, c, φ such that $W, c \models \mathsf{S}_i\varphi$ but $W', c \not\models \mathsf{S}_i\varphi$. By completeness of ρ for W, there is n such that $\rho_n = \langle i, c, \varphi\rangle$. Consequently $W' \notin \mathcal{X}_{\rho[m]}^{\mathsf{snd}}$ for all $m \geq n$. Since $L(\rho[m]) \subseteq \mathcal{X}_{\rho[m]}^{\mathsf{snd}}$, we have $W' \notin L(\rho[m])$ as required.

- **Case 2:** $W \sqsubset W'$. Since $W \in \mathcal{X}_{\rho[n]}^{\mathsf{snd}}$ for all n, W' can never be $\sqsubseteq$-minimal. Thus $W' \notin L(\rho[n])$ for all n.

Note that these cases are exhaustive since $W \not\approx W'$. This completes the proof. □

Putting Propositions 2 and 4 and Lemma 2 together we obtain a characterisation of solvable questions.

Theorem 1. *Q is solvable if and only if $Q^* \preceq Q$.*

Given this result, Q^* is the only question that really matters: any other question is either unsolvable or formed by coarsening Q^*. With this in mind, we make the following definition.

Definition 3. *A method is* truth-tracking *if it solves Q^*.*

Example 5. *We refer back to the questions of Example 4.*

1. *The question $Q_{\varphi,c}$, for any propositional formula $\varphi \in \mathcal{L}_0$, is solvable if and only if either φ is a tautology or a contradiction. To see the "only if" part, consider the contrapositive. For any contingent formula φ, take worlds W_1, W_2 where no source has any expertise (i.e., $\Pi_i^{W_k} = \{\mathcal{V}\}$ for $k \in \{1, 2\}$) but where $v_c^{W_1} \Vdash \varphi$, $v_c^{W_2} \Vdash \neg\varphi$. Then $W_1 \approx W_2$ (e.g., by Proposition 3) but $W_1 \notin Q_{\varphi,c}[W_2]$.*

   Similarly, $Q_{\mathsf{E}_i\varphi,c}$ is solvable iff either φ is a tautology or contradiction, when $|\mathsf{Prop}| \geq 2$.

2. *The finest question $Q_\perp$ is not solvable, since there are always distinct W, W' with $W \approx W'$.*

3. *In general, Q_f is solvable iff $W \approx W'$ implies $f(W) = f(W')$, i.e., iff f takes a unique value on each equivalence class of $\approx$.*

5. What Information can be Learned?

Solving a question Q has a *global* character: we must find the correct answer $Q[W]$ starting from *any* world W. As we saw in Example 5, this rules out the possibility of solving many interesting questions due to the presence of "abnormal" worlds (e.g., those in which no sources have any expertise). In this section we take a more fine-grained approach by looking *locally*: given some *particular* world W, what can we learn about W via truth-tracking methods? Concretely, what properties of W are uniquely defined across $Q^*[W]$? For instance, in Example 1 we took this local perspective and argued that in the particular world W modelling the medical scenario, it is possible to find the true value of q for patient A but not for B. The results and examples of this section will formalise this informal argument. There are two components of a world that need to be identified: the cases (i.e., the valuations) and the expertises (i.e., the partitions). The former will be addressed in Section 5.1, the latter in Section 5.2. The two will be combined in Section 5.3.

In general, the extent to which one can learn depends on W. If no sources have expertise then source partitions are uniquely defined (since *all* consistent formulas are sound, and only the trivial partitions have this property), but any combination of valuations is possible, since if $\Pi_i^W = \{\mathcal{V}\}$ for all $i \in \mathcal{S}$ then $W \approx W'$ for every other W' such that $\Pi_i^{W'} = \{\mathcal{V}\}$ for all $i \in \mathcal{S}$. On the other hand if all sources have total expertise then valuations are uniquely defined, but there may not be enough cases to uniquely identify the source partitions. Of particular interest is the case where $Q^*[W]$ contains only W; starting in such a world, truth-tracking methods are able to find the true world exactly.

In what follows, say $S \subseteq \mathcal{W}$ *decides* Φ *in case* c iff either $S, c \models \Phi$ or $S, c \models \neg\Phi$. That is, the truth value of Φ in case c is unambiguously defined across the worlds in S. If Φ does not depend on the case (e.g., if $\Phi = \mathsf{E}_i\varphi$) we simply say S decides Φ.

5.1 Valuations

We start by considering when $Q^*[W]$ decides a propositional formula φ in case c, i.e., when truth-tracking methods are guaranteed to successfully determine whether or not φ holds in the "actual" world. This leads to a precise characterisation of when $Q^*[W]$ contains a *unique* valuation in case c, so that v_c^W can be found exactly.

We need a notion of *group expertise.* For $\mathcal{S}' \subseteq \mathcal{S}$ and $\Gamma \subseteq \mathcal{L}_0$, write $W \models \mathsf{E}_{\mathcal{S}'}\Gamma$ if for each $\psi \in \Gamma$ there is $i \in \mathcal{S}'$ such that $W \models \mathsf{E}_i\psi$. Then the group $\mathcal{S}'$ has expertise on Γ in a collective sense, even if no single source has expertise on *all* formulas in Γ. We have that φ is decided if $\mathcal{S}$ have group expertise on a set of true formulas $\Gamma \subseteq \mathcal{L}_0$ such that either $\Gamma \Vdash \varphi$ or $\Gamma \Vdash \neg\varphi$.

Theorem 2. *$Q^*[W]$ decides $\varphi \in \mathcal{L}_0$ in case c if and only if there is $\Gamma \subseteq \mathcal{L}_0$ such that (i) $W, c \models \Gamma$; (ii) $W \models \mathsf{E}_{\mathcal{S}}\Gamma$; and (iii) either $\Gamma \Vdash \varphi$ or $\Gamma \Vdash \neg\varphi$.*

$Q^*[W]$ decides *all* propositional formulas – and thus determines the c-valuation v_c^W exactly – iff $\mathcal{S}$ have group expertise on a maximally consistent set of true formulas. For $S \subseteq \mathcal{W}$ and $c \in \mathcal{C}$, write $\mathcal{V}_c^S = \{v_c^W \mid W \in S\}$ for the c-valuations appearing in S.

Theorem 3. *The following are equivalent.*

1. $\mathcal{V}_c^{Q^*[W]} = \{v_c^W\}$.
2. *$Q^*[W]$ decides φ in case c, for all $\varphi \in \mathcal{L}_0$.*
3. *There is $\Gamma \subseteq \mathcal{L}_0$ such that (i) $W, c \models \Gamma$; (ii) $W \models \mathsf{E}_{\mathcal{S}}\Gamma$; and (iii) $\mathrm{Cn}_0(\Gamma)$ is a maximally consistent set.*

We illustrate Theorem 3 with an example.

Example 6. *Consider W from Fig. 1. Then one can show $\mathcal{V}_A^{Q^*[W]} = \{\bar{p}q\} = \{v_A^W\}$, and $\mathcal{V}_B^{Q^*[W]} = \{pq, p\bar{q}\} \neq \{v_B^W\}$. That is, W's A valuation is uniquely determined by truth-tracking methods, but its B valuation is not: there is some world $W' \approx W$ whose B-valuation differs from W's. This matches the informal reasoning in Example 1, in which patient A could be successfully diagnosed on both p and q but B could not.*

Formally, take $\Gamma = \{p \vee q, \neg p\}$. Then $W, A \models \Gamma$, $W \models \mathsf{E}_{\mathcal{S}}\Gamma$ (since D has expertise on $p \vee q$ and T has expertise on $\neg p$), and $\mathrm{Cn}_0(\Gamma) = \mathrm{Cn}_0(\neg p \wedge q)$, which is maximally consistent. This example shows how the expertise of multiple sources can be combined to find valuations uniquely, but that this is not necessarily possible in all cases.

The remainder of this section proves Theorems 2 and 3.

