Studies in Logic, Vol. 17, No. 3 (2024): 24–50
PII: 1674-3202(2024)-03-0024-27

Understanding and Explanation: An Epistemic-Logical Perspective*

Yu Wei

Abstract. Epistemic logicians barely pay any attention to the notion of *understanding*, which stands in sharp contrast to the current situation in philosophy of science and epistemology. This paper proposes an epistemic-logical-style framework for understanding. Since *explanations* aid understanding, our models incorporate varying degrees of explanations, among which a partial order is established. Inspired by philosophical discussions, we syntactically include a spectrum of understanding modalities, ranging from minimal to everyday, demanding, and ideal understanding. A sound and complete axiomatization is provided, followed by discussions on its application to multi-agent scenarios, such as making comparative statements of understanding among different agents and exploring meta-understanding between them.

1 Introduction

1.1 Understanding

Epistemic logic was discussed already by medieval logicians, who not only considered the usual epistemic modalities such as knowing and believing, but also doubting and understanding. ([6]) Among these, the modality of understanding is thought to be a particularly interesting one.[1] However, understanding became a largely forgotten notion in modern epistemic logic research that mainly concerns reasoning patterns of knowledge and belief.

Nevertheless, recent epistemology and philosophy of science have witnessed a resurgence of interest in the nature of understanding. Some philosophers believe

Received 2024-04-12

Yu Wei Department of Philosophy, East China Normal University
ywei@philo.ecnu.edu.cn

*The author would like to thank the two anonymous reviewers for their helpful comments on the earlier version. Many thanks to Qiang Wang for her inspired comments and careful inspection. Last but not least, the support from Shanghai Pujiang Program (No. 22PJC034) and Fundamental Research Funds for the Central Universities (No. 2022ECNU-YYJ038) is acknowledged.

[1] Interested readers are referred to [5, 6] for further discussions of Medieval epistemic logic. Concerning *understanding*, for example, the Middle Ages introduced many non-standard modalities into epistemic logic, among which "understanding" was a natural and common one. Besides, understanding is thought to be a prerequisite for knowledge and has interaction relations with many other epistemic modalities; and based on views from the philosophy of mind in the Middle Ages, the KK-like principle for understanding (i.e., one understands implies that she understands that she understands) is proposed. We will revisit some of them in the Sect. 3 below.

that understanding promises to be a lively topic throughout the twenty-first century.[2] While different uses of "understanding" seem to mean many different things, there are three main types of understanding ([10]):

- Propositional understanding or understanding-that: "S understands that ...."
- Atomistic understanding or understanding-wh: "S understands why/how/where/what ...."
- Objectual understanding or holistic understanding: "S understands X."

Among all these types, much of the recent philosophical discussion focuses on understanding-wh. Philosophers like Gordon ([10]) and Baumberger ([3]) argue that genuine instances of propositional understanding are quite rare, and for those rare instances of epistemologically relevant usage, many are just synonymous with propositional knowledge, others are actually identical to understanding-wh or objectual understanding. As for objectual understanding, it is roughly viewed as the understanding one has of a subject matter, typically expressed through a noun phrase (e.g., "Alice understands quantum mechanics"). Philosophers like Khalifa ([14]) contend that the notion of understanding-wh already captures anything philosophically important about objectual understanding.

The goal of this paper is to investigate epistemic-logically the typical understanding-wh termed *understanding why*, which is also referred to as a narrow conception of understanding in [17]. In the view of philosophers such as Pritchard (e.g., [20]), the usual use of "understanding" is "understanding why," as in "I understand why the house caught fire" or "Alice understands why Bob did this," etc. Understanding why is a paradigmatic expression of understanding. For the sake of simplicity, the notion of understanding will be treated as understanding why in the following.

Understanding why is widely named as "*explanatory understanding*"[3], which indicates the close relations between understanding and the notion of explanation. The author of [30] argues that explanations are precisely those sorts of things that bring about understanding. Alternatively, consider the slogan by Strevens in [27]: No understanding without explanation. Discussions about explanation and understanding will guide us in defining the language and semantics of our framework.

[2] See, for example [4, 13] and the bibliographies therein.

[3] For example, see [4, 14] and the bibliographies therein. It is also noted that we should not treat the "why" in "understanding why" as an implicit restriction: some kinds of understanding why might be more idiomatically expressed as *understanding how* ([14]). For example, it might be more natural to talk about understanding how the dinosaurs went extinct rather than why they went extinct, although there is no significant difference between the two. In either case, what is required is a correct explanation of the extinction.

