In Data We Trust: The Logic of Trust-Based Beliefs

Junli Jiang[1] and **Pavel Naumov**[2]
[1]Institute of Logic and Intelligence, Southwest University, China
[2]University of Southampton, the United Kingdom
walk08@swu.edu.cn, p.naumov@soton.ac.uk

Abstract

The paper proposes a data-centred approach to reasoning about the interplay between trust and beliefs. At its core, is the modality "under the assumption that one dataset is trustworthy, another dataset informs a belief in a statement". The main technical result is a sound and complete logical system capturing the properties of this modality.

1 Introduction

With the constant growth of data available to human and artificial agents, the knowledge and beliefs of these agents are becoming primarily informed not by their own experiences, but by the data they have access to and by whether they trust these data. In the case of algorithmic trading agents, an important source of the data is news analytics [Mitra *et al.*, 2011; von Beschwitz *et al.*, 2020].

As an example, consider a situation when a regulatory agency is about to announce its decision on an application to approve a new drug developed by a small start-up company. If the agency announces the approval of the drug the next day, the stock of the company will rise sharply. If not, the stock will become essentially worthless.

Suppose that the day before the announcement two widely-read newspapers, The Post and The Times, publish contradictory predictions. While The Post article p assures its readers that the drug will be approved, The Times article t insists that the application will be denied.

Note that anyone who *reads* article p and *trusts* it would conclude that the agency has decided to approve the drug. In other words, under the assumption that article p is trustworthy, it informs the belief that the agency has decided to approve the drug:

$$\mathsf{B}^p_p(\text{“the drug is approved”}). \tag{1}$$

We will call article p a *data variable*. In our case, the value of this data variable is the string representing the content of the article. In other cases, data variable can have other types such as numerical or Boolean.

In general, instead of a single data variable, we consider sets of data variables. We call such sets *datasets*. For any datasets X and T, we write $\mathsf{B}^T_X\varphi$ if under the assumption that dataset T is trustworthy, dataset X informs the belief in statement φ.

Going back to our example, note that without the assumption that The Post article p is trustworthy, it does not inform the belief that the drug is approved:

$$\neg\mathsf{B}_p(\text{“the drug is approved”}). \tag{2}$$

At the same time, anyone who trusts The Post, but has not necessarily read the article p, would believe that if article p says that the drug is approved, then it must be approved. We say that just the assumption of the trustworthiness of data variable p informs the belief that if p says the drug is approved, then it should be true:

$$\mathsf{B}^p(\text{“if } p \text{ says drug is approved, then it is approved”}). \tag{3}$$

Next, note that anyone who reads The Post's article (but not necessarily trusts it!) would conclude that its content, under the assumption of the trustworthiness, informs the belief that the drug is approved:

$$\mathsf{B}_p\mathsf{B}^p_p(\text{“the drug is approved”}).$$

Also, anyone who trusts the article (but not necessarily read it) would believe that if under the assumption of the trustworthiness, its content informs the belief that the drug is approved, then the drug indeed must be approved:

$$\mathsf{B}^p(\mathsf{B}^p_p(\text{“drug is approved”}) \to \text{“drug is approved”}). \tag{4}$$

Recall from statement (1) that anyone who reads and trusts article p would belief that the drug is approved. This, however, does not imply that the start-up company would make a good investment. Indeed, if all investors agree that the drug will be approved, then the company's stock price is already adjusted and it is no longer undervalued. Thus, even under the assumption of the trustworthiness of article p, the article does not inform the believe that the start-up company makes a good investment:

$$\neg\mathsf{B}^p_p(\text{“good time to invest in the start-up company”}).$$

Let us assume that it is well-known that a significant number of investors trust The Times rather than The Post. Anyone who reads The Times would conclude that a significant portion of the investors assumes the application is rejected:

$$\mathsf{B}_t\mathsf{B}^t_t(\text{“the drug is not approved”}).$$

Hence, anyone who reads both articles, but trusts p, would conclude that the stock of the start-up company is currently undervalued and it is a good time to invest:

$$\mathsf{B}^{p}_{p,t}(\text{“good time to invest in the start-up company”}). \quad (5)$$

Furthermore, anyone who reads both articles, but not necessarily trusts either of them, would conclude that anyone who reads both articles and trusts p, would believe that the stock is undervalued:

$$\mathsf{B}_{p,t}\mathsf{B}^{p}_{p,t}(\text{“good time to invest in the start-up company”}).$$

In this paper, we formally define trust-based belief modality $\mathsf{B}^T_X\varphi$ and give its sound and complete axiomatisation.

The rest of the paper is structured as follows. First, we define trustworthiness models that are used later to give a formal semantics of our logical system. Then, we introduce the syntax and the formal semantics of the system. In Section 4, we list and discuss its axioms and the inference rules. We review related literature on data-informed knowledge, beliefs, and trust in Section 5. Section 6 contains the proof of the completeness theorem. Section 7 discusses possible extensions of our system with Armstrong's functional dependency relation and a public announcement modality. Section 8 concludes.

2 Trustworthiness Model

The models that we introduce here are Kripke-style models with possible worlds. In our introductory example, the set of possible worlds could be thought of as a set of triples (a, p, t) where $a \in \{\text{approved}, \text{denied}\}$ is the agency decision while p and t is the content of The Post and The Times articles, respectively. A distinctive property of the trustworthiness models is that for each world w the model specifies the set $\mathcal{T}_w$ of variables which are "trustworthy" in the world w. For example, in the world

$$w_1 = (\text{approved}, \text{“It's a strong YES!”}, \text{“Agency says no!”})$$

variable p is trustworthy and t is not. Thus, $\mathcal{T}_{w_1} = \{p\}$.

