A Logic of Trust-Based Beliefs

Junli Jiang[1] and Pavel Naumov[2*]

[1]Institute of Logic and Intelligence, Southwest University, Chongqing, China.

[2*]University of Southampton, Southampton, United Kingdom.

*Corresponding author(s). E-mail(s): p.naumov@soton.ac.uk;
Contributing authors: walk08@swu.edu.cn;

Abstract

Traditionally, knowledge and beliefs are attributed to agents. The article explores an alternative approach where knowledge is informed by data and belief comes from trust in, not necessarily reliable, data. At the core of the article, is the modality "if one dataset is trusted, then another dataset informs a belief". The main technical result is a sound and complete logical system capturing the properties of this modality and its connection with functional dependency between datasets.

1 Introduction

1.1 Trust, Trustworthiness, and Beliefs

The distinction between *trust* towards people and their *trustworthiness* has been discussed in econimics (Chaudhuri & Gangadharan, 2007), political science (Hardin, 2002; Levi & Stoker, 2000), and psychology (Ben-Ner & Halldorsson, 2010; Posten & Mussweiler, 2019) literature. In the words of The Stanford Encyclopedia of Philosophy,

> Trust is an attitude we have towards people whom we hope will be trustworthy, where trustworthiness is a property not an attitude. Trust and trustworthiness are therefore distinct although, ideally, those whom we trust will be trustworthy, and those who are trustworthy will be trusted (McLeod, 2023).

In this article, we propose to extend this distinction from people to information. We capture the information through a set of variables whose values can vary from one possible world to another. We refer to such variables as data variables and to their sets

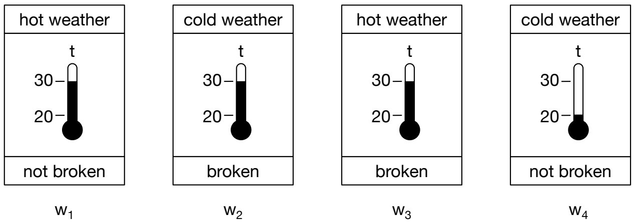

Fig. 1 A fragment of a trustworthiness model.

as *datasets*. An agent might trust a dataset no matter if the dataset is trustworthy or not. Based on this trust, the agent can form a *belief* about the world. If the dataset based on which the belief is formed is trustworthy, then the belief is true. Otherwise, it might be false.

1.2 Motivational Example

Consider a setting where the weather outside could be either hot (more than 25°C) or cold (at most 25°C). There is a thermometer that is supposed to show the outside temperature, but it might be broken. Figure 1 depicts *some* of the possible worlds in a model capturing this setting. For example, in world w_1, it is hot outside, the thermometer is not broken, and it shows 30°C. Note that, in world w_3, it is hot outside and the broken thermometer happens to show 30°C (as the saying goes, even a broken clock is right twice a day). We represent the reading of the thermometer by data variable t. We say that variable t is *trustworthy* in worlds w_1 and w_4 where the thermometer is not broken. We refer to the model whose fragment is depicted in Figure 1 as *trustworthiness model*.

Let us now consider the statement "if $t > 25$, then it is hot outside". Note that this statement is true in worlds w_1, w_3, and w_4 and is false in world w_2. Any agent that *trusts* the thermometer would assume that the thermometer is not broken and, thus, exclude worlds such as w_2 and w_3 from the consideration. In the remaining worlds, the statement "if $t > 25$, then it is hot outside" is true. We say that *trust* in data variable t forms the belief that the statement is true and write this as

$$\mathsf{B}^t(\text{“if } t > 25\text{, then it is hot outside”}). \tag{1}$$

Observe that this belief is formed by trust in data variable t no matter what is the current world. In worlds w_1, w_3, and w_4 this belief is true and in world w_2 it is false. Intuitively, the reason why the belief is false in world w_2 is that the belief is based on *trust* in data variable which is *not trustworthy* in w_2.

Next, let us consider any agent in world w_1 that trusts data variable t and also knows the value of this variable. Such an agent will exclude from consideration the worlds in which the thermometer is broken, such as worlds w_2 and w_3. In addition, such an agent would exclude worlds in which the value of data variable t is different from its value in the current world w_1. An example of a latter world is w_4. In the remaining worlds, such as world w_1 in our figure, $t = 30$ and the thermometer is not broken. In such worlds, it must be hot outside. Thus, in world w_1, trust in data variable

t and the knowledge of its value form a belief that it is hot outside. We write this as

$$w_1 \Vdash \mathsf{B}_t^t(\text{“it is hot outside”}) \tag{2}$$

and say that, in world w_1, if data variable t is trusted, then t *informs* the belief that it is hot outside. Later in this section, we will see examples when the superscript (what is trusted) of modality B is different from its subscript (what is known). Note that statement (2) is true not only for world w_1, but also for worlds w_2 and w_3. Just like in our previous example, the belief in world w_2 is a false belief because in world w_2 it is actually cold outside. The statement (2) is not true for world w_4 because the value of variable t in world w_4 is 20. The weather is cold in all worlds where t is trustworthy and has a value of 20. Thus,

$$w_4 \Vdash \mathsf{B}_t^t(\text{“it is cold outside”}).$$

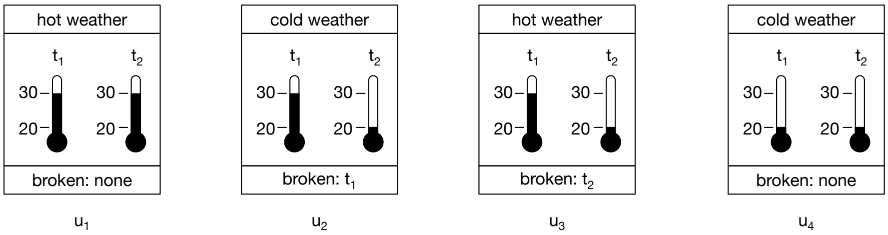

Fig. 2 A fragment of a trustworthiness model with two thermometers.

Let us now consider a more interesting example where there are two thermometers either of which (or both) can be broken. Some of the possible worlds in this setting are shown in Figure 2. Note that any agent that trusts both thermometers would eliminate from consideration all worlds where at least one of the thermometers is broken, such as worlds u_2 and u_3. In the remaining worlds, such as u_1 and u_4, the two thermometers must show the same temperature. Thus, just trust in data variables t_1 and t_2 forms the belief that these two variables must have equal values:

$$\mathsf{B}^{t_1 t_2}(t_1 = t_2). \tag{3}$$

Just like in the case of the belief from statement (1), the above belief forms no matter what is the current world. Next, let's assume that, in world u_1, one knows the value of t_1, which is 30. If this *data knowledge* is combined with *trust* in data variables t_1 and t_2, then in addition to elimination of the world where either of the thermometers is broken (such as u_2 and u_3) one would also eliminate the worlds where $t_1 \neq 30$ (such as u_4). In the remaining worlds, just like in u_1 itself, $t_1 = t_2 = 30$. In all such worlds, for example, the statement $t_2 > 25$ holds. Thus, if data set $\{t_1, t_2\}$ is trusted, then the data variable t_1 informs the belief $t_2 > 25$:

$$u_1 \Vdash \mathsf{B}_{t_1}^{t_1 t_2}(t_2 > 25). \tag{4}$$

The same belief is also informed in worlds u_2 and u_3, where the reading of the first thermometer is also 30°C, but it is not informed in world u_4.