Lemma 3. *For $W \approx W'$, $i \in \mathcal{S}$ and $\varphi \in \mathcal{L}_0$,*

$$W, c \models \varphi \wedge \mathsf{E}_i\varphi \implies W', c \models \varphi.$$

Proof. From $W, c \models \varphi$ we have $v_c^W \in \|\varphi\|$, so $\Pi_i^W[v_c^W] \subseteq \Pi_i^W[\varphi]$. But $W, c \models \mathsf{E}_i\varphi$ means $\Pi_i^W[\varphi] = \|\varphi\|$, so in fact $\Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Now using $W \approx W'$, we find $v_c^{W'} \in \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Hence $W', c \models \varphi$. □

Lemma 4. $\mathcal{V}_c^{Q^*[W]} = \bigcap_{i \in \mathcal{S}} \Pi_i^W[v_c^W]$.

Proof. "$\subseteq$": Suppose $u \in \mathcal{V}_c^{Q^*[W]}$. Then there is $W' \approx W$ such that $u = v_c^{W'}$. Let $i \in \mathcal{S}$. Then $u \in \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W]$ by Proposition 3, as required.

"$\supseteq$": Suppose $u \in \bigcap_{i \in \mathcal{S}} \Pi_i^W[v_c^W]$. Let W' be the world obtained from W by setting the c-valuation to u, keeping partitions and other valuations the same. We need to show $W' \approx W$. We do so via Proposition 3, by showing condition (3). Take any $i \in \mathcal{S}$ and $d \in \mathcal{C}$. If $d \neq c$ then $v_d^{W'} = v_d^W$; since partitions are the same in W' as in W we get $\Pi_i^W[v_d^W] = \Pi_i^{W'}[v_d^{W'}]$. For $c = d$, note $\Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[u]$. By assumption $u \in \Pi_i^W[v_c^W]$, so $\Pi_i^W[u] = \Pi_i^W[v_c^W]$. Hence $\Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W]$ as required. $\square$

Proof of Theorem 2. "if": Take $W' \in Q^*[W]$. Note that since $W, c \models \Gamma$ and $W, c \models \mathsf{E}_{\mathcal{S}}\Gamma$, we may apply Lemma 3 to each formula in Γ in turn to find $W', c \models \Gamma$. Now, if $W, c \models \varphi$ then we must have $\Gamma \Vdash \varphi$, so $W', c \models \varphi$ too. Otherwise $W, c \not\models \varphi$, so we must have $\Gamma \Vdash \neg\varphi$ and $W', c \not\models \varphi$. This shows $W', c \models \varphi$ if and only if $W, c \models \varphi$. Since $W' \in Q^*[W]$ was arbitrary, $Q^*[W]$ decides φ in case c.

"only if": Suppose $Q^*[W]$ decides φ in case c. For each $i \in \mathcal{S}$, take some $\psi_i \in \mathcal{L}_0$ such that $\|\psi_i\| = \Pi_i^W[v_c^W]$. Then $W \models \mathsf{E}_i\psi_i$. Set $\Gamma = \{\psi_i\}_{i \in \mathcal{S}}$. Clearly $W, c \models \Gamma$ and $W \models \mathsf{E}_{\mathcal{S}}\Gamma$. Now, take any $u \in \|\Gamma\|$. By Lemma 4, $\|\Gamma\| = \bigcap_{i \in \mathcal{S}} \Pi_i^W[v_c^W] = \mathcal{V}_c^{Q^*[W]}$. Hence there is some $W' \in Q^*[W]$ such that $u = v_c^{W'}$. But $Q^*[W]$ decides φ in case c, so $W', c \models \varphi$ iff $W, c \models \varphi$. Thus $u \Vdash \varphi$ iff $W, c \models \varphi$. Since $u \in \|\Gamma\|$ was arbitrary, we have $\Gamma \Vdash \varphi$ if $W, c \models \varphi$, and $\Gamma \Vdash \neg\varphi$ otherwise. $\square$

Proof of Theorem 3. (1) implies (2): If $W' \in Q^*[W]$ then W and W' share the same c-valuation by (1), so clearly $W, c \models \varphi$ iff $W', c \models \varphi$, for any φ. Hence $Q^*[W]$ decides φ in case c.

(2) implies (1): Clearly $v_c^W \in \mathcal{V}_c^{Q^*[W]}$. Suppose $u \in \mathcal{V}_c^{Q^*[W]}$. Then there is $W' \in Q^*[W]$ such that $u = v_c^{W'}$. Let $p \in \mathsf{Prop}$. Since $W, W' \in Q^*[W]$ and $Q^*[W]$ decides p in case c, we have $u \Vdash p$ iff $v_c^W \Vdash p$. Since p was arbitrary, $u = v_c^W$.

(2) implies (3): Applying Theorem 2 to each $\varphi \in \mathcal{L}_0$, there is a set $\Gamma_\varphi \subseteq \mathcal{L}_0$ such that $W, c \models \Gamma_\varphi$, $W \models \mathsf{E}_{\mathcal{S}}\Gamma_\varphi$, and either $\Gamma_\varphi \Vdash \varphi$ or $\Gamma_\varphi \Vdash \neg\varphi$. Set $\Gamma = \bigcup_{\varphi \in \mathcal{L}_0} \Gamma_\varphi$. Clearly $W, c \models \Gamma$ – so Γ is consistent – and $W \models \mathsf{E}_{\mathcal{S}}\Gamma$. To show $\mathrm{Cn}_0(\Gamma)$ is *maximally* consistent, suppose $\varphi \notin \mathrm{Cn}_0(\Gamma)$. From monotonicity of classical consequence and $\Gamma_\varphi \subseteq \Gamma$, we get $\varphi \notin \mathrm{Cn}_0(\Gamma_\varphi)$. Hence $\Gamma_\varphi \Vdash \neg\varphi$, and $\Gamma \Vdash \neg\varphi$ too. This means $\mathrm{Cn}_0(\Gamma) \cup \{\varphi\}$ is inconsistent, and we are done.

(3) implies (2): Take $\varphi \in \mathcal{L}_0$. Then we may apply Theorem 2 with Γ from (3) – noting that the maximal consistency property ensure either $\Gamma \Vdash \varphi$ or $\Gamma \models \neg\varphi$ – to see that $Q^*[W]$ decides φ in case c. $\square$

5.2 Source Partitions

We now consider when $Q^*[W]$ decides an expertise formula $\mathsf{E}_i\varphi$, i.e., when truth-tracking methods are guaranteed to successfully determine the expertise of source i. This leads to a precise characterisation of when $Q^*[W]$ contains a *unique* partition for source i. We apply the analysis of the previous section to the set of source partitions $\{\Pi_i^W\}_{i \in \mathcal{S}}$. For $S \subseteq \mathcal{W}$ and $i \in \mathcal{S}$, write $\mathcal{P}_i^S = \{\Pi_i^W \mid W \in S\}$ for the i-partitions appearing in S. When $S = Q^*[W]$,

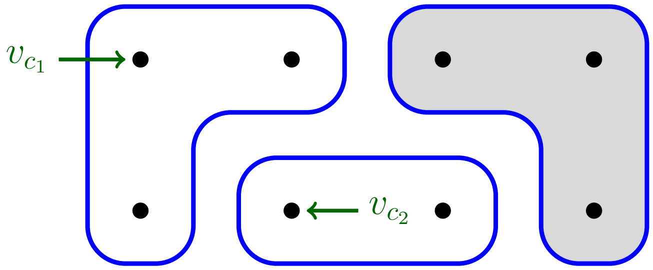

Figure 2: World W from Example 7. Note that for brevity we do not label the valuations.

these are exactly those partitions which agree with Π_i^W at each valuation v_c^W. In other words:

Lemma 5. *$\Pi \in \mathcal{P}_i^{Q^*[W]}$ if and only if $\{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}} \subseteq \Pi$.*

Proof. "if": Suppose $\{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}} \subseteq \Pi$. Let W' be obtained from W by setting i's partition to Π, keeping valuations and other source partitions the same. We claim $W' \approx W$. Indeed, take any $j \in \mathcal{S}$ and $c \in \mathcal{C}$. If $j \neq i$ then $\Pi_j^{W'} = \Pi_i^W$; since valuations are the same we get $\Pi_j^W[v_c^W] = \Pi_j^{W'}[v_c^{W'}]$. For $j = i$, note that since $\Pi_i^W[v_c^W] \in \Pi$ by assumption, and $v_c^W \in \Pi_i^W[v_c^W]$, we have $\Pi[v_c^W] = \Pi_i^W[v_c^W]$. By construction of W', this means $\Pi_i^W[v_c^W] = \Pi[v_c^{W'}] = \Pi_i^{W'}[v_c^{W'}]$. By Proposition 3, $W' \approx W$. Hence $\Pi \in \mathcal{P}_i^{Q^*[W]}$.