As we will see later, for the semantics of the modality $\mathsf{B}^T_X\varphi$, it is not significant what the actual values of data variables are. It is only important if these values are different or the same in any two given worlds. Thus, for the sake of simplicity, in our formal definition of a trustworthiness model below, we do not associate a domain of values with data variables. Instead, we specify an "indistinguishability" equivalence relation $\sim_x$ on the worlds for each data variable x. Informally, $w \sim_x u$ if data variable x has the same values in worlds w and u. For example, if world w_1 is the one specified above and

$$w_2 = (\text{denied}, \text{“It's a strong YES!”}, \text{“Approved!”}),$$

then $w_1 \sim_p w_2$ and $w_1 \not\sim_t w_2$. Note that $\mathcal{T}_{w_2} = \varnothing$.

Throughout the rest of the paper, we assume a fixed set of data variables V. By a dataset we mean any subset of V.

Definition 1. *A tuple* $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$ *is called a trustworthiness model if*

1. *W is a (possibly empty) set of worlds,*
2. *$\sim_x$ is an "indistinguishability" equivalence relation on set W for each $x \in V$,*
3. *$\mathcal{T}_w \subseteq V$ is a set of data variables that are "trustworthy" in world $w \in W$,*
4. *$\pi(p)$ is a subset of W for each propositional variable p.*

3 Syntax and Semantics

In this paper, we assume an arbitrary fixed set V of "data variables" as well as a fixed set of propositional variables. The language Φ of our logical system is defined by the grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{B}^T_X\varphi,$$

where p is a propositional variable and $X, T \subseteq V$ are datasets. We read $\mathsf{B}^T_X\varphi$ as "under the assumption of trustworthiness of dataset T, dataset X informs the belief in φ".

We assume that $\perp$ is formula $\neg(p \to p)$, where p is one of propositional variables. In addition, for any dataset X and any worlds $w, u \in W$, we write $w \sim_X u$ if $w \sim_x u$ for each data variable $x \in X$.

Definition 2. *For any world $w \in W$ of any trustworthiness model $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$ and any formula $\varphi \in \Phi$, satisfaction relation $w \Vdash \varphi$ is defined as follows:*

1. *$w \Vdash p$ if $w \in \pi(p)$,*
2. *$w \Vdash \neg\varphi$ if $w \nVdash \varphi$,*
3. *$w \Vdash \varphi \to \psi$ if $w \nVdash \varphi$ or $w \Vdash \psi$,*
4. *$w \Vdash \mathsf{B}^T_X\varphi$ if $u \Vdash \varphi$ for each world $u \in W$ such that $w \sim_X u$ and $T \subseteq \mathcal{T}_u$.*

To understand item 4 of the above definition, let us go back to our running example and consider worlds w_1 and w_2 defined in Section 2. In world w_1, article p does not inform, without the assumption that this article is trustworthy, the belief that the drug is approved. This is because among the worlds indistinguishable by p from w_1 there is world w_2 where the drug is not approved. Thus, statement (2) is satisfied in world w_1. If the assumption of trustworthiness of data p is added, then the worlds like w_2 are excluded from the consideration (p is not trustworthy there). In the remaining worlds, the drug is approved. Hence, under the assumption that data p is trustworthy, in world w_1, this data informs the belief that the drug is approved. Thus, statement (1) is also satisfied in world w_1. Observe that statement $w \Vdash \mathsf{B}^T_X\perp$ is true if there is no X-indistinguishable from w world in which all variables in dataset T are trustworthy.

Note that the expression $w \Vdash \mathsf{B}^{\varnothing}_X\varphi$ says that statement φ is true in all worlds X-indistinguishable from world w. In other words, it says that statement φ is true as long as the values of variables in dataset X are the same as in world w. In such a situation, we say that the *knowledge* of φ is informed by dataset X in world w. It is easy to see that modality $\mathsf{B}^{\varnothing}_X$ satisfies all standard S5 properties from epistemic logic.

4 Axioms

In addition to propositional tautologies in language Φ, our *Logic of Trust-Based Beliefs* contains the axioms listed below.

A1. Truth: $\mathsf{B}^{\varnothing}_X\varphi \to \varphi$, A2. Distributivity: $\mathsf{B}^T_X(\varphi \to \psi) \to (\mathsf{B}^T_X\varphi \to \mathsf{B}^T_X\psi)$,

A3. Negative Introspection of Beliefs: $\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$,

A4. Monotonicity: $\mathsf{B}^T_X\varphi \to \mathsf{B}^{T'}_{X'}\varphi$, where $X \subseteq X'$, $T \subseteq T'$,

A5. Trust: $\mathsf{B}^T_X(\mathsf{B}^T_Y\varphi \to \varphi)$.

To understand the meaning of the Truth and the Negative Introspection of Beliefs axioms, recall from Section 3 that $\mathsf{B}^{\varnothing}_X\varphi$ is the *knowledge* modality "dataset X informs the knowledge of statement φ". Hence, the Truth axiom is the standard Truth axiom from the epistemic logic. The Negative Introspection of Beliefs axiom states that if dataset X does not inform the belief in φ under the assumption of trustworthiness of dataset T, then dataset X informs the knowledge of this. Note that the standard Negative Introspection axiom from the epistemic logic is a special case of our axiom when set T is empty. The positive introspection of beliefs also holds. We prove it from the above axioms in Lemma 1.

The Trust axiom is a general form of statement (4) in the introduction. Informally, it states that anyone trusting dataset T believes that any belief based on trust in T must be true.