Let us now show that both data variables, t_1 and t_2, are needed in the superscript of the modality B in order for statement (4) to be true. Indeed, if data variable t_1 is not trusted, then the world u_2 is not eliminated from the consideration. Since $t_2 = 20$ in world u_2,

$$u_1 \nVdash \mathsf{B}^{t_2}_{t_1}(t_2 > 25).$$

Similarly, if data variable t_2 is not trusted, then world u_3 is not eliminated. Since $t_2 = 20$ in world u_3,

$$u_1 \nVdash \mathsf{B}^{t_1}_{t_1}(t_2 > 25).$$

Finally, similarly to (2), note that if an agent trusts data variable t_2 and knows that its value is 30, then the agent forms a belief that it is hot outside:

$$v \Vdash \mathsf{B}^{t_2}_{t_2}(\text{“it is hot outside”}) \quad \text{if } t_2 \text{ has value 30 in } v. \tag{5}$$

Let us now consider an agent in world u_1 that trusts t_1 and t_2, but knows only the value of t_1. Because in u_1 data variable t_1 has value 30 and the agent knows this, the agent will eliminate from consideration all worlds in which $t_1 \neq 30$, leaving only the worlds where $t_1 = 30$. Because the agent trusts t_1 and t_2, the agent will eliminate all worlds where at least one of the thermometers is broken, leaving only worlds where $t_2 = t_1$. Thus, the agent will only consider worlds where $t_2 = t_1 = 30$. As we observed in (5), the formula $\mathsf{B}^{t_2}_{t_2}(\text{“it is hot outside”})$ holds in all such worlds. Therefore,

$$u_1 \Vdash \mathsf{B}^{t_1 t_2}_{t_1}\mathsf{B}^{t_2}_{t_2}(\text{“it is hot outside”}). \tag{6}$$

It is interesting to point out that the same second-order belief is also informed in worlds u_2 and u_3. However, unlike in world u_1, in worlds u_2 and u_3 this is a false second-order belief.

1.3 Contribution

The main contribution of this work is a formal semantics and a complete axiomatisation of modality $\mathsf{B}^T_X\varphi$ which stands for “if dataset T is trusted, then dataset X informs the belief φ”.

The modality $\mathsf{B}^T_X\varphi$ is closely related to conditional belief modality $\mathsf{B}^\psi_a\varphi$ which is completely axiomatised by Board (2004). In turn, conditional belief modality is similar to Lewis' (1973) counterfactual modality $\psi \,\Box\!\!\rightarrow \varphi$. Informally, $\mathsf{B}^\psi_a\varphi$ means that the agent believes φ after receiving the information that ψ. Using this modality, statement (2) from the introduction can be written as

$$w_1 \Vdash \mathsf{B}_a^{\text{“the thermometer is not broken”}}(\text{“it is hot outside”})$$

for some agent a who can see the thermometer. The major advantage of the modality $\mathsf{B}^T_X\varphi$, that we propose, is the ability to separate the two components that form a belief: the subjective mindset of the agent captured by parameter T and the objective information about the world specified by parameter X.

The other advantage of our approach is that it can be combined with data-specific modalities and operators. One such operator is Armstrong's (1974) functional dependency expression $X \rhd Y$. This expression denotes the fact that the values of the variables in dataset X functionally determine the values of the variables in dataset Y. In the current work, we included the expression $X \rhd Y$ into the language of our logical system. In Section 8 we discuss other possible extensions by data-specific modalities and operators.

The rest of the article is structured as follows. First, we review related literature on trust and beliefs. Then, We define trustworthiness models that are used later to give a formal semantics of our logical system. In Section 4, we introduce the syntax and the formal semantics of the system. In Section 5, we list and discuss its axioms and the inference rules. Their soundness is proven in Section 6. Section 7 contains the proof of the completeness theorem. Section 8 discusses possible extensions of our system. Section 9 concludes. The initial version of this work, without the operator $\rhd$ and the proof of completeness, appeared as (Jiang & Naumov, 2022b).

2 Literature Review

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

The closest works to ours are (Liau, 2003) and (Perrotin, Galimullin, Canu, & Alechina, 2019). Liau (2003) introduced a logical system describing the interplay bewteen 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, Galimullin, Canu, and Alechina (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. 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 an atomic trust proposition $T_{a,b}$ (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 the trust atomic proposition and the belief modality is the axiom $T_{a,b} \to (B_aB_b\varphi \to B_a\varphi)$. Intuitively, this axiom corresponds to the formula $\mathsf{B}_X^T\mathsf{B}_Y^T\varphi \to \mathsf{B}_X^T\varphi$ in our language. The last statement is provable in our system through a combination of the Trust and the Distributivity axioms. Unlike our work, (Perrotin, Galimullin, Canu, & Alechina, 2019) and (Liau, 2003) do not consider data-informed beliefs.

3 Trustworthiness Model

Lewis (1973) used sphere semantics for modality $\Box\!\!\rightarrow$. This semantics has been later generalised to neighbourhood semantics (Girlando, Lellmann, & Olivetti, 2019; Girlando, Negri, Olivetti, & Risch, 2016; van Eijck & Li, 2017). Another type of semantics for modality $\Box\!\!\rightarrow$ is plausibility semantics (Baltag & Smets, 2006, 2008; Board, 2004; Boutilier, 1994; Friedman & Halpern, 1997, 1999). These semantics do not capture trust and, thus, can not be used to model trust-based beliefs.

To model trust-based beliefs, we propose trustworthiness models inspired by our informal models in Figure 1 and Figure 2. Trust is a broad term with multiple meanings that can be formalised in many different ways. The focus of the current work is on trust-based beliefs, not trust. Our trustworthiness models are not meant to provide a way to model trust in general. Instead, they aim to capture the aspect of the trust needed to define the formal semantics of trust-based beliefs.

Throughout the rest of the article, we assume a fixed finite set of data variables V and an arbitrary set of atomic propositions. 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. *relation $\sim_x$ is an "indistinguishability" equivalence relation on set W for each data variable $x \in V$,*
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 atomic proposition p.*

4 Syntax and Semantics

The language Φ of our logical system is defined by the grammar:

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

where p is an atomic proposition and $X, Y, T \subseteq V$ are datasets. We read $X \rhd Y$ as "the values of the variables in dataset X *inform* the values of the variables in dataset Y", and we read $\mathsf{B}^T_X\varphi$ as "if dataset T is trusted, then dataset X informs the belief in φ". We assume that $\bot$ is formula $\neg(p \to p)$, where p is one of the atomic propositions. When the arguments in the expression $X \rhd Y$ are given explicitly, we omit curly brackets to improve readability. For example, we write $x_1, x_2 \rhd y$ instead $\{x_1, x_2\} \rhd \{y\}$.

In the definition below and the rest of the article, 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 formula $\varphi \in \Phi$ and any world $w \in W$ of any trustworthiness model $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$, the satisfaction relation $w \Vdash \varphi$ is defined as follows:*

1. *$w \Vdash p$ if $w \in \pi(p)$, where p is an atomic proposition,*
2. *$w \Vdash X \rhd Y$ if $w \sim_Y u$ for each world $u \in W$ such that $w \sim_X u$,*
3. *$w \Vdash \neg\varphi$ if $w \nVdash \varphi$,*

4. $w \Vdash \varphi \to \psi$ *if* $w \nVdash \varphi$ *or* $w \Vdash \psi$,
5. $w \Vdash \mathsf{B}^T_X\varphi$ *if* $u \Vdash \varphi$ *for each world* $u \in W$ *such that* $w \sim_X u$ *and* $T \subseteq \mathcal{T}_u$.