"only if": This is clear from Proposition 3. □

Example 7. *Suppose $|\mathsf{Prop}| = 3$, $\mathcal{C} = \{c_1, c_2\}$ and $i \in \mathcal{S}$. Consider a world W whose i-partition is shown in Fig. 2. By Lemma 5, a partition Π appears as $\Pi_i^{W'}$ for some $W' \approx W$ if and only if it contains the leftmost and bottommost sets. Any such Π consists of these cells together with a partition of the shaded area. Since there are 5 possible partitions of a 3-element set, it follows that $|\mathcal{P}_i^{Q^*[W]}| = 5$.*

Example 7 hints that if the cells containing the valuations v_c^W cover the whole space of valuations $\mathcal{V}$, or just omit a single valuation, then i's partition is uniquely defined in $Q^*[W]$. That is, truth-tracking methods can determine the full extent of i's expertise if the "actual" world is W. Indeed, we have the following analogue of Theorem 3 for partitions.

Theorem 4. *The following are equivalent.*

1. $\mathcal{P}_i^{Q^*[W]} = \{\Pi_i^W\}$.
2. $Q^*[W]$ *decides* $\mathsf{E}_i\varphi$ *for all* $\varphi \in \mathcal{L}_0$.
3. $|\mathcal{V} \setminus R| \leq 1$, *where* $R = \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]$.

Note that $R = \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]$ is the set of valuations indistinguishable from the actual state at some case c. Clause 3 of Theorem 4 says this set needs to essentially cover the whole space $\mathcal{V}$, omitting at most a single point. In this sense, it is easier to find Π_i^W uniquely when i has *less expertise*, since the cells $\Pi_i^W[v_c^W]$ will be larger. In the extreme case where i has total expertise, i.e., $\Pi_i^W = \{\{v\} \mid v \in \mathcal{V}\}$, we need at least $2^{|\mathsf{Prop}|} - 1$ cases with distinct valuations in order to find Π_i^W exactly.

Example 8. *In Example 7 we have already seen a world W for which $\mathcal{P}_i^{Q^*[W]}$ does not contain a unique partition. For a positive example, consider the world W from Fig. 1. Then $\mathcal{V} \setminus R_{\mathsf{D}} = \{\bar{p}\bar{q}\}$ and $\mathcal{V} \setminus R_{\mathsf{T}} = \emptyset$, so both the partitions of D and T can be found uniquely by truth-tracking methods.*

The remainder of this section proves Theorem 4.

Lemma 6. *Let $i \in \mathcal{S}$ and $U \subseteq \mathcal{V}$. Then $U \subseteq \bigcup_{c \in \mathcal{C}} \Pi_i^W[v_c^W]$ and $W \approx W'$ implies $\Pi_i^W[U] = \Pi_i^{W'}[U]$.*

Proof. It suffices to show that for all $u \in U$ we have $\Pi_i^W[u] = \Pi_i^{W'}[u]$, since by definition $\Pi[U] = \bigcup_{u \in U} \Pi[u]$. Let $u \in U$. Then there is $c \in \mathcal{C}$ such that $u \in \Pi_i^W[v_c^W]$. Hence $\Pi_i^W[u] = \Pi_i^W[v_c^W]$. But since $W \approx W'$, $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$. This means $u \in \Pi_i^{W'}[v_c^{W'}]$, so $\Pi_i^{W'}[u] = \Pi_i^{W'}[v_c^{W'}] = \Pi_i^W[v_c^W] = \Pi_i^W[u]$, as required, □

Lemma 7. *$Q^*[W]$ decides $\mathsf{E}_i\varphi$ if and only if, writing $R = \bigcup_{c \in \mathcal{C}} \Pi_i^W[v_c^W]$, either (i) $\|\varphi\| \subseteq R$; (ii) $\|\neg\varphi\| \subseteq R$; or (iii) there is some $c \in \mathcal{C}$ such that $\Pi_i^W[v_c^W]$ intersects with both $\|\varphi\|$ and $\|\neg\varphi\|$.*

Proof. "if": First suppose (i) holds. Take $W' \in Q^*[W]$. From $\|\varphi\| \subseteq R$, $W \approx W'$ and Lemma 6 we get $\Pi_i^W[\varphi] = \Pi_i^{W'}[\varphi]$. Consequently, $W' \models \mathsf{E}_i\varphi$ iff $W \models \mathsf{E}_i\varphi$. Since W' was arbitrary, either all worlds in $Q^*[W]$ satisfy $\mathsf{E}_i\varphi$, or all do not. Hence $Q^*[W]$ decides $\mathsf{E}_i\varphi$.

If (ii) holds, a similar argument shows that $Q^*[W]$ decides $\mathsf{E}_i\neg\varphi$. But it is easily checked that $\mathsf{E}_i\varphi \equiv \mathsf{E}_i\neg\varphi$, so $Q^*[W]$ also decides $\mathsf{E}_i\varphi$.

Finally, suppose (iii) holds. Then there is $c \in \mathcal{C}$ and $u \in \|\varphi\|$, $v \in \|\neg\varphi\|$ such that $u, v \in \Pi_i^W[v_c^W]$. We claim $Q^*[W] \models \neg\mathsf{E}_i\varphi$. Indeed, take $W' \in Q^*[W]$. Then $\Pi_i^W[v_c^W] = \Pi_i^{W'}[v_c^{W'}]$, so $u, v \in \Pi_i^{W'}[v_c^{W'}]$. In particular, u and v differ on φ but are contained in the same cell in $\Pi_i^{W'}$. Hence $W' \models \neg\mathsf{E}_i\varphi$.

"only if": We show the contrapositive. Suppose none of (i), (ii), (iii) hold. Then there is $u \in \|\varphi\| \setminus R$ and $v \in \|\neg\varphi\| \setminus R$. Let us define two worlds W_1, W_2 from W by modifying i's partition:

$$\begin{aligned}\Pi_i^{W_1} &= \{\Pi_i^W[v_c^W]\}_{c \in \mathcal{C}} \cup \{\mathcal{V} \setminus R\},\\ \Pi_i^{W_2} &= \{\Pi_i^W[v_c^W]\}_{c \in \mathcal{C}} \cup \{\{w\} \mid w \in \mathcal{V} \setminus R\}.\end{aligned}$$

Then $W_1, W_2 \in Q^*[W]$ by Lemma 5. We claim that $W_1 \models \neg\mathsf{E}_i\varphi$ but $W_2 \models \mathsf{E}_i\varphi$, which will show $Q^*[W]$ does not decide $\mathsf{E}_i\varphi$.

First, note that since $u, v \notin R$, we have $\Pi_i^{W_1}[u] = \Pi_i^{W_1}[v] = \mathcal{V} \setminus R$. Since u and v differ on φ but share the same partition cell, $W_1 \models \neg\mathsf{E}_i\varphi$.

To show $W_2 \models \mathsf{E}_i\varphi$, take $w \in \|\varphi\|$. If $w \notin R$ then $\Pi_i^{W_2}[w] = \{w\} \subseteq \|\varphi\|$. Otherwise there is $c \in \mathcal{C}$ such that $w \in \Pi_i^W[v_c^W]$. Thus $\Pi_i^W[v_c^W]$ intersects with $\|\varphi\|$. Since (iii) does not hold, this in fact implies $\Pi_i^W[v_c^W] \subseteq \|\varphi\|$, and consequently $\Pi_i^{W_2}[w] = \Pi_i^W[v_c^W] \subseteq \|\varphi\|$. Since $w \in \|\varphi\|$ was arbitrary, we have shown $\Pi_i^{W_2}[\varphi] = \bigcup_{w \in \|\varphi\|} \Pi_i^{W_2}[w] \subseteq \|\varphi\|$. Since the reverse inclusion always holds, this shows $W_2 \models \mathsf{E}_i\varphi$, and we are done. □

Proof of Theorem 4. The implication (1) to (2) is clear since if $W' \in Q^*[W]$ then $\Pi_i^{W'} = \Pi_i^W$ by (1), so $W' \models \mathsf{E}_i\varphi$ iff $W \models \mathsf{E}_i\varphi$, and thus $Q^*[W]$ decides $\mathsf{E}_i\varphi$.