We write $\vdash \varphi$ and say that formula φ is a *theorem* if φ is provable from the above axioms using the Modus Ponens and the Necessitation

$$\frac{\varphi, \varphi \to \psi}{\psi} \qquad\qquad \frac{\varphi}{\mathsf{B}^T_X\varphi}$$

inference rules. In addition to the unary relation $\vdash \varphi$, we also consider a binary relation $F \vdash \varphi$. We write $F \vdash \varphi$ if formula φ is derivable from the *theorems* of our logical system and the set of additional assumptions F using the Modus Ponens inference rule only. Note that statement $\varnothing \vdash \varphi$ is equivalent to $\vdash \varphi$. We say that a set of formulae F is *inconsistent* if $F \vdash \varphi$ and $F \vdash \neg\varphi$ for some formula $\varphi \in \Phi$.

Lemma 1. $\vdash \mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\mathsf{B}^T_X\varphi$.

Proof. Formula $\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi \to \neg\mathsf{B}^T_X\varphi$ is an instance of the Truth axiom. Thus, $\vdash \mathsf{B}^T_X\varphi \to \neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$, by contraposition. Hence, taking into account the following instance $\neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$ of the Negative Introspection axiom, we have

$$\vdash \mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi. \tag{6}$$

At the same time, formula $\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$ is also an instance of the Negative Introspection axiom. Thus, $\vdash \neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^T_X\varphi$ by the law of contrapositive in the propositional logic. Hence, by the Necessitation inference rule, $\vdash \mathsf{B}^{\varnothing}_X(\neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^T_X\varphi)$. Thus, by the Distributivity axiom and the Modus Ponens inference rule, $\vdash \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\mathsf{B}^T_X\varphi$. The latter, together with statement (6), implies the statement of the lemma by propositional reasoning. $\square$

The proofs of the next two results are omitted due to the space constraint.

Theorem 1 (strong soundness)**.** *For any world w of a trustworthiness model, any set of formulae $F \subseteq \Phi$, and any formula $\varphi \in \Phi$, if $w \Vdash f$ for each formula $f \in F$ and $F \vdash \varphi$, then $w \Vdash \varphi$.*

Lemma 2. *If $\varphi_1, .., \varphi_n \vdash \psi$, then $\mathsf{B}^T_X\varphi_1, .., \mathsf{B}^T_X\varphi_n \vdash \mathsf{B}^T_X\psi$.*

5 Literature Review

The modality $\mathsf{B}^T_X\varphi$ is closely connected to counterfactual modality $\psi \mathrel{\Box\!\!\to} \varphi$ [Lewis, 1973], also known as conditional belief modality. Informally, $\psi \mathrel{\Box\!\!\to} \varphi$ states that the assumption of ψ leads to a belief in φ. Using counterfactual modality, statement (1) from the introduction can be written as

"The Post article is true" $\mathrel{\Box\!\!\to}$ "the drug is approved".

The latter statement, essentially, says that under the assumption of trustworthiness of The Post article, the knowledge of the content of this article informs the belief that the drug will be approved. The major advantage of the modality $\mathsf{B}^T_X\varphi$, that we propose, is the ability to separate what is trustworthy from what is known. Without this separation, one would not be able to express beliefs that are based on pure trust of the source without the knowledge of the content, as in statements (3) and (4). One also would not be able to express beliefs based on a mix of trusted and untrusted sources as in statement (5).

Lewis [1973] used sphere semantics for modality $\mathrel{\Box\!\!\to}$. This semantics has been later generalised to neighbourhood semantics [Girlando *et al.*, 2016; Girlando *et al.*, 2019; van Eijck and Li, 2017]. Another type of semantics for modality $\mathrel{\Box\!\!\to}$ is plausibility semantics [Board, 2004; Baltag and Smets, 2006; Baltag and Smets, 2008; Boutilier, 1994; Friedman and Halpern, 1997; Friedman and Halpern, 1999]. Trustworthiness models are original to the current paper. It is interesting to point out that, properly modified, axioms of our logical system are valid for modality $\mathrel{\Box\!\!\to}$ under sphere and plausibility semantics. For example, our Trust axiom is sound in the form: $\psi \mathrel{\Box\!\!\to} ((\psi \mathrel{\Box\!\!\to} \varphi) \to \varphi)$.

In Section 3, we noticed that $w \Vdash \mathsf{B}^T_X\bot$ is true if there is no X-indistinguishable from w world in which all variables in dataset T are trustworthy. This corresponds to statement $\psi \mathrel{\Box\!\!\to} \bot$ being true if formula ψ is not satisfied in any worlds.

Recall from our discussion in Section 3 that statement $\mathsf{B}^{\varnothing}_X\varphi$ captures data-informed *knowledge*. This modality is essentially the same as "dependence" modality $D_X\varphi$ recently introduced by Baltag and van Benthem [2021]. However, due to slight difference in semantics, their modality $D_X\varphi$ in addition to S5 properties also satisfies axiom $\varphi \to D_X\varphi$. Under the semantics of Definition 2, property $\varphi \to \mathsf{B}^{\varnothing}_X\varphi$ is not universally true.

Many other logical systems for reasoning about values of data variables have been proposed before. Armstrong considered relation $X \rhd Y$ that means "the values of the variables in set X functionally determine the values of the variables in set Y" [1974]. His axioms became known in database literature as Armstrong's axioms [Garcia-Molina *et al.*, 2009, p. 81]. Baltag proposed a logical system for expression $X \rhd_a Y$, that stands for "agent a knows how to compute dataset Y based on dataset X" [2016]. More and Naumov gave axiomatisation of "no-information-flow" relation [2009], proposed in [Sutherland, 1986]. Modality "public inspection" of a dataset was introduced in [van Eijck *et al.*, 2017]. Wang and Fan gave axiomatisation of "conditionally knowing value" modality [2014].