1.2 Explanation

Serious why-questions demand explanations. Why does a piece of iron rust? The explanation for this phenomenon, for instance, is that iron and oxygen undergo a redox reaction when they come into contact in the presence of H_2O (whether in liquid form or as moisture in the air). Why is the sky blue? It is due to the way sunlight interacts with our atmosphere. The capacity to construct explanations is widely recognized as a fundamental feature of scientific theorization. It is commonly held that to provide an explanation is to respond to a why-question. Engaging in this process is thought to enhance our understanding of the world.

There are numerous contemporary philosophical studies on *explanation*. [4]As usual, we use the Latin words *explanandum* and *explanans* to cover what is being explained and whatever doing the explaining respectively. If we ask "why X?" then X is the explanandum. If we answer "because Y," then Y is the explanans. For instance, given an explanation E, it could be represented as $Y \Rightarrow X$, where X is the explanandum (e.g. the phenomenon of a piece of iron rusting), Y is the explanans (e.g. a redox reaction), and $\Rightarrow$ symbolizes the explanatory relations between Y and X. A theoretical account of explanation would specify the nature of $\Rightarrow$ and Y, X.

Accordingly, the distinctions among various theories of explanation (so-called models of explanation in the literature) primarily concern the acceptable types of explananda and explanantia and the nature of their relations. For example, in Hempel's deductive-nomological (DN) model of explanation ([12]), considered one of the foundations of modern discussions on explanation, the explanans and explanandum consist of specific statement sets, with the explanatory relation being logical entailment; and in the inductive-statistical (IS) model ([11]), inductive support is emphasized as the main explanatory relation. In Salmon's causal-mechanical (CM) model ([23]), the explanandum, which is an event, is explained by demonstrating how it fits within the causal nexus referred to by the explanans, so the explanatory relation is causal.

A full account of explanation is beyond the scope of this paper. Instead, we consider explanation at an abstract level and focus on the dichotomy of elements (*explanandum* and *explanans*) and their relations, regardless of the nature of the components. There is already some formal work on explanation along this line:

- Frameworks for explanations: It is argued that an explanation is a certain type of argument, thus the abstract argumentation framework introduced in [8] is applied to model explanation. The explanatory argumentation framework in

[4] Although explanations are commonly referred to as "scientific explanations" (especially in the field of philosophy of science), it is widely accepted that explanations given in daily life share significant similarities with explanations given in the sciences ([30, 31, 32]). Scientific explanations are typically more precise and rigorous than our explanations in everyday, non-scientific contexts, but the distinction between these two types of explanations is largely a matter of degree, rather than a fundamental difference in nature. That is, everyday explanations is continuous with scientific explanations. Thus explanation is thought to be a unified notion in this paper.

[25] is a tuple $\langle \mathcal{A}, \mathcal{X}, \rightarrow, \dashrightarrow, \sim \rangle$, where $\langle \mathcal{A}, \rightarrow \rangle$ is the argumentation system consisting of the set of arguments (explanations) $\mathcal{A}$ and the attack relation $\rightarrow \subseteq \mathcal{A} \times \mathcal{A}$. $\mathcal{X}$ is the set of explananda, $\dashrightarrow \subseteq (\mathcal{A} \times \mathcal{X}) \cup (\mathcal{A} \times \mathcal{A})$ is the explanatory relation holding between an argument and an explanandum, and between two arguments, $\sim \subseteq \mathcal{A} \times \mathcal{A}$ is the incompatibility relation between arguments. [5] Similarly, the abstract explanation framework in [24] is a tuple $\langle P, K, E \rangle$ where P is a set of explananda and explanantia, K is a set of criteria imposed on explanations, and $E \subseteq K \rightarrow (P \times P)$ is the explanatory-relation function from K to binary relations on P. [6]

- Frameworks for explanations and epistemic notions: Xu, Wang and Studer [33] take the ideas similar to justification logic together with the standard epistemic logic to capture agent i knowing why p (in formula: $\mathrm{Ky}_i p$), which is taken as agent i knowing an explanation of p semantically. The explanatory relation is characterized by $t : p$, which is a formula from justification logic originally stating that "t is a justification of p". Thus the semantical analysis of "knowing why p" is $\exists t \mathrm{K}_i(t : p)$.
  Philosophically inspired by [16] and technically by [33], Wei [29] analyzes the expression understanding why φ (in formula: $\mathrm{Uy}\varphi$) as $\exists t_1 \exists t_2(\mathrm{K}(t_2 : (t_1 : \varphi)),$ where $t_1 : \varphi$ means t_1 is an explanation for φ and $t_2 : (t_1 : \varphi)$ expresses t_2 is a *higher-order* explanation for "t_1 is an explanation for φ". This framework thereby embodies the philosophical idea that understanding why requires at least two explanations at different levels, as opposed to merely knowing why.