To show (2) implies (3) we show the contrapositive. Suppose $|\mathcal{V} \setminus R| > 1$. Then there are distinct $u, v \in \mathcal{V} \setminus R$. Let φ be any propositional formula with $\|\varphi\| = \{u\}$. We show by Lemma 7 that $Q^*[W]$ does not decide $\mathsf{E}_i\varphi$. Indeed, all three conditions fail: $\|\varphi\| \not\subseteq R$ (since $u \notin R$), $\|\neg\varphi\| \not\subseteq R$ (since $v \in \|\neg\varphi\| \setminus R$) and no $\Pi_i^W[v_c^W]$ intersects with $\|\varphi\|$ (otherwise $u \in \Pi_i^W[v_c^W] \subseteq R$).

Finally, for (3) implies (1) we also show the contrapositive. Suppose there is $\Pi \in \mathcal{P}_i^{Q^*[W]} \setminus \{\Pi_i^W\}$. Write $\mathcal{R} = \{\Pi_i^W[v_c^W]\}_{c\in\mathcal{C}}$, so that $\mathcal{R}$ is a partition of R. By Lemma 5, $\mathcal{R} \subseteq \Pi$. Note that $\mathcal{R} \subseteq \Pi_i^W$ too. Since $\Pi \neq \Pi_i^W$, we in fact have $\mathcal{R} \subset \Pi$ and $\mathcal{R} \subset \Pi_i^W$. Hence $\Pi \setminus \mathcal{R}$ and $\Pi_i^W \setminus \mathcal{R}$ are distinct partitions of $\mathcal{V} \setminus R$. Since a one-element set has a unique partition, $\mathcal{V} \setminus R$ must contain at least two elements. ☐

5.3 Learning the Actual World Exactly

Putting Theorems 3 and 4 together we obtain a precise characterisation of when W can be found *exactly* by truth-tracking methods, i.e when $Q^*[W] = \{W\}$.

Corollary 1. $Q^*[W] = \{W\}$ *if and only if*

1. *There is a collection* $\{\Gamma_c\}_{c\in\mathcal{C}} \subseteq \mathcal{L}_0^{\mathcal{C}}$ *such that for each* c, *(i)* $W, c \models \Gamma_c$; *(ii)* $W \models \mathsf{E}_{\mathcal{S}}\Gamma_c$; *(iii)* $\mathrm{Cn}_0(\Gamma_c)$ *is maximally consistent; and*
2. *For each each* $i \in \mathcal{S}$, $|\mathcal{V} \setminus \bigcup_{c\in\mathcal{C}} \Pi_i^W[v_c^W]| \leq 1$.

6. Truth-Tracking Methods

So far we have focussed on solvable questions, and the extent to which they reveal information about the actual world. We now turn to the methods which solve them. We give a general characterisation of truth-tracking methods under mild assumptions, before discussing the families of *conditioning* and *score-based* methods from Singleton and Booth (2022).

6.1 A General Characterisation

For sequences σ, δ, write $\sigma \equiv \delta$ iff δ is obtained from σ by replacing each report $\langle i, c, \varphi\rangle$ with $\langle i, c, \psi\rangle$, for some $\psi \equiv \varphi$. For $k \in \mathbb{N}$, let σ^k denote the k-fold repetition of σ. Consider the following properties which may hold of a learning method L.

Equivalence *If* $\sigma \equiv \delta$ *then* $L(\sigma) = L(\delta)$.

Repetition $L(\sigma^k) = L(\sigma)$.

Soundness $L(\sigma) \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$.

Equivalence says that L should not care about the syntactic form of the input. *Repetition* says that the output from L should not change if each source repeats their reports k times. Note this is a weaker requirement than saying that L should be invariant under duplication of *some* reports in σ. In that case the number of times a source i repeats a particular report could, conceivably, be taken as an indicator of the *confidence* that i has in that report, and so the duplication represents extra context that we may want L to take into account. *Soundness* says that all reports in σ are conjectured to be sound.[5]

For methods satisfying these properties, we have a precise characterisation of truth-tracking, i.e., necessary and sufficient conditions for L to solve Q^*. First, some new notation is required. Write $\delta \preceq \sigma$ iff for each $\langle i, c, \varphi \rangle \in \delta$ there is $\psi \equiv \varphi$ such that $\langle i, c, \psi \rangle \in \sigma$. That is, σ contains everything δ does, up to logical equivalence. Set

$$T_\sigma = \mathcal{X}_\sigma^{\mathsf{snd}} \setminus \bigcup \left\{ \mathcal{X}_\delta^{\mathsf{snd}} \mid \delta \not\preceq \sigma \right\}$$

Then $W \in T_\sigma$ iff σ is sound for W and any δ sound for W has $\delta \preceq \sigma$. In this sense σ contains *all* soundness statements for W – up to equivalence – so can be seen as a finite version of a stream. Let us call σ a *pseudo-stream* for W whenever $W \in T_\sigma$.

Theorem 5. *A method L satisfying* Equivalence*,* Repetition *and* Soundness *is truth-tracking if and only if it satisfies the following property.*

Credulity *If $T_\sigma, c \not\models \mathsf{S}_i\varphi$ then $L(\sigma), c \models \neg\mathsf{S}_i\varphi$.*

Before the proof, we comment on our interpretation of *Credulity*. It says that whenever $\neg\mathsf{S}_i\varphi$ is consistent with T_σ – those W for which σ is a pseudo-stream – $L(\sigma)$ should imply $\neg\mathsf{S}_i\varphi$. Since the number of sound statements *decreases* with increasing expertise, this is a principle of *maximal trust*: we should believe i has the expertise to rule out φ in case c, whenever this is consistent with T_σ. That is, some amount of *credulity* is required to find the truth. Our assumption that learning methods receive complete streams ensures that, if a source in fact lacks this expertise, they will eventually report φ and this belief can be be retracted. A stronger version of *Credulity* spells this out explicitly in terms of expertise:

$$\text{If } T_\sigma, c \not\models \neg\mathsf{E}_i\varphi \text{ then } L(\sigma), c \models \mathsf{E}_i\varphi. \tag{1}$$

The above property (1) implies *Credulity* in the presence of *Soundness*, and is thus a sufficient condition for truth-tracking (when also taken with *Equivalence* and *Repetition*).[6]

Theorem 5 also shows truth-tracking cannot be performed *deductively*: the method $L(\sigma) = \mathcal{X}_\sigma^{\mathsf{snd}}$, which does not go beyond the mere information that each report is sound, fails *Credulity*. Some amount of *inductive* or *non-monotonic* reasoning, as captured by *Credulity*, is necessary.

Example 9. *For a simple example of* Credulity *assume* $\mathsf{Prop} = \{p\}$*,* $\mathcal{S} = \{i\}$ *and* $\mathcal{C} = \{c\}$*. In this case $\mathcal{V}$ contains just two possible valuations v_p and $v_{\neg p}$ according to whether p is true or false, respectively. There are also just two possible partitions for i, namely $\Pi_{\mathsf{E}} = \{\{v_p\}, \{v_{\neg p}\}\}$ and $\Pi_{\neg\mathsf{E}} = \{\mathcal{V}\}$ according to whether i is able to distinguish between the two possible valuations or not. Thus $\mathcal{W}$ contains four possible worlds $W_1 = \langle v_p, \Pi_{\mathsf{E}} \rangle$, $W_2 = \langle v_p, \Pi_{\neg\mathsf{E}} \rangle$, $W_3 = \langle v_{\neg p}, \Pi_{\mathsf{E}} \rangle$ and $W_4 = \langle v_{\neg p}, \Pi_{\neg\mathsf{E}} \rangle$. There are just three distinct reports,*

---

[5] Note that we are using the term "soundness" in three different, but related, ways in this paper. This usage refers to a property of learning methods, in contrast to our earlier usages as *(i)* a property of a formula being "sound" for a source to report, and *(ii)* a stream ρ being "sound" for a world W.