Multiple logical systems capturing properties of trust have been proposed. Castelfranchi and Falcone suggested to treat trust as a mental state and define it through beliefs. Very roughly, I trust you to do something if I belief that you will do it [1998]. This approach has been further developed in [Herzig *et al.*, 2010]. Tagliaferri and Aldini introduced trust as a modality whose semantics is defined through numerical trustworthiness threshold functions [2019]. They did not consider a connection between trust and beliefs. Primiero proposed a trust logic for reasoning about communications [2020].

The closest works to ours are [Liau, 2003] and [Perrotin *et al.*, 2019]. In [Liau, 2003], the author introduced a logical system containing modalities $B_a\varphi$ (agent a beliefs in φ), $I_{a,b}\varphi$ (agent a acquires information φ from b), and $T_{a,b}\varphi$ (agent a trusts the judgement of b on the truth of φ). The semantics of modalities B and I are Kripke-style, while the one for modality T is neighbourhood-based. Certain connections between these semantics are assumed. [Perrotin *et al.*, 2019] proposed a logical system that describes the interplay between beliefs, trust, and public group announcements. In their system, trust is semantically modelled through set T_a^w of all agents whom agent a trusts in state w. This set resembles set $\mathcal{T}_w$ in our semantics. In their semantics, beliefs are defined using belief bases. As public announcements are made, the set of agents T_a^w to whom agent a trusts is updated based on the agent's belief base. Thus, in their system, beliefs define trust, while in ours trust defines beliefs. The syntax of their system includes trust propositional variable $T_{a,b}$ (agent a trusts agent b) and belief modality $B_a\varphi$ (agent a beliefs in statement φ). The only axiom of their system that includes both trust propositional variable and belief modality is the axiom $T_{a,b} \to (B_aB_b\varphi \to B_a\varphi)$. It is interesting to note that this axiom could be roughly translated into our language as $\mathsf{B}^T_X\mathsf{B}^T_Y\varphi \to \mathsf{B}^T_X\varphi$. The last statement is provable in our system through a combination of the Trust and the Distributivity axioms. Unlike our work, [Perrotin *et al.*, 2019] and [Liau, 2003] do not consider data-informed beliefs.

6 Completeness

In this section, we prove the completeness of our system.

6.1 Canonical Model

As usual, at the core of the proof of the completeness is the construction of a canonical model. The goal of this section is to define canonical trustworthiness model $M(T_0, F_0) = (W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$ for any dataset $T_0 \subseteq V$ and any maximal consistent set of formula $F_0 \subseteq \Phi$.

Definition 3. *Set of worlds W is the set of all sequences $T_0, F_0, X_1, T_1, F_1, \dots, X_n, T_n, F_n$ such that $n \geq 0$ and, for each i where $0 \leq i \leq n$,*

1. *$X_i, T_i \subseteq V$ are datasets,*
2. *F_i is a maximal consistent set of formulae such that*
   (a) *$\psi \in F_i$ for each formula $\mathsf{B}^{\varnothing}_{X_i}\psi \in F_{i-1}$, if $i > 0$,*
   (b) *$\mathsf{B}^{T_i}_Y\varphi \to \varphi \in F_i$ for each dataset $Y \subseteq V$ and each formula $\varphi \in \Phi$.*

For any worlds $w', w \in W$ such that

$$
\begin{aligned}
w' &= T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}\\
w &= T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}, X_n, T_n, F_n
\end{aligned}
$$

we say that worlds w' and w are *adjacent*. The adjacency relation defines a tree structure on set W. We say that the *edge* between nodes w' and w of this tree is *labelled* with all variables in dataset X_n and that the *node* w is *labelled* with the pair T_n, F_n. By $T(w)$ and $F(w)$ we mean sets T_n and F_n respectively.

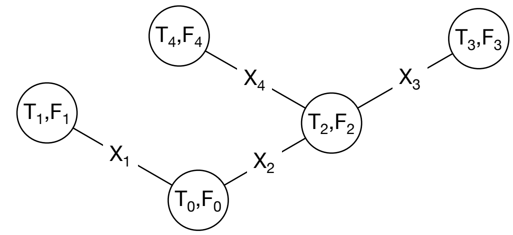

Figure 1: Fragment of tree W.

It will be convenient to visualise tree W as shown in Figure 1. In this figure, the world $T_0, F_0, X_2, T_2, F_2, X_4, T_4, F_4$ is adjacent to the world T_0, F_0, X_2, T_2, F_2. The edge between these two worlds is labelled by all variables in dataset X_4.

Definition 4. *For any worlds $u, w \in W$ and any data variable $x \in V$, let $u \sim_x w$ if every edge along the unique simple path between vertices u and w is labelled with variable x.*

Lemma 3. *Relation $\sim_x$ is an equivalence relation on set W for each data variable $x \in V$.*

Definition 5. $\mathcal{T}_w = T(w)$.

Definition 6. $\pi(p) = \{w \in W \mid p \in F(w)\}$.

This concludes the definition of the canonical trustworthiness model $M(F_0) = (W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$.

6.2 Properties of the Canonical Model

As common in modal logic, at the core of the proof of the completeness is a truth lemma. In our case, this is Lemma 7. Lemma 5 and Lemma 6 are used in the induction step of the proof of the truth lemma. Lemma 4 below is an auxiliary result used in the proof of Lemma 5.