Observe that statement $w \Vdash \mathsf{B}^T_X\bot$ is true if there is no X-indistinguishable from w world in which all variables in dataset T are trustworthy.

Note also 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 may say that the *knowledge* of φ is informed by dataset X in world w. Modality $\mathsf{B}^{\varnothing}_X\varphi$ has been first introduced by Grossi, Lorini, and Schwarzentruber (2015) in the special case when X is a set of *Boolean* variables. Baltag and van Benthem (2021) generalised their approach to arbitrary variables that are not necessarily Boolean. Because of the technical choices made by Baltag and van Benthem in the semantics of their system, their version of modality $\mathsf{B}^{\varnothing}_X\varphi$ is not an S5-modality. We proposed an alternative semantics under which the properties of $\mathsf{B}^{\varnothing}_X$ modality are exactly those captured in modal logic S5 (Jiang & Naumov, 2022a). We also introduced the term "data-informed knowledge".

Finally, observe that item 2 of the above definition states that the X-equivalence class of current world w is a subset of Y-equivalence class of w. Thus, knowing the values of all variables in dataset X would restrict the set of all possible worlds to those where the values of all variables in dataset Y are the same as in the current world. In other words, knowing the values of all variables in dataset X informs the knowledge of the values of all variables in dataset Y in *the current world* w. This is the definition of the *local* functional dependency between two datasets in the current world. One can also consider *global* functional dependency: "in *each world* of the model, knowing the values of all variables in dataset X informs the knowledge of the values of all variables in dataset Y". This dependency is expressible in our language by the formula $\mathsf{B}^{\varnothing}_{\varnothing}(X \rhd Y)$.

5 Axioms

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

1. Reflexivity: $X \rhd Y$, where $Y \subseteq X$,
2. Transitivity: $X \rhd Y \to (Y \rhd Z \to X \rhd Z)$,
3. Augmentation: $X \rhd Y \to (X \cup Z) \rhd (Y \cup Z)$,
4. Truth: $\mathsf{B}^{\varnothing}_X\varphi \to \varphi$,
5. Distributivity: $\mathsf{B}^T_X(\varphi \to \psi) \to (\mathsf{B}^T_X\varphi \to \mathsf{B}^T_X\psi)$,
6. Negative Introspection of Beliefs: $\neg\mathsf{B}^T_X\varphi \to \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$,
7. Trust: $\mathsf{B}^T_X(\mathsf{B}^T_Y\varphi \to \varphi)$,
8. Monotonicity: $X \rhd Y \to (\mathsf{B}^T_Y\varphi \to \mathsf{B}^T_X\varphi)$ and
   $\mathsf{B}^T_X\varphi \to \mathsf{B}^{T'}_X\varphi$, where $T \subseteq T'$,
9. Introspection of Dependency: $X \rhd Y \to \mathsf{B}^{\varnothing}_X(X \rhd Y)$.

The Reflexivity, the Transitivity, and the Augmentation axioms are the standard Armstrong's (1974) axioms of functional dependency.

To understand the meaning of the Truth and the Negative Introspection of Beliefs axioms, recall from Section 4 that $\mathsf{B}_X^{\varnothing}\varphi$ 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 when dataset T is trusted, 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.

Note that statement $\mathsf{B}_Y^T\varphi \to \varphi$, generally speaking, is not true. However, by item 5 of Definition 2, this statement is true in all worlds of the model in which dataset T is trustworthy. We capture this observation by the Trust axiom of our system. Note that there is no connection between datasets X and Y in this axiom. In particular, dataset X could be the empty set $\varnothing$. Informally, the axiom states that anyone trusting dataset T believes that any belief based on trust in T must be true.

The meaning of the two Monotonicity axioms is straightforward. Note that by item 2 of Definition 2, if statement $X \rhd Y$ is true in any world w, then it is also true in any world u such that $w \sim_X u$. We capture this in the Introspection of Dependency axiom.

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

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

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 $\vdash \mathsf{B}_X^T\varphi \to \mathsf{B}_X^{\varnothing}\mathsf{B}_X^T\varphi$.

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

$$\vdash \mathsf{B}_X^T\varphi \to \mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^T\varphi. \tag{7}$$

At the same time, the formula $\neg\mathsf{B}_X^T\varphi \to \mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^T\varphi$ is also an instance of the Negative Introspection axiom. Thus, $\vdash \neg\mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^T\varphi \to \mathsf{B}_X^T\varphi$ by the law of contrapositive in the propositional logic. Hence, by the Necessitation inference rule, $\vdash \mathsf{B}_X^{\varnothing}(\neg\mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^T\varphi \to \mathsf{B}_X^T\varphi)$. Thus, by the Distributivity axiom and the Modus Ponens inference rule, $\vdash \mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^{\varnothing}\neg\mathsf{B}_X^T\varphi \to \mathsf{B}_X^{\varnothing}\mathsf{B}_X^T\varphi$. The latter, together with statement (7), implies the statement of the lemma by propositional reasoning. ⊠

We show the soundness of our axioms in Section 6.

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

Lemma 3 (Lindenbaum) *Any consistent subset of Φ can be extended to a maximal consistent subset of Φ.*
PROOF. The standard proof of Lindenbaum's lemma (Mendelson, 2009, Proposition 2.14) applies here. ⊠

6 Soundness

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

The soundness of the Truth, the Distributivity, and the Monotonicity axioms is straightforward. Below we prove the soundness of the remaining axioms as separate lemmas.

Lemma 4 *If $Y \subseteq X$, then $w \Vdash X \rhd Y$.*
PROOF. Consider any world $u \in W$ such that $w \sim_X u$. By item 2 of Definition 2, it suffices to prove that $w \sim_Y u$, which is true by the assumption $w \sim_X u$ and the assumption $Y \subseteq X$ of the lemma. ⊠

Lemma 5 *If $w \Vdash X \rhd Y$ and $w \Vdash Y \rhd Z$, then $w \Vdash X \rhd Z$.*
PROOF. Consider any world $u \in W$ such that $w \sim_X u$. By item 2 of Definition 2, it suffices to prove that $w \sim_Z u$. The assumption $w \sim_X u$ implies that $w \sim_Y u$ by the assumption $w \Vdash X \rhd Y$ of the lemma and item 2 of Definition 2. Therefore, $w \sim_Z u$ by the assumption $w \Vdash Y \rhd Z$ of the lemma and item 2 of Definition 2. ⊠

Lemma 6 *If $w \Vdash X \rhd Y$, then $w \Vdash X \cup Z \rhd Y \cup Z$.*
PROOF. Consider any world $u \in W$ such that $w \sim_{X\cup Z} u$. By item 2 of Definition 2, it suffices to prove that $w \sim_{Y\cup Z} u$. Indeed, the assumption $w \sim_{X\cup Z} u$ implies that

$$w \sim_X u, \tag{8}$$
$$w \sim_Z u. \tag{9}$$

The assumption $w \Vdash X \rhd Y$ of the lemma implies $w \sim_Y u$ by item 2 of Definition 2 and statement (8). Then, $w \sim_{Y\cup Z} u$ due to statement (9). ⊠