[6] We conjecture (1) is strictly stronger than *Credulity*.

up to logical equivalence, that i can possibly give about c, namely p, $\neg p$ and the tautological $p \vee \neg p$. Let's assume i reports the first and the third of these, i.e., $\sigma = (\langle i, c, p\rangle, \langle i, c, p \vee \neg p\rangle)$. *We have* $\mathcal{X}_\sigma^{\mathsf{snd}} = \{W_1, W_2, W_4\}$. *The missing report* $\langle i, c, \neg p\rangle$ *is sound for* W_2 *and* W_4 *but not for* W_1. *Thus* $T_\sigma = \{W_1\}$ *and so* $T_\sigma, c \not\models \mathsf{S}_i \neg p$. *Therefore* Credulity *says we should have* $L(\sigma), c \models \neg \mathsf{S}_i \neg p$, *which means that, on the basis of just* σ, *we should assume* i *is expert on* p.

The rest of this section works towards the proof of Theorem 5. We collect some useful properties of pseudo-streams. First, pseudo-streams provide a way of accessing Q^* via a finite sequence: T_σ is a cell in Q^* whenever it is non-empty.

Lemma 8. *If* $W \in T_\sigma$, *then (i)* $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ *iff* $W \sqsubseteq W'$; *and (ii)* $T_\sigma = Q^*[W]$.

Proof. Suppose $W \in T_\sigma$. For (i), first suppose $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ and $W, c \models \mathsf{S}_i\varphi$. Considering the singleton sequence $\delta = \langle i, c, \varphi\rangle$ we have $W \in \mathcal{X}_\delta^{\mathsf{snd}}$. From $W \in T_\sigma$ we get $\delta \preceq \sigma$, i.e., there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. From $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ and $\mathsf{S}_i\varphi \equiv \mathsf{S}_i\psi$ we get $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$.

Now suppose $W \sqsubseteq W'$ and let $\langle i, c, \varphi\rangle \in \sigma$. Then since $W \in T_\sigma \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ we have $W, c \models \mathsf{S}_i\varphi$, and $W \sqsubseteq W'$ gives $W', c \models \mathsf{S}_i\varphi$. Consequently $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$.

Now for (ii), first suppose $W' \in Q^*[W]$. Then W and W' satisfy exactly the same soundness statements, so $W' \in T_\sigma$ also. Conversely, suppose $W' \in T_\sigma$. Then $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$, so (i) gives $W \sqsubseteq W'$. But we also have $W' \in T_\sigma$ and $W \in \mathcal{X}_\sigma^{\mathsf{snd}}$, so (i) again gives $W' \sqsubseteq W$. Hence $W \approx W'$, i.e., $W' \in Q^*[W]$. ☐

We can now show that property (1) implies *Credulity* together with *Soundness*.

Proposition 5. *Suppose* L *satisfies* Soundness *and property (1). Then* L *satisfies* Credulity.

Proof. Suppose $T_\sigma, c \not\models \mathsf{S}_i\varphi$. By assumption, there is some $W \in T_\sigma$ such that $W, c \not\models \mathsf{S}_i\varphi$. Take some $W' \in L(\sigma)$. We need to show that $W', c \not\models \mathsf{S}_i\varphi$.

Now, from $W, c \not\models \mathsf{S}_i\varphi$ we get $\Pi_i^W[v_c^W] \cap \|\varphi\| = \emptyset$. Taking ψ such that $\|\psi\| = \Pi_i^W[v_c^W]$, we have $\|\psi\| \cap \|\varphi\| = \emptyset$ and $W, c \models \mathsf{E}_i\psi$. Thus $T_\sigma, c \not\models \neg\mathsf{E}_i\psi$, so property (1) gives $L(\sigma), c \models \mathsf{E}_i\psi$. Since $W' \in L(\sigma)$, we get $W', c \models \mathsf{E}_i\psi$, i.e., $\Pi_i^{W'}[\psi] = \|\psi\|$.

On the other hand, from *Soundness* we have $L(\sigma) \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$, so $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$. Since $W \in T_\sigma$, Lemma 8 gives $W \sqsubseteq W'$, and so $\|\psi\| = \Pi_i^W[v_c^W] \subseteq \Pi_i^{W'}[v_c^{W'}]$ by Lemma 1. Now since $\|\psi\|$ is a subset of the cell $\Pi_i^{W'}[v_c^{W'}]$, its expansion under $\Pi_i^{W'}$ is equal to this cell, i.e., $\Pi_i^{W'}[\psi] = \Pi_i^{W'}[v_c^{W'}]$. But we showed above that $\Pi_i^{W'}[\psi] = \|\psi\|$. Hence $\Pi_i^{W'}[v_c^{W'}] = \|\psi\|$. In particular, $\Pi_i^{W'}[v_c^{W'}] \cap \|\varphi\| = \emptyset$, and $W', c \not\models \mathsf{S}_i\varphi$ as required. ☐

The next two results show that initial segments of streams are (eventually) pseudo-streams, and that any pseudo-stream gives rise to a stream.

Lemma 9. *If* ρ *is a stream for* W, *there is* n *such that* $W \in T_{\rho[m]}$ *for all* $m \geq n$.

Proof. Let $\widehat{\cdot}$ be a function which selects a representative formula for each equivalence class of $\mathcal{L}_0/{\equiv}$, so that $\varphi \equiv \widehat{\varphi}$ and $\varphi \equiv \psi$ implies $\widehat{\varphi}$ is equal to $\widehat{\psi}$. Note that since Prop is finite, and since $\mathcal{S}$ and $\mathcal{C}$ are also finite, there are only finitely many reports of the form $\langle i, c, \widehat{\varphi}\rangle$. By completeness of ρ for W, we may take n sufficiently large so that $W, c \models \mathsf{S}_i\widehat{\varphi}$ implies

$\langle i, c, \widehat{\varphi}\rangle \in \rho[n]$, for all i, c, φ. Now, take $m \geq n$. We need to show $W \in T_{\rho[m]}$. Clearly $W \in \mathcal{X}^{\mathsf{snd}}_{\rho[m]}$, since ρ is sound for W. Suppose $W \in \mathcal{X}^{\mathsf{snd}}_{\delta}$. We need to show $\delta \preceq \rho[m]$. Indeed, take $\langle i, c, \varphi\rangle \in \delta$. Then $W, c \models \mathsf{S}_i\varphi$. Since $\mathsf{S}_i\varphi \equiv \mathsf{S}_i\widehat{\varphi}$, we have $W, c \models \mathsf{S}_i\widehat{\varphi}$. Hence $\langle i, c, \widehat{\varphi}\rangle$ appears in $\rho[n]$, and consequently in $\rho[m]$ too. Since $\varphi \equiv \widehat{\varphi}$, this shows $\delta \preceq \rho[m]$. □

Lemma 10. *If $W \in T_\sigma$ and $N = |\sigma|$, there is a stream ρ for W such that $\rho[Nk] \equiv \sigma^k$ for all $k \in \mathbb{N}$.*

Proof. First note that $W \in T_\sigma$ implies $\sigma \neq \emptyset$, so $N > 0$. Since $\mathcal{L}_0$ is countable, we may index the set of $\mathcal{L}_0$ formulas equivalent to $\varphi \in \mathcal{L}_0$ as $\{\varphi_n\}_{n\in\mathbb{N}}$. Let σ_n be obtained from σ by replacing each report $\langle i, c, \varphi\rangle$ with $\langle i, c, \varphi_n\rangle$. Then $\sigma \equiv \sigma_n$. Let ρ be the sequence obtained as the infinite concatenation $\sigma_1 \circ \sigma_2 \circ \sigma_3 \circ \cdots$ (this is possible since σ is of positive finite length). Then $\rho[Nk] = \sigma_1 \circ \cdots \circ \sigma_k$, and consequently $\rho[Nk] \equiv \sigma^k$.