Lemma 4. *For any formula $\mathsf{B}^T_Y\varphi \in \Phi$ and any worlds*

$$
\begin{aligned}
w' &= T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}\\
w &= T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}, X_n, T_n, F_n
\end{aligned}
$$

if $Y \subseteq X_n$, then $\mathsf{B}^T_Y\varphi \in F(w')$ iff $\mathsf{B}^T_Y\varphi \in F(w)$.

Proof. $(\Rightarrow)$: Suppose $\mathsf{B}^T_Y\varphi \in F(w')$. Thus, $\mathsf{B}^T_Y\varphi \in F_{n-1}$. Then, by Lemma 1 and the Modus Ponens inference rule, $F_{n-1} \vdash \mathsf{B}^{\varnothing}_Y\mathsf{B}^T_Y\varphi$. Hence, $F_{n-1} \vdash \mathsf{B}^{\varnothing}_{X_n}\mathsf{B}^T_Y\varphi$ by the assumption $Y \subseteq X_n$ of the lemma, the Monotonicity axiom, and the Modus Ponens inference rule. Then, $\mathsf{B}^{\varnothing}_{X_n}\mathsf{B}^T_Y\varphi \in F_{n-1}$ because F_{n-1} is a maximal consistent set. Thus, $\mathsf{B}^T_Y\varphi \in F_n$ by item 2(a) of Definition 3. Therefore, $\mathsf{B}^T_Y\varphi \in F(w)$.

$(\Leftarrow)$: Suppose $\mathsf{B}^T_Y\varphi \notin F(w')$. Then, $\mathsf{B}^T_Y\varphi \notin F_{n-1}$. Thus, $\neg\mathsf{B}^T_Y\varphi \in F_{n-1}$ because F_{n-1} is a maximal consistent set of formulae. Hence, $F_{n-1} \vdash \mathsf{B}^{\varnothing}_Y\neg\mathsf{B}^T_Y\varphi$ by the Negative Introspection axiom and the Modus Ponens inference rule. Thus, $F_{n-1} \vdash \mathsf{B}^{\varnothing}_{X_n}\neg\mathsf{B}^T_Y\varphi$ by the assumption $Y \subseteq X_n$ of the lemma, the Monotonicity axiom, and the Modus Ponens inference rule. Then, again because set F_{n-1} is maximal, $\mathsf{B}^{\varnothing}_{X_n}\neg\mathsf{B}^T_Y\varphi \in F_{n-1}$. Thus, $\neg\mathsf{B}^T_{X_n}\varphi \in F_n$ by item 2(a) of Definition 3. Hence, $\mathsf{B}^T_Y\varphi \notin F_n$, because set F_n is consistent. Therefore, $\mathsf{B}^T_Y\varphi \notin F(w)$. □

Lemma 5. *For any worlds $w, u \in W$ and any formula $\mathsf{B}^T_X\varphi \in F(w)$, if $w \sim_X u$ and $T \subseteq \mathcal{T}_u$, then $\varphi \in F(u)$.*

Proof. By Definition 4, the assumption $w \sim_X u$ implies that each edge along the unique path between nodes w and u is labelled with each variable in dataset X. Then, the assumption $\mathsf{B}^T_X\varphi \in F(w)$ implies $\mathsf{B}^T_X\varphi \in F(u)$ by applying Lemma 4 to each edge along this path. Note that the assumption $T \subseteq \mathcal{T}_u$ of the lemma implies that $T \subseteq T(u)$ by Definition 5. Thus, $F(u) \vdash \mathsf{B}^{T(u)}_X\varphi$ by the Monotonicity axiom and the Modus Ponens inference rule. Hence, $F(u) \vdash \varphi$ by item 2(b) of Definition 3 and the Modus Ponens inference rule. Therefore, $\varphi \in F(u)$ because the set $F(u)$ is maximal. □

Lemma 6. *For any $w \in W$ and any formula $\mathsf{B}^T_X\varphi \notin F(w)$, there exists a world $u \in W$ such that $w \sim_X u$, $T \subseteq \mathcal{T}_u$, and $\varphi \notin F(u)$.*

Proof. Consider the set of formulae

$$\begin{aligned} G \;=\; & \{\neg\varphi\} \cup \{\psi \mid \mathsf{B}^{\varnothing}_X\psi \in F(w)\} \\ & \cup\{\mathsf{B}^T_Y\chi \to \chi \mid Y \subseteq V, \chi \in \Phi\} \end{aligned} \tag{7}$$

Claim 1. *Set G is consistent.*

PROOF OF CLAIM. Assume the opposite. Thus, there are formulae $\chi_1, \dots, \chi_n \in \Phi$, datasets $Y_1, \dots, Y_n \subseteq V$, and formulae

$$\mathsf{B}^{\varnothing}_X\psi_1, \dots, \mathsf{B}^{\varnothing}_X\psi_m \in F(w) \tag{8}$$

such that

$$\mathsf{B}^T_{Y_1}\chi_1 \to \chi_1, \dots, \mathsf{B}^T_{Y_n}\chi_n \to \chi_n, \psi_1, \dots, \psi_m \vdash \varphi.$$

Hence, by Lemma 2,

$$\begin{aligned} &\mathsf{B}^T_X(\mathsf{B}^T_{Y_1}\chi_1 \to \chi_1), \dots, \mathsf{B}^T_X(\mathsf{B}^T_{Y_n}\chi_n \to \chi_n), \\ &\qquad\qquad \mathsf{B}^T_X\psi_1, \dots, \mathsf{B}^T_X\psi_m \vdash \mathsf{B}^T_X\varphi. \end{aligned}$$