Lemma 7 *If $w \nVdash \mathsf{B}^T_X\varphi$, then $w \Vdash \mathsf{B}^{\varnothing}_X\neg\mathsf{B}^T_X\varphi$.*
PROOF. By item 5 of Definition 2, the assumption $w \nVdash \mathsf{B}^T_X\varphi$ implies that there is a world $u \in W$ such that

$$w \sim_X u, \tag{10}$$
$$T \subseteq \mathcal{T}_u, \tag{11}$$

and

$$u \nVdash \varphi. \tag{12}$$

Consider any world $v \in W$ such that $w \sim_X v$. By item 5 of Definition 2, it suffices to show that $v \nVdash \mathsf{B}^T_X\varphi$. Assume the opposite. Then, $v \Vdash \mathsf{B}^T_X\varphi$. Note that statement (10) and the assumption $w \sim_X v$ imply that $v \sim_X u$ because $\sim_X$ is an equivalence relation. Therefore, $u \Vdash \varphi$ by item 5 of Definition 2 and statement (11), which contradicts statement (12). $\boxtimes$

Lemma 8 *$w \Vdash \mathsf{B}_X^T(\mathsf{B}_Y^T\varphi \to \varphi)$.*

PROOF. Consider any world $u \in W$ such that $w \sim_X u$ and $T \subseteq \mathcal{T}_u$. By item 5 of Definition 2, it suffices to show that $u \Vdash \mathsf{B}_Y^T\varphi \to \varphi$. Suppose that $u \Vdash \mathsf{B}_Y^T\varphi$. By item 4 of Definition 2, it is enough to prove that $u \Vdash \varphi$.

Note that $u \sim_Y u$ because relation $\sim_Y$ is reflexive. Also, $T \subseteq \mathcal{T}_u$ by the choice of world u. Then, the assumption $u \Vdash \mathsf{B}_Y^T\varphi$ implies that $u \Vdash \varphi$ by item 5 of Definition 2. $\boxtimes$

Lemma 9 *If $w \Vdash X \rhd Y$, then $w \Vdash \mathsf{B}_X^{\varnothing}(X \rhd Y)$.*

PROOF. Let $u \in W$ be any world such that $w \sim_X u$. By item 5 of Definition 2, it suffices to show that $u \Vdash X \rhd Y$. Consider any world $v \in W$ such that $u \sim_X v$. By item 2 of Definition 2, it suffices to prove that $u \sim_Y v$.

Indeed, the assumptions $w \sim_X u$ and $u \sim_X v$ imply that $w \sim_X v$ because relation $\sim_X$ is transitive. Hence, by item 2 of Definition 2 and the assumption $w \Vdash X \rhd Y$ of the lemma,

$$w \sim_Y v. \tag{13}$$

At the same time, the assumption $w \sim_X u$ implies $w \sim_Y u$ again by item 2 of Definition 2 and the assumption $w \Vdash X \rhd Y$ of the lemma. Therefore, $u \sim_Y v$ by statement (13) and symmetry and transitivity of the relation $\sim_Y$. $\boxtimes$

7 Completeness

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

7.1 Dataset Closure

An important idea used in our proof of completeness is "dataset closure". Informally, for each set of formulae F and each dataset X, by closure X_F^* we denote the set of all data variables about which set F can prove that they are informed by set X. This notion goes back to "saturated" sets in Armstrong's article on functional dependency (Armstrong, 1974, Section 6). Closures are used in Definition 5 of the next section to specify the labels of the edges of a tree.

Definition 3 *$X_F^* = \{x \in V \mid X \rhd x \in F\}$ for any datasets $X, T \subseteq V$ and any maximal consistent set of formulae $F \subseteq \Phi$.*

In other words, the closure X_F^* is the set of all data variables that, according to set F, are functionally determined by dataset X. Intuitively, such set must include variables from the dataset X itself. Next, we formally prove this.

Lemma 10 *$X \subseteq X_F^*$.*

PROOF. Consider any data variable $x \in X$. Thus, $\vdash X \rhd x$ by the Reflexivity axiom. Hence, $\vdash X \rhd x$ by the Necessitation inference rule. Then, $(X \rhd x) \in F$ because F is a maximal consistent set of formulae. Therefore, $x \in X_F^*$ by Definition 3. $\boxtimes$

Note that $(X \rhd x) \in F$ for each data variable $x \in X^*_F$ by Definition 3. The next lemma shows that all such variables x could be brought together on the right-hand-side of $\rhd$ expression.

Lemma 11 $F \vdash X \rhd X^*_F$.

PROOF. The set X^*_F is finite by Definition 3 and the assumption of the article that set V is finite. Let $X^*_F = \{x_1, \dots, x_n\}$. Note that $F \vdash X \rhd x_i$ for each $i \leq n$ by Definition 3. We prove by induction that $F \vdash X \rhd x_1, \dots, x_k$ for each integer k such that $0 \leq k \leq n$.

Base Case: $F \vdash X \rhd \varnothing$ by the Reflexivity axiom.

Induction Step: Suppose that $F \vdash X \rhd x_1, \dots, x_k$. Then, by the Augmentation axiom and the Modus Ponens inference rule,

$$F \vdash X \cup \{x_{k+1}\} \rhd x_1, \dots, x_k, x_{k+1}. \tag{14}$$

Recall that $F \vdash X \rhd x_{k+1}$. Hence, $F \vdash X \cup X \rhd X \cup \{x_{k+1}\}$ by the Augmentation axiom and the Modus Ponens inference rule. Then, $F \vdash X \rhd X \cup \{x_{k+1}\}$. Therefore,

$$F \vdash X \rhd x_1, \dots, x_k, x_{k+1}$$

by the Transitivity axiom, statement (14), and the Modus Ponens rule applied twice. ⊠

7.2 Canonical Model

As usual, at the core of the proof of 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 $F_0 \subseteq \Phi$.

Usually, possible worlds in modal logics are modelled using maximal consistent sets of formulae. In the case of epistemic logic S5, we say that two worlds are $\sim_a$-equivalent if they contain the same K_a-formulae. Unfortunately, this construction does not work for the distributed knowledge version of S5. Indeed, if two worlds are $\sim_a$-equivalent and $\sim_b$-equivalent, then they share K_a-formulae and K_b-formulae, but not necessarily K_{ab}-formulae. However, due to the semantics of distributed knowledge, any two worlds that are simultaneously $\sim_a$-equivalent and $\sim_b$-equivalent must share K_{ab}-formulae. In the case of distributed knowledge, this problem was solved in (Fagin, Halpern, & Vardi, 1992) by using a tree construction. The tree construction specifies a tree whose nodes are labelled with maximal consistent sets and whose edges are labelled with sets of agents. The construction guarantees that maximal consistent sets at any two adjacent nodes share K_C-formulae, where C is any subset of the label on the edge connecting the two nodes. Nodes represent possible worlds. Two nodes are $\sim_a$-indistinguishable if each edge along the simple path between the two nodes is labelled with a set containing agent a. The desired property about K_{ab}-formulae follows from the fact that the simple path between any two nodes in a tree is unique.

In this article, we adopt the tree construction to data-informed beliefs. The agents in (Fagin, Halpern, & Vardi, 1992) are replaced in our construction by data variables.

Also, in addition to maximal consistent sets of formulae, we label the nodes with sets of data variables that are trustworthy in those nodes (possible worlds). Just like in (Fagin, Halpern, & Vardi, 1992), we formally define nodes as sequences.