We will borrow some ideas from the work in frameworks for explanations to expand and enhance the logical framework on understanding why by [29] .

1.3 Basic ideas of the syntax and semantics

The relation between understanding and knowing has been a prominent theme in the search for a satisfactory account of understanding, as noted by [22]. Our point of departure is the widely held assumption that understanding why requires "more" than knowing why. As a case in point, Pritchard [19] introduces a scenario in which a child knows, via testimony, that a house burned down due to faulty wiring. The child then knows why the house burned down. She could answer a corresponding why-question since she accepts the information and is ready to repeat it to her friends. However, she does not understand why the house burned down because she has no conception of how the faulty wiring caused the fire. Thus, if the child were asked why

[5] In the explanatory argumentation framework, $a \dashrightarrow x$ is designated as "a explains x", where $a \in \mathcal{A}$, $x \in \mathcal{A} \cup \mathcal{X}$. The explanatory relation between arguments themselves allows for explanations to be deepened. See [25] for more details.

[6] In the the abstract explanation framework, xE_iy is read as "x explains y according to the criterion i" or "x is an i-explanans of y", where $x, y \in P$ and $i \in K$. See [24] for more details.

the introduction of faulty wiring caused the fire, she would be unable to respond. The idea is that one having an understanding of why could also answer a kind of "vertical" follow-up why-question (see [16]), which seeks a higher-order explanation, namely, why a particular explanation (e.g., faulty wiring caused the fire) is the explanation.

Based on the logic of knowing why in [33], the author of [29] define a new "packed" modality Uy in the language, and conceal the information of high-level explanations by virtue of the existence quantifications of $\exists t_1 \exists t_2(\mathrm{K}(t_2 : (t_1 : \varphi)))$ in the semantics. An understanding why model $\mathfrak{M}$ is defined as a tuple $(W, E, R, \mathcal{E}, V)$ where (W, R, V) is an epistemic model, E is a non-empty set of explanations, and $\mathcal{E}$ is an admissible explanation function specifying the set of worlds for both first-level explanations ($\mathcal{E}(t, \varphi)$) and second-level explanations ($\mathcal{E}(t_2, \langle t_1, \psi\rangle)$). If $w \in \mathcal{E}(t, \varphi)$ then t is a (first-level) explanation for φ in the world w, and if $v \in \mathcal{E}(t_2, \langle t_1, \psi\rangle)$ then v is a world where t_2 is a second-level explanation for that t_1 is an explanation of ψ.

The truth conditions for the standard operators are routine, and with:

- Kyφ holds at $\mathfrak{M}, w$ iff (1) there exists $t \in E$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(t, \varphi)$; and (2) for all $v \in W$ with wRv, φ holds at v.
- Uyφ holds at $\mathfrak{M}, w$ iff (1) there exist $t_1, t_2 \in E$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(t_2, \langle t_1, \varphi\rangle)$; and (2) for all $v \in W$ with wRv, $\mathfrak{M}, v \vDash \varphi$.

Therefore, while the formula Kyφ is roughly $\exists t\mathrm{K}(t : \varphi) \wedge \mathrm{K}\varphi$, the structure of the Uyφ can be displayed as $\exists t_1 \exists t_2 \mathrm{K}(t_2 : (t_1 : \varphi)) \wedge \mathrm{K}\varphi$. The framework limits itself to two levels of explanation because it investigates "understanding why" by identifying what distinguishes it from "knowing why". Based on insights from philosophical viewpoints, two levels of explanations suffice.

Nevertheless, in the framework presented in [29], the notion of understanding why appears to be an all-or-nothing matter. As noted in [26], much of the recent literature argues that understanding why is a distinct cognitive state from knowing why for the reason that, unlike knowledge, understanding comes in degrees. It is possible to more or less understand something.[7] Immediately, two questions arise:

1. What does it mean to understand better?
2. How do we navigate varying degrees of understanding?

As for the first question, according to the philosophical views, the distinctions among degrees of understanding can be drawn from explanations of the same phenomena, with some considered deeper or better than others ([27]). For example, Railton ([21]) suggests that an explanation which traces an event's causal history further back