Then, $\mathsf{B}^T_X\psi_1, \dots, \mathsf{B}^T_X\psi_m \vdash \mathsf{B}^T_X\varphi$ by the Trust axiom applied n times. Thus, $\mathsf{B}^{\varnothing}_X\psi_1, \dots, \mathsf{B}^{\varnothing}_X\psi_m \vdash \mathsf{B}^T_X\varphi$ by the Monotonicity axiom and the Modus Ponens inference rule applied m times. Hence, $F(w) \vdash \mathsf{B}^T_X\varphi$ due to statement (8). Then, $\mathsf{B}^T_X\varphi \in F(w)$ because the set $F(w)$ is maximal, which contradict the assumption $\mathsf{B}^T_X\varphi \notin F(w)$ of the lemma. □

Let G' be any maximal consistent extension of set G. Suppose that $w = T_0, F_0, \dots, X_n, T_n, F_n$. Consider sequence

$$u = T_0, F_0, \dots, X_n, T_n, F_n, X, T, G'. \tag{9}$$

Note that $u \in W$ by Definition 3, equation (7), and the choice of set G' as an extension of set G. Also, observe that $w \sim_X u$ by Definition 4 and equation (9). In addition, $T = T(u) = \mathcal{T}_u$ by equation (9) and Definition 5. Finally, $\neg\varphi \in G \subseteq G' = F(u)$ by equation (7), the choice of G' as an extension of G, and equation (9).Therefore, $\varphi \notin F(u)$ because the set $F(u)$ is consistent. This concludes the proof of the lemma. □

Lemma 7. *$w \Vdash \varphi$ iff $\varphi \in F(w)$, for any world $w \in W$ and any formula $\varphi \in \Phi$.*

Proof. We prove the lemma by induction on structural complexity of formula φ. If formula φ is a propositional variable, then the statement of the lemma follows from Definition 6 and item 1 of Definition 2.

If formula φ is a negation or an implication, then the statement of the lemma follows from the induction hypothesis, items 2 and 3 of Definition 2 and the maximality and consistency of set $F(w)$ in the standard way.

Let us now suppose that formula φ has the form $\mathsf{B}^T_X\psi$.
$(\Rightarrow)$: If $\mathsf{B}^T_X\psi \notin F(w)$ then, by Lemma 6, there exists a world $u \in W$ such that $w \sim_X u$, $T \subseteq \mathcal{T}_u$, and $\psi \notin F(u)$. Thus, $u \nVdash \psi$ by the induction hypothesis. Therefore, $w \nVdash \mathsf{B}^T_X\psi$ by item 4 of Definition 2.
$(\Leftarrow)$: Consider any world u such that $w \sim_X u$ and $T \subseteq \mathcal{T}_u$. By item 4 of Definition 2, it suffices to show that $u \Vdash \psi$. By Lemma 5, the assumptions $\mathsf{B}^T_X\psi \in F(w)$, $w \sim_X u$, and $T \subseteq \mathcal{T}_u$ imply $\psi \in F(u)$. Therefore, $u \Vdash \psi$ by the induction hypothesis. □

6.3 Completeness: Final Step

Theorem 2 (strong completeness)**.** *For any set of formulae $F \subseteq \Phi$ and any formula $\varphi \in \Phi$, if $F \nvdash \varphi$, then there is a world w of a trustworthiness model such that $w \Vdash f$ for each formula $f \in F$ and $w \nVdash \varphi$.*

Proof. The assumption $F \nvdash \varphi$ implies that the set $F \cup \{\neg\varphi\}$ is consistent. Let F_0 be any maximal consistent extension of this set. Consider the canonical model $M(\varnothing, F_0)$.

First, we show that the sequence $\varnothing, F_0$ is a world of this canonical model. By Definition 3, it suffices to show that $\mathsf{B}^{\varnothing}_Y\psi \to \psi \in F_0$ for each dataset $Y \subseteq V$ and each formula $\psi \in \Phi$. The last statement is true by the Truth axiom and because set F_0 is maximal.

Next, note that $\varphi \notin F_0$ because set F_0 is consistent and $\neg\varphi \in F_0$. Then, by Lemma 7 and because $F \subseteq F_0$, it follows that $\varnothing, F_0 \Vdash f$ for each formula $f \in F$ and $\varnothing, F_0 \nVdash \varphi$. □

7 Future Work

By focusing on data instead of agents, we are able to significantly simplify the settings of the previous works on beliefs and trust. This creates an opportunity for extensions of the proposed logical system in the future.

7.1 Functional Dependency

One of such extension is the incorporation of Armstrong's functional dependency relation $X \rhd Y$ [Armstrong, 1974] into the logic. Informally, $X \rhd Y$ means that the values of the variables in dataset X inform (or functionally determine) the values of the variables in dataset Y. The term "inform" can be interpreted in two ways: globally (in each world) and locally (in the current world). By $X \rhd Y$ we denote the local interpretation. The global interpretation could be expressed as $\mathsf{B}^{\varnothing}_{\varnothing}(X \rhd Y)$. The formal definition of relation $X \rhd Y$ is below.