Definition 4 *Set W of worlds is the set of all sequences $T_0, F_0, X_1, T_1, F_1, \ldots, X_n, T_n, F_n$ 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 $w = T_0, F_0, \ldots, X_{n-1}, T_{n-1}, F_{n-1}$ and $u = T_0, F_0, \ldots, F_{n-1}, X_n, T_n, F_n$, we say that worlds w and u are *adjacent*. The adjacency relation defines a tree structure on set W. By $T(u)$ and $F(u)$ we mean sets T_n and F_n respectively.

Definition 5 *For any worlds $w, u \in W$ such that*

$$w = T_0, F_0, \ldots, X_{n-1}, T_{n-1}, F_{n-1}$$
$$u = T_0, F_0, \ldots, X_{n-1}, T_{n-1}, F_{n-1}, X_n, T_n, F_n,$$

*the **edge** between nodes w and u of this tree is **labelled** with all variables in dataset $(X_n)^*_{F_{n-1}}$ and that the **node** u is **labelled** with the pair T_n, F_n.*

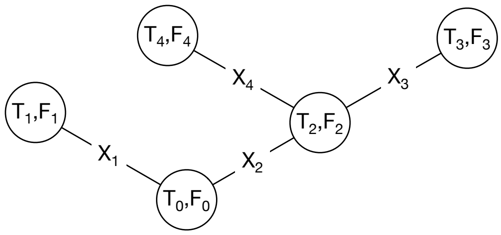

Fig. 3 Fragment of tree W.

It will be convenient to visualise tree W as shown in Figure 3. 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 the set $(X_4)^*_{F_2}$.

Definition 6 *For any worlds $w, u \in W$ and any data variable $x \in V$, let $w \sim_x u$ if every edge along the unique simple path between vertices w and u is labelled with variable x.*

Lemma 12 *Relation $\sim_x$ is an equivalence relation on set W for each data variable $x \in V$.*

Proof. The relation $\sim_x$ is reflexive because the simple path connecting any node to itself has no edges. Thus, each edge along this path is vaciously labelled with variable x. The relation is symmetric because the simple path from a node w to a node u contains the same edges as the simple path from node u to node w.

To prove that relation $\sim_x$ is transitive, suppose that all edges along the simple path from a node w to a node u are labelled with variable x and the same is true for the simple path from the node u to a node v. Note that the combination of these two paths forms a (not necessarily simple) path such that all edges along this path are labelled with variable x. By removing the loops from this path, one can obtain a *simple* path from the node w to node the v such that all edges along this simple path are labelled with variable x. ⊠

Definition 7 $\mathcal{T}_w = T(w)$.
Definition 8 $\pi(p) = \{w \in W \mid p \in F(w)\}$.

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

7.3 Properties of the Canonical Model

As common in modal logic, at the core of the proof of completeness is a truth lemma. In our case, this is Lemma 18. Lemma 14 and Lemma 15 are used in the induction step of the proof of the truth lemma. Lemma 13 below is an auxiliary result used in the proof of Lemma 14.

Lemma 13 *For any formula $\mathsf{B}^T_Y\varphi \in \Phi$ and any worlds*

$$w = T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}$$
$$u = T_0, F_0, \dots, X_{n-1}, T_{n-1}, F_{n-1}, X_n, T_n, F_n$$

if $Y \subseteq (X_n)^*_{F_{n-1}}$, then $\mathsf{B}^T_Y\varphi \in F(w)$ iff $\mathsf{B}^T_Y\varphi \in F(u)$.

PROOF. $(\Rightarrow)$: Suppose that $\mathsf{B}^T_Y\varphi \in F_{n-1}$. Thus, by Lemma 1 and the Modus Ponens inference rule

$$F_{n-1} \vdash \mathsf{B}^{\varnothing}_Y \mathsf{B}^T_Y\varphi. \tag{15}$$

Note that $F_{n-1} \vdash X_n \rhd (X_n)^*_{F_{n-1}}$ by Lemma 11. Also, $\vdash (X_n)^*_{F_{n-1}} \rhd Y$ by the assumption $Y \subseteq (X_n)^*_{F_{n-1}}$ of the lemma and the Reflexivity axiom. Hence, by the Transitivity axiom and the Modus Ponens rules applied twice, $F_{n-1} \vdash X_n \rhd Y$. Then, $F_{n-1} \vdash \mathsf{B}^{\varnothing}_{X_n}\mathsf{B}^T_Y\varphi$ by the Monotonicity axiom and statement (15). Thus, because F_{n-1} is a maximal consistent set, $\mathsf{B}^{\varnothing}_{X_n}\mathsf{B}^T_Y\varphi \in F_{n-1}$. Therefore, $\mathsf{B}^T_Y\varphi \in F_n$ by item 2(a) of Definition 4.

$(\Leftarrow)$: Suppose that $\mathsf{B}^T_Y\varphi \notin F_{n-1}$. 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. Then, again because set F_{n-1} is maximal, $\mathsf{B}^{\varnothing}_Y\neg\mathsf{B}^T_Y\varphi \in F_{n-1}$. Thus, $\neg\mathsf{B}^T_Y\varphi \in F_n$ by item 2(a) of Definition 4. Therefore, $\mathsf{B}^T_Y\varphi \notin F_n$ because set F_n is consistent. ⊠

Lemma 14 *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 6, 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 13 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 7. 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 4 and the Modus Ponens inference rule. Therefore, $\varphi \in F(u)$ because the set $F(u)$ is maximal. $\boxtimes$

Lemma 15 *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

$$G = \{\neg\varphi\} \cup \{\psi \mid \mathsf{B}^{\varnothing}_X\psi \in F(w)\} \cup \{\mathsf{B}^T_Y\chi \to \chi \mid Y \subseteq V, \chi \in \Phi\} \tag{16}$$

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

$$\mathsf{B}^{\varnothing}_X\psi_1, \dots, \mathsf{B}^{\varnothing}_X\psi_m \in F(w) \tag{17}$$

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,

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

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 (17). Then, $\mathsf{B}^T_X\varphi \in F(w)$ because the set $F(w)$ is maximal, which contradicts the assumption $\mathsf{B}^T_X\varphi \notin F(w)$ of the lemma. $\boxtimes$

Let G' be any maximal consistent extension of set G. Such an extension exists by Lemma 3. Suppose that $w = T_0, F_0, \dots, X_n, T_n, F_n$. Consider sequence

$$u = T_0, F_0, \dots, X_n, T_n, F_n, X, T, G'. \tag{18}$$

Note that $u \in W$ by Definition 4, equation (16), and the choice of set G' as an extension of set G. Also, observe that the edge between nodes w and u is labelled with each variable in set X by Definition 5, equation (18), and Lemma 10. Thus, $w \sim_X u$ by Definition 6. In addition, $T = T(u) = \mathcal{T}_u$ by equation (18) and Definition 7. Finally, $\neg\varphi \in G \subseteq G' = F(u)$ by equation (16), the choice of G' as an extension of G, and equation (18).Therefore, $\varphi \notin F(u)$ because the set $F(u)$ is consistent. This concludes the proof of the lemma. $\boxtimes$

Lemma 16 *For any world $w \in W$ and any formula $\neg(X \rhd Y) \in F(w)$, there is a world $w' \in W$ such that $w \sim_X w'$, and $w \nsim_Y w'$.*

PROOF. Let world w be sequence $T_0, F_0, X_1, \dots, X_n, T_n, F_n$. Consider sequence

$$w' = T_0, F_0, X_1, \dots, X_n, T_n, F_n, X, T_n, F_n.$$

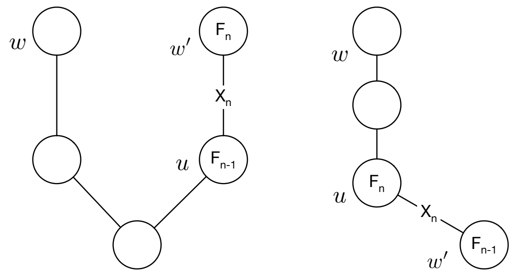

Fig. 4 Case I (left) and Case II (right).