It remains to show ρ is a stream for W. Soundness of ρ follows from $W \in T_\sigma \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$, since every report in ρ is equivalent to some report in σ by construction. For completeness, suppose $W, c \models \mathsf{S}_i\varphi$. As in the proof of Lemma 8, considering the singleton sequence $\delta = \langle i, c, \varphi\rangle$, we get from $W \in T_\sigma$ that there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. Hence there is $n \in \mathbb{N}$ such that $\varphi = \psi_n$, so $\langle i, c, \varphi\rangle \in \sigma_n$, and thus $\langle i, c, \varphi\rangle \in \rho$. □

Next we obtain an equivalent formulation of *Credulity* which is less transparent as a postulate for learning methods, but easier to work with.

Lemma 11. *Suppose L satisfies* Soundness*. Then L satisfies* Credulity *if and only if $L(\sigma) \subseteq T_\sigma$ for all σ with $T_\sigma \neq \emptyset$.*

Proof. “if”: Suppose $T_\sigma, c \not\models \mathsf{S}_i\varphi$. Then there is $W \in T_\sigma$ such that $W, c \not\models \mathsf{S}_i\varphi$. By our assumption and Lemma 8, $L(\sigma) \subseteq T_\sigma = Q^*[W]$. Thus every world in $L(\sigma)$ agrees with W on soundness statements, so $L(\sigma), c \models \neg\mathsf{S}_i\varphi$.

“only if”: Suppose there is some $W \in T_\sigma$, and take $W' \in L(\sigma)$. We need to show $W' \in T_\sigma$; by Lemma 8, this is equivalent to $W \approx W'$. First suppose $W, c \models \mathsf{S}_i\varphi$. Then $W \in T_\sigma$ implies there is $\psi \equiv \varphi$ such that $\langle i, c, \psi\rangle \in \sigma$. By *Soundness* for L, we have $W' \in L(\sigma) \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$. Consequently $W', c \models \mathsf{S}_i\psi$ and thus $W', c \models \mathsf{S}_i\varphi$. This shows $W \sqsubseteq W'$. Now suppose $W, c \not\models \mathsf{S}_i\varphi$. Then $T_\sigma, c \not\models \mathsf{S}_i\varphi$. By *Credulity*, $L(\sigma), c \models \neg\mathsf{S}_i\varphi$. Hence $W', c \not\models \mathsf{S}_i\varphi$. This shows $W' \sqsubseteq W$. Thus $W \approx W'$ as required. □

Finally, we prove the characterisation of truth-tracking.

Proof of Theorem 5. Suppose L satisfies *Equivalence*, *Repetition* and *Soundness*.

“if”: Suppose *Credulity* holds. We show L solves Q^*. Take any world W and stream ρ for W. By Lemma 9, there is n such that $W \in T_{\rho[m]}$ for all $m \geq n$. By Lemma 8, $T_{\rho[m]} = Q^*[W]$ for such m. In particular, $T_{\rho[m]} \neq \emptyset$. By *Credulity* and Lemma 11, we get $L(\rho[m]) \subseteq T_{\rho[m]} = Q^*[W]$.

“only if”: Suppose L solves Q^*. We show *Credulity* via Lemma 11. Suppose there is some $W \in T_\sigma$, and write $N = |\sigma| > 0$. By Lemma 10, there is a stream ρ for W such that $\rho[Nk] \equiv \sigma^k$ for all $k \in \mathbb{N}$. By *Repetition* and *Equivalence*, $L(\sigma) = L(\sigma^k) = L(\rho[Nk])$. But L solves Q^*, so for k sufficiently large we have $L(\rho[Nk]) \subseteq Q^*[W] = T_\sigma$. Hence, going via some large k, we obtain $L(\sigma) \subseteq T_\sigma$ as required. □

6.2 Conditioning Methods

In this section we turn to the family of *conditioning* methods, proposed by Singleton and Booth (2022) and inspired by similar methods in the belief change literature (Spohn, 1988), which have also been studied in the framework of learning of Gierasimczuk (2010) and Baltag et al. (2019). While our interpretation of input sequences is different – we read $\langle i, c, \varphi \rangle$ as i reporting φ is *possible* in case c, whereas Singleton and Booth (2022) read this as i *believes* φ – this class of methods can still be applied in our setting.

Conditioning methods operate by successively restricting a fixed *plausibility total preorder*[7] to the information corresponding to each new report $\langle i, c, \varphi \rangle$. In this paper, we take a report $\langle i, c, \varphi \rangle$ to correspond to the fact that $\mathsf{S}_i\varphi$ holds in case c; this fits with our assumption throughout that sources only report sound statements.[8] Thus, the worlds under consideration given a sequence σ are exactly those satisfying all soundness statements in σ, i.e., $\mathcal{X}^{\mathsf{snd}}_\sigma$. Note that $\mathcal{X}^{\mathsf{snd}}_\sigma$ represents the *indefeasible knowledge* given by σ: worlds outside $\mathcal{X}^{\mathsf{snd}}_\sigma$ are eliminated and cannot be recovered with further reports, since for any sequence δ, $\mathcal{X}^{\mathsf{snd}}_{\sigma\circ\delta} \subseteq \mathcal{X}^{\mathsf{snd}}_\sigma$. The plausibility order allows us to represent *defeasible beliefs* about the most plausible worlds within $\mathcal{X}^{\mathsf{snd}}_\sigma$.

Definition 4. *For a total preorder $\leq$ on $\mathcal{W}$, the* conditioning method $L_\leq$ *is given by* $L_\leq(\sigma) = \min_\leq \mathcal{X}^{\mathsf{snd}}_\sigma$.

Note that $\mathcal{X}^{\mathsf{snd}}_\sigma \neq \emptyset$ for all σ. For example, if $\Pi^W_i = \{\mathcal{V}\}$ for all i then $W \in \mathcal{X}^{\mathsf{snd}}_\sigma$. From this and the fact that $\mathcal{W}$ is finite, we know $L_\leq$ is consistent. Moreover, $L_\leq$ satisfies *Equivalence*, *Repetition* and *Soundness*.

Example 10. *We recall two concrete choices of $\leq$ given by Singleton and Booth (2022).*

1. *Set $W \leq_{\mathsf{vbc}} W'$ iff $r_{\mathsf{vbc}}(W) \leq r_{\mathsf{vbc}}(W')$, where*

$$r_{\mathsf{vbc}}(W) = -\sum_{i \in \mathcal{S}} |\{p \in \mathsf{Prop} \mid \Pi^W_i[p] = \|p\|\}|.$$

   The most plausible worlds in this order are those in which sources have as much expertise on the propositional variables as possible, on aggregate. The subscript vbc *here stands for* variable-based conditioning*, and we denote the corresponding conditioning method by L_{vbc}.*

2. *Set $W \leq_{\mathsf{pbc}} W'$ iff $r_{\mathsf{pbc}}(W) \leq r_{\mathsf{pbc}}(W')$, where*

$$r_{\mathsf{pbc}}(W) = -\sum_{i \in \mathcal{S}} |\Pi^W_i|.$$

   This order aims to maximise the number of cells in each source's partitions, thereby maximising the number of propositions on which they have expertise. Note that the propositional variables play no special role. The subscript pbc *here stands for* partition-based conditioning*, and we denote the corresponding conditioning operator by L_{pbc}.*

---

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

[8] Singleton and Booth (2022) consider more general conditioning methods in which this choice is not fixed.

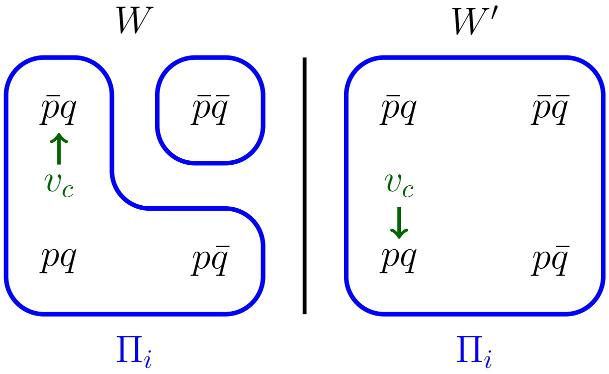

Figure 3: Worlds which demonstrate L_{vbc} is not truth-tracking.