Definition 7. *For any world $w \in W$ of trustworthiness model $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$ and any datasets $X, Y \subseteq V$, let $w \Vdash X \rhd Y$ when for each world $u \in W$, if $w \sim_X u$, then $w \sim_Y u$.*

If the language of our logical system is extended with the primitive proposition $X \rhd Y$ and Definition 7 is incorporated into Definition 2, then the following additional axioms are sound with respect to the modified semantics:

A6. Reflexivity: $X \rhd Y$, where $Y \subseteq X$,

A7. Transitivity: $X \rhd Y \to (Y \rhd Z \to X \rhd Z)$,

A8. Augmentation: $X \rhd Y \to X \cup Z \rhd Y \cup Z$,

A9. Monotonicity: $X \rhd Y \to (\mathsf{B}^T_Y\varphi \to \mathsf{B}^T_X\varphi)$.

The first three of these axioms are known in database theory as Armstrong's axioms [Garcia-Molina *et al.*, 2009, p. 81]. The complete axiomatisation of the interplay between relation $X \rhd Y$ and the data-informed belief modality $\mathsf{B}^T_X\varphi$ remains an open question.

7.2 Public Announcements

Another interesting possible extension of our logic is by a public announcement modality. Given the data focus of our logical system, it makes sense to consider public announcement of values of datasets rather than of true formulae. Such modality has been first introduced in [van Eijck *et al.*, 2017] under name "public inspection". We use notation $[X]\varphi$ for modality "formula φ holds after the values of all variables in dataset X are publicly announced". To formally define the semantics of this modality, we modify satisfaction relation from a binary relation $w \Vdash \varphi$ to a ternary relation $w, U \Vdash \varphi$. It reads "formula φ is satisfied in world w after a public announcement of the values of all variables in dataset U".

To change from binary form of relation $\Vdash$ to the ternary one, we first need to slightly modify Definition 1. Namely, in item 4 we will assume that $\pi(p)$ is a set of pairs (w, U), where $w \in W$ is a world and $U \subseteq V$ is a dataset. Informally, $(w, U) \in \pi(p)$ if propositional variable p holds in world w after a public announcement of the values of all variables in dataset U. Then, Definition 2 could be modified as follows to define the ternary form of the satisfaction relation.

Definition 8. *For any world $w \in W$ of any trustworthiness model $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$, any dataset $U \subseteq V$, and any formula $\varphi \in \Phi$, satisfaction relation $w, U \Vdash \varphi$ is defined as follows:*

1. *$w, U \Vdash p$ if $(w, U) \in \pi(p)$,*
2. *$w, U \Vdash \neg\varphi$ if $w, U \nVdash \varphi$,*
3. *$w, U \Vdash X \rhd Y$ when for each $v \in W$, if $w \sim_{X\cup U} v$, then $w \sim_Y v$,*
4. *$w, U \Vdash \varphi \to \psi$ if $w, U \nVdash \varphi$ or $w, U \Vdash \psi$,*
5. *$w, U \Vdash \mathsf{B}^T_X\varphi$ if $v, U \Vdash \varphi$ for each world $v \in W$ such that $v \sim_{X\cup U} u$ and $T \subseteq \mathcal{T}_u$,*
6. *$w, U \Vdash [X]\varphi$ if $w, U \cup X \Vdash \varphi$,*

In the classical logic of public announcements it is assumed that only true formulae can be announced [van Ditmarsch *et al.*, 2007, Chapter 4]. Similar, in the logic of public inspections, the "true" values of the variables are announced [van Eijck *et al.*, 2017]. The same is technically true in our semantics given above. However, in our setting the announced values *do not have to be trustworthy*. For example, a newspaper prediction could be publicly announced even if the prediction is wrong. Such announcement is "true" because the newspaper indeed made such a prediction, but this data is not trustworthy because the prediction itself is wrong. The ability to reason about such announcements is a unique feature of our approach that distinguishes it from the previous works.

The following addition axioms capture the interplay between data-informed beliefs, functional dependency, and public announcements:

A10. Distributivity: $[X](\varphi \to \psi) \to ([X]\varphi \to [X]\psi)$,

A11. Combination: $[X][Y]\varphi \leftrightarrow [X \cup Y]\varphi$,

A12. Duality: $\neg[X]\varphi \leftrightarrow [X]\neg\varphi$,

A13. Introspection of Dependency: $X \rhd Y \to \mathsf{B}^T_X(X \rhd Y)$,

A14. Perfect Recall: $\mathsf{B}^T_X[Y]\varphi \to [Y]\mathsf{B}^T_X\varphi$,

A15. Public Knowledge: $[X](\mathsf{B}^T_{X\cup Y}\varphi \to \mathsf{B}^T_Y\varphi)$,

A16. Prior Belief: $[X]\mathsf{B}^T_Y\varphi \to \mathsf{B}^T_{X\cup Y}[X]\varphi$,

A17. Partial Announcement: $(X \cup Y) \rhd Z \leftrightarrow [X](Y \rhd Z)$,

A18. Empty Announcement: $\varphi \to [\varnothing]\varphi$.

The complete axiomatisation of these properties (or even the properties of modalities B^T_X and $[X]$ without the functional dependency) is another question that we leave for the future.

8 Conclusion

The existing literature on the logics of trust assumes that trust is a relation between agents. In this paper, we proposed to study trustworthiness as a property of data. We introduced trust-based beliefs that are formed by the interplay between the known data on one hand and the trusted data on the other. We proposed a sound and complete logical system describing such beliefs. We also discussed possible extensions of this system with Armstrong's function dependency relation and public announcement modality.

Acknowledgments

Junli Jiang acknowledges the support of the Key Research Funds for the Key Liberal Science Research Base of Chongqing (NO.16SKB036) and the Fundamental Research Funds for the Central Universities (SWU1409412).