[7] Borrowing an example from [14], consider Alice, a leading atmospheric physicist. Presumably, her understanding of why the sky is blue is extensive, encompassing a range of causal factors, deep theoretical principles, experimental results, and methodologies. Contrast this with Alice's freshman student, Bob, who is credited with understanding why the sky is blue even though he knows only a small fraction of what Alice does. In the story, Alice's understanding is considered better than Bob's. We will return to this example in Sect. 5.

is deeper. Similarly, Thagard ([28]) argues that deepening occurs in a causal explanation when it provides an underlying causal basis for the causal hypothesis. Recall the explanatory argumentation framework by [25], the explanatory relation holding between an argument and an explanandum, as well as between two arguments themselves. The explanatory relation between arguments allows for explanations to be deepened. In the framework, $c \dashrightarrow b \dashrightarrow a \dashrightarrow e$ and $b \dashrightarrow a \dashrightarrow e$ can be viewed as two explanations (where $a, b, c \in \mathcal{A}$ and $e \in \mathcal{X}$), and the former is deeper than the latter. The argument c can be used to explain one of the premises of argument b or the link between the premises and the conclusion of b.

It suggests that we should release the restriction to two levels of explanations in the model, allowing for deeper explanations and capturing reasoning about explanations and different degrees of understanding at an abstract level. This adds interesting layers to the analysis. Given a φ-phenomenon to be explained, we refer to $t : \varphi$ as an atomic explanation. By contrast, $s : t : \varphi$ represents a deeper explanation than the atomic one, as it involves more levels. Furthermore, we have the following:

- An explanation $t_n : \cdots : t_1 : \varphi$ is *deeper* than an explanation $t_m : \cdots : t_1 : \varphi$ iff $n > m$.
- $t_n : \cdots : t_1 : \varphi$ and $s_m : \cdots : s_1 : \varphi$ are *alternative explanations* of φ iff neither $t_n : \cdots : t_1 : \varphi$ is deeper than $s_m : \cdots : s_1 : \varphi$, nor $s_m : \cdots : s_1 : \varphi$ is deeper than $t_n : \cdots : t_1 : \varphi$.
- An explanation $t_n : \cdots : t_1 : \varphi$ is called *demanding* iff $n \geqslant 2$.
- An explanation $t_n : \cdots : t_1 : \varphi$ is *ideal* iff it cannot be deepen.

We therefore establish a strict partial order among different explanations of φ by applying the notion of deepening, and thereby facilitating comparative understanding.

Now, regarding the second question, Khalifa ([14]) suggests two approaches: we could analyze a minimal understanding by identifying the necessary conditions for any understanding; alternatively, we might explore a maximal or ideal kind of understanding, which would serve as the mirror image of the minimal understanding. In either case, developing a method to compare different degrees of understanding would allow us to describe the full spectrum, as represented by [14]:

Minimal understanding < Everyday understanding < Typical scientist's understanding < Ideal understanding.

Based on the framework in [29], identifying an explanation is the necessary conditions for any understanding, thus knowing why φ (Kyφ) could be regarded as the minimal understanding why φ occurs. [8] In addition, the understanding why ψ (Uyψ)

[8] It resembles the definition of minimal understanding in [14]: "One has minimal understanding of why p if and only if, for some q, she believes that q explains why p, and q explains why p is approximately true". Note that the notion of minimal understanding is analyzed by identifying only the necessary conditions for any understanding, which is not a typical notion of understanding.

characterized by only two levels of explanations could be viewed as the everyday understanding. So, we redefine multiple different degrees of understanding modalities in the language, collectively referred to as U, which are respectively:

- Minimal Understanding: $\mathrm{U}^{\mathrm{M}}\varphi$, requires an atomic explanation of φ, i.e., $\exists t\mathrm{K}(t:\varphi)$.
- Everyday Understanding: $\mathrm{U}^{\mathrm{E}}\varphi$, requires a two-level explanation of φ, i.e., $\exists t_1\exists t_2\mathrm{K}(t_2:t_1:\varphi)$.
- Demanding Understanding: $\mathrm{U}^{\mathrm{D}}\varphi$, requires a demanding explanation deeper than two-level of φ, i.e., $\exists t_1\cdots\exists t_m\mathrm{K}(t_m:\cdots:t_1:\varphi)$ $(m>2)$.
- Ideal Understanding: $\mathrm{U}^{\mathrm{I}}\varphi$, requires an ideal explanation of φ, i.e., $\exists t_1\cdots\exists t_n\exists c$ $\mathrm{K}(c:t_n:\cdots:t_1:\varphi)$ ($n\geqslant 2$, c denotes a self-evident explanation that cannot be deepen anymore).