A straightforward property of $\leq$ characterises truth-tracking for conditioning methods. For a generic total preorder $\leq$, let $<$ denote its strict part.

Theorem 6. *$L_{\leq}$ is truth-tracking if and only if*

$$W \sqsubset W' \implies \exists W'' \approx W \text{ such that } W'' < W'. \tag{2}$$

Like *Credulity*, (2) is a principle of maximising trust in sources. Recall from Lemma 1 that $W \sqsubset W'$ means all sources are more knowledgeable in each case in W than in W', and there is at least one source and case for which this holds strictly. If we aim to trust sources as much as possible, we might impose $W < W'$ here; then W' is strictly less plausible and will be ruled out in favour of W. This yields a sufficient condition for truth-tracking, but to obtain a necessary condition we need to allow a "surrogate" world $W'' \approx W$ to take the place of W.

Proof of Theorem 6. Write $L = L_{\leq}$. Since L satisfies *Equivalence*, *Repetition* and *Soundness*, we may use Theorem 5. Furthermore, it is sufficient by Lemma 11 to show that (2) holds if and only if $L(\sigma) \subseteq T_\sigma$, whenever $T_\sigma \neq \emptyset$.

"if": Suppose $W \sqsubset W'$. Let σ be some pseudo-stream for W, so that $W \in T_\sigma$.[9] Note that since $W \in T_\sigma \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$ and $W \sqsubset W'$, we have $W' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ also. By assumption, $L(\sigma) \subseteq T_\sigma = Q^*[W]$. Since $W \not\approx W'$, this means $W' \in \mathcal{X}_\sigma^{\mathsf{snd}} \setminus L(\sigma)$. That is, W' lies in $\mathcal{X}_\sigma^{\mathsf{snd}}$ but is not $\leq$-minimal. Consequently there is $W'' \in \mathcal{X}_\sigma^{\mathsf{snd}}$ such that $W'' < W'$. Since L is consistent, we may assume without loss of generality that $W'' \in L(\sigma)$. Hence $W'' \in Q^*[W]$, so $W'' \approx W$.

"only if": Suppose there is some $W \in T_\sigma$, and let $W' \in L(\sigma)$. We need to show $W' \in T_\sigma = Q^*[W]$, i.e., $W \approx W'$. Since $W' \in L(\sigma) \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$, Lemma 8 gives $W \sqsubseteq W'$. Suppose for contradiction that $W \not\approx W'$. Then $W \sqsubset W'$. By (2), there is $W'' \approx W$ such that $W'' < W'$. But W' is $\leq$-minimal in $\mathcal{X}_\sigma^{\mathsf{snd}}$, so this must mean $W'' \notin \mathcal{X}_\sigma^{\mathsf{snd}}$. On the other hand, $W'' \in Q^*[W] = T_\sigma \subseteq \mathcal{X}_\sigma^{\mathsf{snd}}$: contradiction. □

Example 11. *We revisit the methods of Example 10.*

---

[9] For example, pick some stream ρ and apply Lemma 9 to obtain a pseudo-stream.

1. *The variable-based conditioning method L_{vbc} is* not *truth-tracking. Indeed, consider the worlds W and W' shown in Fig. 3, where we assume $\mathsf{Prop} = \{p, q\}$, $\mathcal{S} = \{i\}$ and $\mathcal{C} = \{c\}$. Then $W \sqsubset W'$ (e.g., by Lemma 1). Note that i does not have expertise on p or q in both W and W', so $r_{\mathsf{vbc}}(W) = r_{\mathsf{vbc}}(W') = 0$. Moreover, i's partition is uniquely determined in $Q^*[W]$ by Theorem 4, so if $W'' \approx W$ then $r_{\mathsf{vbc}}(W'') = 0$ also. That is, there is no $W'' \approx W$ such that $W'' <_{\mathsf{vbc}} W'$. Hence (2) fails, and L_{vbc} is not truth-tracking. Intuitively, the problem here is that since i's expertise is not split along the lines of the propositional variables when W is the actual world, L_{vbc} will always maintain W' as a possibility.*

2. *The partition-based conditioning method L_{pbc}* is *truth-tracking. Indeed, if $W \sqsubset W'$ we may construct W'' from W by modifying the partition of each source i so that all valuations outside of $\bigcup_{c \in \mathcal{C}} \Pi_i^W[v_c^W]$ lie in their own cell. Then $W \approx W''$. One can show that $\Pi_i^{W''}$ refines $\Pi_i^{W'}$ for all $i \in \mathcal{S}$, and there is some i for which the refinement is strict. Hence the partitions in W'' contain strictly more cells, so $W'' <_{\mathsf{pbc}} W'$.*

6.3 Score-based Methods

In this section we consider the other class of methods introduced by Singleton and Booth (2022): *score-based* methods. These methods make use of a function d that represents a measure of “disagreement” between a world W and a report $\langle i, c, \varphi \rangle$; the greater $d(W, \langle i, c, \varphi \rangle)$, the less plausible it is deemed for i to report φ in case c.

Definition 5. *For a function $d : \mathcal{W} \times (\mathcal{S} \times \mathcal{C} \times \mathcal{L}_0) \to \mathbb{N}_0$ and a sequence σ, write*

$$r_d^\sigma(W) = \sum_{\langle i, c, \varphi \rangle \in \sigma} d(W, \langle i, c, \varphi \rangle).$$

The score-based method *L_d is then given by $L_d(\sigma) = \mathrm{argmin}_{W \in \mathcal{X}_\sigma^{\mathsf{snd}}} r_d^\sigma(W)$.*

The conjecture $L_d(\sigma)$ consists of the worlds W satisfying the soundness constraints of σ with minimal disagreement score, computed as the sum $r_d^\sigma(W)$ of the disagreement on each report.[10] L_d is consistent and satisfies both *Repetition* and *Soundness* for any choice of d.

Example 12. *Adapting the score-based example from (Singleton & Booth, 2022) to this setting, take*

$$d_{\mathsf{exm}}(W, \langle i, c, \varphi \rangle) = |\Pi_i^W[\varphi] \setminus \|\varphi\||.$$

The corresponding method aims to minimise the “excess” valuations in $\Pi_i^W[\varphi]$ which are not themselves models of φ. We denote it by L_{exm}, standing for excess-minimisation.

Truth-tracking for score-based operators satisfying *Equivalence* can be characterised in almost exactly the same way as for conditioning operators, using a property similar to (2).

Theorem 7. *Suppose d is such that $d(W, \langle i, c, \varphi \rangle) = d(W, \langle i, c, \psi \rangle)$ whenever $\varphi \equiv \psi$. Then L_d is truth-tracking if and only if*

$$W \in T_\sigma \text{ and } W \sqsubset W' \implies \exists W'' \approx W \text{ such that } r_d^\sigma(W'') < r_d^\sigma(W'). \tag{3}$$

---

[10] Singleton and Booth (2022) actually consider a more general class of score-based methods in which $r_d^\sigma(W)$ also includes a summand $r_0(W)$ coming from some given prior plausibility ranking r_0.

The proof is essentially identical to that of Theorem 6, and is thus omitted.