To prove that $w' \in W$, consider any formula $\mathsf{B}_X^{\varnothing}\varphi \in F_n$. By Definition 4, it suffices to show that $\varphi \in F_n$. Indeed, the assumption $\mathsf{B}_X^{\varnothing}\varphi \in F_n$ implies $F_n \vdash \varphi$ by the Truth axiom and the Modus Ponens inference rule. Therefore, $\varphi \in F_n$ because set F_n is maximal.

To prove $w \sim_X w'$, note that $X \subseteq X^*_{F_n}$ by Lemma 10. Thus, the edge between vertices w and w' is labelled with each data variable in set X. Therefore, $w \sim_X w'$ by Definition 6.

Finally, we show that $w \nsim_Y w'$. By Definition 6, it suffices to prove that the simple path between vertices w and w' is *not* labelled by at least one variable from set Y. Then, by the choice of sequence w' and Definition 5, it suffices to show that $Y \nsubseteq X^*_{F_n}$. Suppose the opposite. Thus, $\vdash X^*_{F_n} \rhd Y$ by the Reflexivity axiom. Note that $F_n \vdash X \rhd X^*_{F_n}$ by Lemma 11. Hence, $F_n \vdash X \rhd Y$ by the Transitivity axiom and the Modus Ponens inference rule applied twice. Thus, $\neg(X \rhd Y) \notin F_n = F(w)$ because set F_n is consistent, which contradicts the assumption $\neg(X \rhd Y) \in F(w)$ of the lemma. ⊠

Lemma 17 *For any worlds $w, w' \in W$, if $X \rhd Y \in F(w)$, and $w \sim_X w'$, then $w \sim_Y w'$.*

Proof. We prove the lemma by induction on the length of the simple path between vertices w and w'. If $w = w'$, then, vacuously, each edge along the simple path between vertices w and w' is labelled with each data variable. Hence, $w \sim_Y w'$ by Definition 6.

Suppose that $w \neq w'$. Consider the unique simple path between vertices w and w'. By the assumption $w \sim_X w'$ of the lemma and Definition 6, each edge along this path is labelled with each data variable in set X. Because $w \neq w'$, there must exist a vertex $u \in W$ on the unique simple path between w and w' such that vertices u and w' are adjacent. Note that the unique simple path between vertices w' and u is a part of the unique simple path between vertices w and w'. Thus, each edge along the simple path between vertices w and u is labelled with each data variable in set X. Hence, by Definition 6,

$$w \sim_X u. \tag{19}$$

Claim 2 *The edge between vertices u and w' is labelled with each data variable in set Y.*

Proof of Claim. We consider the following two cases separately, see Figure 4:

Case I: $u = T_0, F_0, X_1, T_1, F_1, \dots, T_{n-1}, F_{n-1}$ and $w' = T_0, F_0, X_1, T_1, F_1, \dots, X_n, T_n, F_n$. Consider any data variable $y \in Y$. By Definition 5, it suffices to show that $y \in (X_n)^*_{F(u)}$. Note that $X \rhd Y \in F(w)$ by the assumption of the lemma. Thus, $F(w) \vdash \mathsf{B}^{\varnothing}_X(X \rhd Y)$ by the Introspection of Dependency axiom and the Modus Ponens inference rule. Hence, $\mathsf{B}^{\varnothing}_X(X \rhd Y) \in F(w)$ because set $F(w)$ is maximal. Then, $X \rhd Y \in F(u)$ by Lemma 14 and statement (19). Note that $\vdash Y \rhd \{y\}$ by the Reflexivity axiom. Hence, by the Transitivity axiom and the Modus Ponens inference rule applied twice,

$$F(u) \vdash X \rhd y. \tag{20}$$

Recall that u is a vertex on the simple path connecting vertices w and w' and all edges along this path are labelled with all variables in dataset X. Hence, $X \subseteq (X_n)^*_{F(u)}$ by Definition 5. Then, by the Reflexivity axiom and the Modus Ponens inference rule. $\vdash (X_n)^*_{hd(u)} \rhd X$. Thus, $\vdash (X_n)^*_{F(u)} \rhd y$ by the Transitivity axiom and statement (20). Hence, $\vdash X_n \rhd y$ by the Transitivity axiom and Lemma 11. Therefore, $y \in (X_n)^*_{F(u)}$ by Definition 3.

Case II: $w' = T_0, F_0, X_1, \dots, F_{n-1}$ and $u = T_0, F_0, X_1, \dots, X_n, T_n, F_n$. This case is similar to the previous one, except that it uses the set $F(w')$ instead of the set $F(u)$ everywhere in the proof. ⊠

To finish the proof of the lemma, note that the simple path between vertices w and u is shorter than the simple path between vertices w and w'. Hence, $w \sim_Y u$, by the induction hypothesis. Also, $u \sim_Y w'$ by Claim 2 and Definition 6. Therefore, $w \sim_Y w'$ because relation $\sim_Y$ is transitive. ⊠

Lemma 18 *$w \Vdash \varphi$ 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 an atomic proposition, then the statement of the lemma follows from Definition 8 and item 1 of Definition 2.

Let formula φ have the form $X \rhd Y$.

$(\Rightarrow)$: Assume $X \rhd Y \notin F(w)$. Thus, $\neg(X \rhd Y) \in F(w)$ because set $F(w)$ is maximal. Hence, by Lemma 16, there is a world $w' \in W$ such that $w \sim_X w'$, and $w \nsim_Y w'$. Therefore, $w \nVdash X \rhd Y$ by item 2 of Definition 2.

$(\Leftarrow)$: Assume that $X \rhd Y \in F(w)$. Then, by Lemma 17, for any world $w' \in W$, if $w \sim_X w'$, then $w \sim_Y w'$. Therefore, $w \Vdash X \rhd Y$ by item 2 of Definition 2.

If formula φ is a negation or an implication, then the statement of the lemma follows from the induction hypothesis, items 3 and 4 of Definition 2 and the maximality and consistency of the set $F(w)$ in the standard way.

Finally, 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 15, 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 5 of Definition 2.

$(\Leftarrow)$: Consider any world u such that $w \sim_X u$ and $T \subseteq \mathcal{T}_u$. By item 5 of Definition 2, it suffices to show that $u \Vdash \psi$. By Lemma 14, 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. ⊠

7.4 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 4, 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.

Finally, note that $\varphi \notin F_0$ because set F_0 is consistent and $\neg\varphi \in F_0$. Then, by Lemma 18 and because $F \subseteq F_0$, it follows that $\varnothing, F_0 \Vdash f$ for each formula $f \in F$ and $\varnothing, F_0 \nVdash \varphi$. ⊠

8 Future Work

The syntax and the semantics of our formal modelling of trust and beliefs are relatively simple. This creates an opportunity for extensions of the proposed logical system. In (Jiang & Naumov, 2024) we propose one such extension with doxastic strategies. Below we list several other possible directions.