References

[Armstrong, 1974] W. W. Armstrong. Dependency structures of data base relationships. In *Information Processing 74 (Proc. IFIP Congress, Stockholm, 1974)*, pages 580–583. North-Holland, Amsterdam, 1974.

[Baltag and Smets, 2006] Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5 – 21, 2006.

Proceedings of the 13th Workshop on Logic, Language, Information and Computation (WoLLIC 2006).

[Baltag and Smets, 2008] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the foundations of game and decision theory (LOFT 7)*, 3:9–58, 2008.

[Baltag and van Benthem, 2021] Alexandru Baltag and Johan van Benthem. A simple logic of functional dependence. *Journal of Philosophical Logic*, 50:1–67, 2021.

[Baltag, 2016] Alexandru Baltag. To know is to know the value of a variable. In Lev D. Beklemishev, Stéphane Demri, and András Maté, editors, *Advances in Modal Logic 11, proceedings of the 11th conference on Advances in Modal Logic, Budapest, Hungary, August 30 – September 2, 2016*, pages 135–155. College Publications, 2016.

[Board, 2004] Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49(1):49–80, 2004.

[Boutilier, 1994] Craig Boutilier. Conditional logics of normality: a modal approach. *Artificial intelligence*, 68(1):87–154, 1994.

[Castelfranchi and Falcone, 1998] Cristiano Castelfranchi and Rino Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160)*, pages 72–79. IEEE, 1998.

[Friedman and Halpern, 1997] Nir Friedman and Joseph Y Halpern. Modeling belief in dynamic systems, part I: foundations. *Artificial intelligence*, 95(2):257–316, 1997.

[Friedman and Halpern, 1999] Nir Friedman and Joseph Y Halpern. Modeling belief in dynamic systems, part II: Revision and update. *Journal of Artificial Intelligence Research*, 10:117–167, 1999.

[Garcia-Molina *et al.*, 2009] Hector Garcia-Molina, Jeffrey Ullman, and Jennifer Widom. *Database Systems: The Complete Book*. Prentice-Hall, second edition, 2009.

[Girlando *et al.*, 2016] Marianna Girlando, Sara Negri, Nicola Olivetti, and Vincent Risch. The logic of conditional beliefs: Neighbourhood semantics and sequent calculus. In *Advances in Modal Logic*. College Publications, 2016.

[Girlando *et al.*, 2019] Marianna Girlando, Bjoern Lellmann, and Nicola Olivetti. Nested sequents for the logic of conditional belief. In *European Conference on Logics in Artificial Intelligence*, pages 709–725. Springer, 2019.

[Herzig *et al.*, 2010] Andreas Herzig, Emiliano Lorini, Jomi F Hübner, and Laurent Vercouter. A logic of trust and reputation. *Logic Journal of the IGPL*, 18(1):214–244, 2010.

[Lewis, 1973] David K Lewis. *Counterfactuals*. Harvard University Press, 1973.

[Liau, 2003] Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems—a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.

[Miner More and Naumov, 2009] Sara Miner More and Pavel Naumov. An independence relation for sets of secrets. In H. Ono, M. Kanazawa, and R. de Queiroz, editors, *Proceedings of 16th Workshop on Logic, Language, Information and Computation (Tokyo, 2009), LNAI 5514*, pages 296–304. Springer, 2009.

[Mitra *et al.*, 2011] Gautam Mitra, Dan diBartolomeo, Ashok Banerjee, and Xiang Yu. Automated analysis of news to compute market sentiment: Its impact on liquidity and trading. Technical Report The Future of Computer Trading in Financial Markets, Foresight Driver Review, DR 8, UK Government Office for Science, July 2011.

[Perrotin *et al.*, 2019] Elise Perrotin, Rustam Galimullin, Quentin Canu, and Natasha Alechina. Public group announcements and trust in doxastic logic. In *International Workshop on Logic, Rationality and Interaction*, pages 199–213. Springer, 2019.

[Primiero, 2020] Giuseppe Primiero. A logic of negative trust. *Journal of Applied Non-Classical Logics*, 30(3):193–222, 2020.

[Sutherland, 1986] David Sutherland. A model of information. In *Proceedings of Ninth National Computer Security Conference*, pages 175–183, 1986.

[Tagliaferri and Aldini, 2019] Mirko Tagliaferri and Alessandro Aldini. A trust logic for the varieties of trust. In *International Conference on Software Engineering and Formal Methods*, pages 119–136. Springer, 2019.

[van Ditmarsch *et al.*, 2007] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer, 2007.

[van Eijck and Li, 2017] Jan van Eijck and Kai Li. Conditional belief, knowledge and probability. In *Proceedings of the 16th conference on Theoretical aspects of rationality and knowledge*, pages 188–206. EPTCS 251, 2017.

[van Eijck *et al.*, 2017] Jan van Eijck, Malvin Gattinger, and Yanjing Wang. Knowing values and public inspection. In *Indian Conference on Logic and Its Applications*, pages 77–90. Springer, 2017.

[von Beschwitz *et al.*, 2020] Bastian von Beschwitz, Donald B Keim, and Massimo Massa. First to "read" the news: News analytics and algorithmic trading. *The Review of Asset Pricing Studies*, 10(1):122–178, 2020.

[Wang and Fan, 2014] Yanjing Wang and Jie Fan. Conditionally knowing what. In Rajeev Goré, Barteld P. Kooi, and Agi Kurucz, editors, *Advances in Modal Logic 10, invited and contributed papers from the tenth conference on Advances in Modal Logic, held in Groningen, The Netherlands, August 5-8, 2014*, pages 569–587. College Publications, 2014.