Example 13. *Revisiting Example 12, we find that* L_{exm} *is truth-tracking. Indeed, it is clear that* d_{exm} *treats equivalent formulas identically, since* $d_{\mathsf{exm}}(W, \langle i, c, \varphi\rangle)$ *only depends on* Π_i^W *and* $\|\varphi\|$*. Given* $W \in T_\sigma$ *and* $W \sqsubset W'$ *one can take* W'' *in the same way as for* L_{pbc} *in Example 11. Then* $\Pi_i^{W''}$ *refines* $\Pi_i^{W'}$ *for all* i*, so* $d_{\mathsf{exm}}(W'', \langle i, c, \varphi\rangle) \leq d_{\mathsf{exm}}(W', \langle i, c, \varphi\rangle)$ *for all* $\langle i, c, \varphi\rangle \in \sigma$*. Consequently,* $r_d^\sigma(W'') \leq r_d^\sigma(W')$*. Moreover, since* $W \sqsubset W'$ *there is some* i *and* c *such that* $\Pi_i^W[v_c^W] \subset \Pi_i^{W'}[v_c^{W'}]$*. Taking any* φ *such that* $\|\varphi\| = \Pi_i^W[v_c^W]$*, we have* $W, c \models \mathsf{S}_i\varphi$*. Since* $W \in T_\sigma$*, there is some* $\psi \equiv \varphi$ *such that* $\langle i, c, \psi\rangle \in \sigma$*. Now, since* $W \approx W''$ *we have* $\Pi_i^{W''}[v_c^{W''}] = \Pi_i^W[v_c^W]$*. Consequently*

$$\Pi_i^{W''}[\psi] = \Pi_i^{W''}[\Pi_i^W[v_c^W]] = \Pi_i^{W''}[\Pi_i^{W''}[v_c^{W''}]] = \Pi_i^{W''}[v_c^{W''}] = \Pi_i^W[v_c^W] = \|\psi\|$$

and thus $d_{\mathsf{exm}}(W'', \langle i, c, \psi\rangle) = 0$*. On the other hand,* $\Pi_i^{W'}[v_c^{W'}] \supset \|\psi\|$ *implies* $\Pi_i^{W'}[\psi] = \Pi_i^{W'}[v_c^{W'}]$*, and so*

$$d_{\mathsf{exm}}(W', \langle i, c, \psi\rangle) = |\Pi_i^{W'}[v_c^{W'}] \setminus \|\psi\|| > 0 = d_{\mathsf{exm}}(W'', \langle i, c, \psi\rangle),$$

which gives $r_d^\sigma(W'') < r_d^\sigma(W')$ *as required.*

7. Conclusion

Summary. In this paper we studied truth-tracking in the presence of non-expert sources. To start with, the model assumes sources report everything true *up to their lack of expertise*, i.e., all that they consider possible. We obtained precise characterisations of when truth-tracking methods can uniquely find the actual valuations of a world W, and in doing so showed how sources may *combine* their expertise to track the truth. Similar results were presented for finding the actual partitions of a world W, i.e., finding the true extent of each source's expertise.

We then presented the *Credulity* postulate, which characterises truth-tracking methods under mild assumptions. Roughly speaking, this postulate says that one needs to *trust* sources to be experts wherever possible. Purely deductive reasoning – in which one does not conjecture beyond the fact that each received report is sound – fails to be credulous enough in this sense, and thus some amount of *non-monotonic* reasoning is required for truth-tracking.

Next, we reconsidered the belief and expertise operators of Singleton and Booth (2022) in the context of truth-tracking. Interestingly, it was seen that the variable-based conditioning method L_{vbc} is *not* truth-tracking, but the partition-based conditioning method L_{pbc} and score-based method L_{exm} are. The success of the latter two methods showed that truth-tracking is compatible with rational belief change as expressed by the postulates of (Singleton & Booth, 2022) (which are satisfied by all three methods).

Limitations and future work. Conceptually, the assumption that streams are complete is very strong. As seen in Example 3, completeness requires sources to give jointly inconsistent reports whenever $\Pi_i[v_c]$ contains more than just v_c. Such reports provide information about the source's expertise: if i reports both φ and $\neg\varphi$ we know $\neg\mathsf{E}_i\varphi$. To provide all sound reports sources must also have *negative introspection* over their own knowledge, i.e., they *know* when they do not know something. Indeed, our use of partitions makes expertise closely related to S5 knowledge (Singleton & Booth, 2022; Singleton, 2021), which has been criticised in the philosophical literature as too strong. In reality, non-expert sources may have *beliefs* about the world, and may prefer to report only that which they believe. A source may even believe a sound report φ is *false*, since soundness only says the source does not *know* $\neg\varphi$. For example, in Example 1 the doctor D may think it is more likely that A suffers from p than q, but we cannot express this in our framework.

On the technical side, our results on solvability of Q^* and the characterisation of Theorem 5 rely on the fact that we only consider finitely many worlds. In a sense this trivialises the problem of induction as studied by Kelly et al. (1997) and Baltag et al. (2015), among others. In future work it would be interesting to see which results can be carried over to the case where Prop is infinite.

Acknowledgements

Thanks are due to the reviewers, both of this journal and of NMR 2022, for some highly useful comments that helped to improve the paper.

References

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 510–530.

Baltag, A., Gierasimczuk, N., Özgün, A., Sandoval, A. L. V., & Smets, S. (2019). A dynamic logic for learning theory. *Journal of Logical and Algebraic Methods in Programming*, *109*, 100485.

Baltag, A., Gierasimczuk, N., & Smets, S. (2015). On the solvability of inductive problems: A study in epistemic topology. In Ramanujam, R. (Ed.), *Proceedings Fifteenth Conference on Theoretical Aspects of Rationality and Knowledge, TARK, 2015*, Vol. 215 of *EPTCS*, pp. 81–98.

Baltag, A., Gierasimczuk, N., & Smets, S. (2019). Truth-Tracking by Belief Revision. *Studia Logica*, *107*(5), 917–947.

de Condorcet, N. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Paris: L'Imprimerie Royale.

Elkind, E., & Slinko, A. (2016). Rationalizations of voting rules. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.), *Handbook of Computational Social Choice*, pp. 169–196. Cambridge University Press.

Eschenbach, C., & Özcep, O. L. (2010). Ontology-revision operators based on reinterpretation. *Logic Journal of the IGPL*, *18*(4), 579–616.

Everaere, P., Konieczny, S., & Marquis, P. (2010). The epistemic view of belief merging: Can we track the truth?. In *Proceedings Nineteenth European Conference on Artificial Intelligence (ECAI)*, pp. 621–626.

Gierasimczuk, N. (2009a). Bridging learning theory and dynamic epistemic logic. *Synthese*, *169*(2), 371–384.

Gierasimczuk, N. (2009b). Learning by erasing in dynamic epistemic logic. In *International Conference on Language and Automata Theory and Applications*, pp. 362–373. Springer.

Gierasimczuk, N. (2010). *Knowing One's Limits: Logical Analysis of Inductive Inference.* Ph.D. thesis, ILLC.

Gierasimczuk, N. (2023). Inductive inference and epistemic modal logic. In Klin, B., & Pimentel, E. (Eds.), *31st EACSL Annual Conference on Computer Science Logic, CSL 2023, February 13-16, 2023, Warsaw, Poland*, Vol. 252 of *LIPIcs*, pp. 2:1–2:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447–474.

Goranko, V., & Passy, S. (1992). Using the universal modality: Gains and questions. *Journal of Logic and Computation*, *2*(1), 5–30.

Hartmann, S., & Sprenger, J. (2012). Judgment aggregation and the problem of tracking the truth. *Synthese*, *187*(1), 209–221.

Jain, S., Osherson, D., Royer, J. S., & Sharma, A. (1999). *Systems that Learn: An Introduction to Learning Theory.* MIT.

Kelly, K., Schulte, O., & Hendricks, V. (1997). Reliable belief revision. In Chiara, M. L., Doets, K., Mundici, D., & van Benthem, J. (Eds.), *Logic and Scientific Methods*, pp. 383–398. Springer.

Konieczny, S., & Pino Pérez, R. (2002). Merging information under constraints: A logical framework. *Journal of Logic and Computation*, *12*(5), 773–808.

Li, Y., Gao, J., Meng, C., Li, Q., Su, L., Zhao, B., Fan, W., & Han, J. (2016). A Survey on Truth Discovery. *SIGKDD Explor. Newsl.*, *17*(2), 1–16.

Singleton, J. (2021). A logic of expertise. *ESSLLI 2021 Student Session.* https://arxiv.org/abs/2107.10832.

Singleton, J., & Booth, R. (2022). Who's the Expert? On Multi-source Belief Change. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*, pp. 331–340.

Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In Harper, W. L., & Skyrms, B. (Eds.), *Causation in Decision, Belief Change, and Statistics*, pp. 105–134. Springer.

Terzopoulou, Z., & Endriss, U. (2019). Optimal truth-tracking rules for the aggregation of incomplete judgments. In *Proceedings of the 12th International Symposium on Algorithmic Game Theory (SAGT-2019).*