8.1 Doxastic Functional Dependency

One such possible extension is a generalisation of functional dependency expression $X \rhd Y$ to *doxastic* functional dependency expression $X \rhd^T Y$. Recall that, informally, $X \rhd Y$ means that "knowing the values of all variables from dataset X is enough to determine the values of all variables from dataset Y". The expression $X \rhd^T Y$, informally, means that "if dataset T is trusted, then knowing the values of all variables from dataset X is enough to determine the values of all variables from dataset Y"[1]. Formally, we propose the following definition of this modality:

$w \Vdash X \rhd^T Y$ when for any worlds $u, u' \in W$ if $w \sim_X u$, $w \sim_X u'$, $T \subseteq \mathcal{T}_u$, and $T \subseteq \mathcal{T}_{u'}$, then $u \sim_Y u'$.

Observe that $X \rhd^{\varnothing} Y$ is equivalent to the original functional dependency expression $X \rhd Y$. Most axioms of our logical system can be generalised from functional dependency to doxastic functional dependency:

1. Reflexivity: $X \rhd^T Y$, where $Y \subseteq X$,
2. Transitivity: $X \rhd^T Y \to (Y \rhd^T Z \to X \rhd^T Z)$,
3. Augmentation: $X \rhd^T Y \to (X \cup Z) \rhd^T (Y \cup Z)$,
4. Introspection of Dependency: $X \rhd^T Y \to \mathsf{B}^{\varnothing}_X (X \rhd^T Y)$.

The only exception is the Monotonicity axiom. Its most natural generalisation, $X \rhd^T Y \to (\mathsf{B}^T_Y \varphi \to \mathsf{B}^T_X \varphi)$, is not universally true. Indeed, consider the trustworthiness

[1] Even more informally, "the one who knows X and trusts T, beliefs that she knows Y".

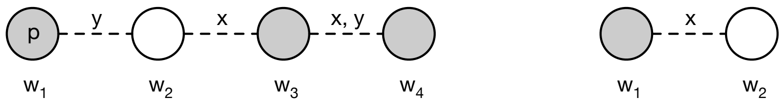

Fig. 5 Two trustworthiness models. The worlds in which data variable t is trustworthy are shaded grey.

model depicted in the left of Figure 5. Observe that $w_2 \Vdash x \rhd^t y$ and $w_2 \Vdash \mathsf{B}^t_y p$, but $w_2 \nVdash \mathsf{B}^t_x p$. The same is also true about the right trustworthiness model from Figure 5, but perhaps it provides less insight due to being too succinct.

A complete axiomatisation of the interplay between trust-based belief modality B^T_X and doxastic functional dependency expression $X \rhd^T Y$ remains an open problem.

8.2 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 a public announcement of the values of datasets rather than of true formulae. Such modality has been first introduced in (van Eijck, Gattinger, & Wang, 2017) under the 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". As we observed in (Deuser, Jiang, Naumov, & Zhang, 2024), to formally define the semantics of this modality, it is easiest to modify the 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 the 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 atomic proposition 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 9 *For any trustworthiness model* $(W, \{\sim_x\}_{x\in V}, \{\mathcal{T}_w\}_{w\in W}, \pi)$*, any world* $w \in W$*, any dataset* $U \subseteq V$*, and any formula* $\varphi \in \Phi$*, the satisfaction relation* $w, U \Vdash \varphi$ *is defined as follows:*

1. $w, U \Vdash p$ *if* $(w, U) \in \pi(p)$,
2. $w, U \Vdash X \rhd Y$ *when for each* $v \in W$*, if* $w \sim_{X\cup U} v$*, then* $w \sim_Y v$,
3. $w, U \Vdash \neg\varphi$ *if* $w, U \nVdash \varphi$,
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* $w \sim_{X\cup U} v$ *and* $T \subseteq \mathcal{T}_v$,
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, van der Hoek, & Kooi, 2007, Chapter 4). Similarly, in the logic of public inspections, the "true" values of the variables are announced (van Eijck, Gattinger, & Wang, 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 an 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 additional axioms capture the interplay between data-informed beliefs, functional dependency, and public announcements:

10. Distributivity: $[X](\varphi \to \psi) \to ([X]\varphi \to [X]\psi)$,
11. Combination: $[X][Y]\varphi \leftrightarrow [X \cup Y]\varphi$,
12. Duality: $\neg[X]\varphi \leftrightarrow [X]\neg\varphi$,
13. Perfect Recall: $\mathsf{B}^T_X[Y]\varphi \to [Y]\mathsf{B}^T_X\varphi$,
14. Public Knowledge: $[X](\mathsf{B}^T_{X\cup Y}\varphi \to \mathsf{B}^T_Y\varphi)$,
15. Prior Belief: $[X]\mathsf{B}^T_Y\varphi \to \mathsf{B}^T_{X\cup Y}[X]\varphi$,
16. Partial Announcement: $(X \cup Y) \rhd Z \leftrightarrow [X](Y \rhd Z)$,
17. Empty Announcement: $\varphi \leftrightarrow [\varnothing]\varphi$.

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.

8.3 De Re Trust

Imagine that you are looking at a broken thermometer that shows 30°C. You know that the thermometer is broken, but you observe that your mobile phone, which you trust, also shows a temperature of 30°C. As a result, you don't trust the thermometer reading as a data variable, but you trust *the value* of this variable. This trust in the value as opposed to the data variable is very different from the trust we formalised in the current article. Perhaps the trust in the value should be called *de re* trust while the trust in the data variable (no matter what its value is) could be referred to as *de dicto* trust. Note that de dicto trust, at least the way we presented it in the current work, leads to the formation of beliefs. On the other hand, it appears that de re trust is formed by the beliefs. It is not clear how the trustworthiness models proposed in the current article can be used to model de re trust. Perhaps a different type of semantics can be developed in the future to study de re trust and its interplay with de dicto trust.

9 Conclusion

In this article, we proposed to extend the distinction between trust and trustworthiness from agents to information expressed through (data) variables. Building on this distinction, we defined beliefs as a combination of subjective mindset (trusted variables) and objective information (values of variables) available to an agent. For this setting, we gave a sound and complete logical system describing the interplay of beliefs and functional dependencies between datasets. We also discussed possible extensions of this system with a doxastic function dependency expression and a public announcement modality, as well as a related notion of de re trust.

Acknowledgements. Junli Jiang acknowledges the support of the Innovation Research 2035 Pilot Plan of Southwest University (No. SWUPilotPlan018), the Fundamental Research Funds for the Central Universities (SWU2409115), and the China Scholarship Council (No. 202306990071).

Declarations

The authors wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

Armstrong, W.W. (1974). Dependency structures of data base relationships. *Information processing 74 (Proc. IFIP Congress, Stockholm, 1974)* (pp. 580–583). Amsterdam: North-Holland.

Baltag, A., & Smets, S. (2006). Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, *165*, 5 - 21, (Proceedings of the 13th Workshop on Logic, Language, Information and Computation (WoLLIC 2006))

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

Baltag, A., & van Benthem, J. (2021). A simple logic of functional dependence. *Journal of Philosophical Logic*, *50*, 1–67,

Ben-Ner, A., & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*(1), 64–79,

Board, O. (2004). Dynamic interactive epistemology. *Games and Economic Behavior*, *49*(1), 49–80,

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

Castelfranchi, C., & Falcone, R. (1998). Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. *Proceedings international* *conference on multi agent systems (cat. no. 98ex160)* (pp. 72–79).

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, *73*(4), 959–985,

Deuser, K., Jiang, J., Naumov, P., Zhang, W. (2024). A dynamic logic of data-informed knowledge. *Journal of Philosophical Logic*, *53*(2), 521-557, https://doi.org/10.1007/s10992-023-09740-5

Fagin, R., Halpern, J.Y., Vardi, M.Y. (1992). What can machines know? on the properties of knowledge in distributed systems. *Journal of the ACM (JACM)*, *39*(2), 328–376,

Friedman, N., & Halpern, J.Y. (1997). Modeling belief in dynamic systems, part I: foundations. *Artificial intelligence*, *95*(2), 257–316,

Friedman, N., & Halpern, J.Y. (1999). Modeling belief in dynamic systems, part II: Revision and update. *Journal of Artificial Intelligence Research*, *10*, 117–167,

Girlando, M., Lellmann, B., Olivetti, N. (2019). Nested sequents for the logic of conditional belief. *European conference on logics in artificial intelligence* (pp. 709–725).

Girlando, M., Negri, S., Olivetti, N., Risch, V. (2016). The logic of conditional beliefs: Neighbourhood semantics and sequent calculus. *Advances in modal logic.* College Publications.

Grossi, D., Lorini, E., Schwarzentruber, F. (2015). The ceteris paribus structure of logics of game forms. *Journal of Artificial Intelligence Research*, *53*, 91–126,

Hardin, R. (2002). *Trust and trustworthiness.* Russell Sage Foundation.

Herzig, A., Lorini, E., Hübner, J.F., Vercouter, L. (2010). A logic of trust and reputation. *Logic Journal of the IGPL*, *18*(1), 214–244,

Jiang, J., & Naumov, P. (2022a). Data-informed knowledge and strategies. *Artificial Intelligence*, *309*, 103727, https://doi.org/https://doi.org/10.1016/j.artint.2022.103727

Jiang, J., & Naumov, P. (2022b). In data we trust: The logic of trust-based beliefs. *the 31st international joint conference on artificial intelligence (IJCAI-22).*

Jiang, J., & Naumov, P. (2024). The logic of doxastic strategies. *Proceedings of 38th aaai conference on artificial intelligence (AAAI-24)* (Vol. 38, p. 19876-19883).

Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. *Annual review of political science*, *3*(1), 475–507,

Lewis, D.K. (1973). *Counterfactuals.* Harvard University Press.

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

McLeod, C. (2023). Trust. E.N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2023/entries/trust/.

Mendelson, E. (2009). *Introduction to mathematical logic.* Boca Raton, Florida: CRC press.

Perrotin, E., Galimullin, R., Canu, Q., Alechina, N. (2019). Public group announcements and trust in doxastic logic. *International workshop on logic, rationality and interaction* (pp. 199–213). Springer.

Posten, A.-C., & Mussweiler, T. (2019). Egocentric foundations of trust. *Journal of Experimental Social Psychology*, *84*, 103820,

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

Tagliaferri, M., & Aldini, A. (2019). A trust logic for the varieties of trust. *International conference on software engineering and formal methods* (pp. 119–136).

van Ditmarsch, H., van der Hoek, W., Kooi, B. (2007). *Dynamic epistemic logic.* Berlin, Germany: Springer.

van Eijck, J., Gattinger, M., Wang, Y. (2017). Knowing values and public inspection. *Indian conference on logic and its applications* (pp. 77–90).

van Eijck, J., & Li, K. (2017). Conditional belief, knowledge and probability. *Proceedings of the 16th conference on theoretical aspects of rationality and knowledge* (p. 188-206).

A Proof of Lemma 2

Before proving Lemma 2, we show the following auxiliary result.

Lemma 19 (deduction) *If $F, \varphi \vdash \psi$, then $F \vdash \varphi \to \psi$.*

Proof. Suppose that sequence $\psi_1, \dots, \psi_n$ is a proof from set $F \cup \{\varphi\}$ and the theorems of our logical system that uses the Modus Ponens inference rule only. In other words, for each $k \leq n$, either

1. $\vdash \psi_k$, or
2. $\psi_k \in F$, or
3. ψ_k is equal to φ, or
4. there are $i, j < k$ such that formula ψ_j is equal to $\psi_i \to \psi_k$.

It suffices to show that $F \vdash \varphi \to \psi_k$ for each $k \leq n$. We prove this by induction on k through considering the four cases above separately.

Case 1: $\vdash \psi_k$. Note that $\psi_k \to (\varphi \to \psi_k)$ is a propositional tautology, and thus, is an axiom of our logical system. Hence, $\vdash \varphi \to \psi_k$ by the Modus Ponens inference rule. Therefore, $F \vdash \varphi \to \psi_k$.

Case 2: $\psi_k \in F$. Note again that $\psi_k \to (\varphi \to \psi_k)$ is a propositional tautology, and thus, is an axiom of our logical system. Therefore, by the Modus Ponens inference rule, $F \vdash \varphi \to \psi_k$.

Case 3: formula ψ_k is equal to φ. Thus, $\varphi \to \psi_k$ is a propositional tautology. Therefore, $F \vdash \varphi \to \psi_k$.

Case 4: formula ψ_j is equal to $\psi_i \to \psi_k$ for some $i, j < k$. Thus, by the induction hypothesis, $F \vdash \varphi \to \psi_i$ and $F \vdash \varphi \to (\psi_i \to \psi_k)$. Note that formula $(\varphi \to \psi_i) \to ((\varphi \to (\psi_i \to \psi_k)) \to (\varphi \to \psi_k))$ is a propositional tautology. Therefore, $F \vdash \varphi \to \psi_k$ by applying the Modus Ponens inference rule twice. ⊠

Lemma 2. If $\varphi_1, .., \varphi_n \vdash \psi$, then $\mathsf{B}^Y_X \varphi_1, .., \mathsf{B}^Y_X \varphi_n \vdash \mathsf{B}^Y_X \psi$. Proof. By Lemma 19 applied n times, the assumption $\varphi_1, \dots, \varphi_n \vdash \psi$ implies that

$$\vdash \varphi_1 \to (\varphi_2 \to \dots (\varphi_n \to \psi) \dots).$$

Thus, by the Necessitation inference rule,

$$\vdash \mathsf{B}^T_X (\varphi_1 \to (\varphi_2 \to \dots (\varphi_n \to \psi) \dots)).$$

Hence, by the Distributivity axiom and the Modus Ponens rule,

$$\vdash \mathsf{B}^T_X \varphi_1 \to \mathsf{B}^T_X (\varphi_2 \to \dots (\varphi_n \to \psi) \dots).$$

Then, again by the Modus Ponens rule,

$$\mathsf{B}^T_X \varphi_1 \vdash \mathsf{B}^T_X (\varphi_2 \to \dots (\varphi_n \to \psi) \dots).$$

Therefore, $\mathsf{B}_X^T\varphi_1, \ldots, \mathsf{B}_X^T\varphi_n \vdash \mathsf{B}_X^T\psi$ by applying the previous steps $(n-1)$ more times. ⊠