Therefore, the spectrum of understanding expressed in our framework is:

Minimal Understanding U^{M} $<$ Everyday understanding U^{E} $<$ Demanding understanding U^{D} $<$ Ideal understanding U^{I}.

For the demanding understanding $\mathrm{U}^{\mathrm{D}}\varphi$, since the comparative measures of explanatory power offered in the extended framework are restricted to explanatory depth, we cannot further compare two alternative explanations, such as stating that one is closer to the correct scientific explanation than the other. Therefore, we are not yet able to properly define a typical scientific understanding, but instead define the demanding understanding $\mathrm{U}^{\mathrm{D}}\varphi$ that is better than the everyday understanding $\mathrm{U}^{\mathrm{E}}\varphi$. We will further discuss the issue of comparative principles for two alternative explanations in Sect. 5.

What follows is a brief summary. To formalize this semantics of different degrees of understanding, we need to deviate from the models defined in [29] by having more levels of explanations and ideal explanations in the models.

Our main contributions are summarized below.

- We formalize the notion of varying degrees of understanding using a logical language featuring four modalities for minimal understanding, everyday understanding, demanding understanding and ideal understanding respectively.
- We expand the existing model of everyday understanding to establish a partial order among different explanations, thereby facilitating comparative understanding.
- We provide a sound and complete axiomatization of our logic to characterize the interplay between different degrees of understanding.
- Using this framework, we explore its application to multi-agent scenarios, specifically for making comparative statements about the understanding among different agents and meta-understanding between them.

The paper is organized as follows. In Sect. 2, we build our logical framework. Sect. 3 gives an axiomatization, and the detailed completeness proof is shown in Sect.

4. Some applications of comparative understanding and meta-understanding in multi-agent scenarios are discussed in Sect. 5. The last section flags further work.

2 Syntax and Semantics

Definition 1 (Epistemic language of comparative understanding). Fix nonempty set P of propositional letters, the language **ELCU** is defined as (where $p \in P$):

$$\varphi ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathrm{K}\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi$$

Intuitively, the formula $\mathrm{U}^{\mathrm{M}}\varphi$ says that the agent has a minimal understanding of why φ. The expression $\mathrm{U}^{\mathrm{E}}\varphi$ indicates that the agent has an everyday understanding of why φ holds. $\mathrm{U}^{\mathrm{D}}\varphi$ suggests that the agent possesses a demanding understanding of why φ holds, surpassing an everyday understanding. $\mathrm{U}^{\mathrm{I}}\varphi$ says that the agent has an ideal understanding of why φ is true. In this paper, the modalities U^{M}, U^{E}, U^{D}, and U^{I} will collectively be referred to using the unified term U.

We accept the view in [33] that although something is a tautology, you may not have a minimal understanding (knowledge why) of that it is a tautology. A special set of "self-evident" tautologies Λ is introduced, which the agent is assumed to minimally understand. For example, we can let all the instances of $\varphi \wedge \psi \to \varphi$ and $\varphi \wedge \psi \to \psi$ be Λ. At present, we do not suppose any necessitation rule for U in general.

Definition 2. An **ELCU** model $\mathcal{M}$ is a tuple $(W, E, R, \mathcal{E}, V)$ where:

- W is a non-empty set of possible worlds.
- E is a non-empty set of explanations equipped with operators $\cdot$, ! and c such that:
  1. If $t, s \in E$, then $t \cdot s \in E$,
  2. If $t \in E$, then $!t \in E$,
  3. A special symbol c is in E satisfying that $c \cdot c = c$.
- $R \subseteq W \times W$ is an equivalence relation over W.
- $\mathcal{E} : E^n \times \textbf{ELCU} \to 2^W$ $(n \geqslant 1)$ is an admissible explanation function satisfying the following conditions:

Explanation Application:

$\mathcal{E}(\langle t_n, \ldots, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}(\langle s_n, \ldots, s_1\rangle, \varphi)$
$\subseteq \mathcal{E}(\langle t_n \cdot s_n, \ldots, t_1 \cdot s_1\rangle, \psi)$.

Constant Specification:

If $\varphi \in \Lambda$, then $\mathcal{E}(c, \varphi) = W$,

Higher-level Explanation Factivity:

$\mathcal{E}(\langle t_{n+1}, t_n, \ldots, t_1\rangle, \varphi) \subseteq \mathcal{E}(\langle t_n, \ldots, t_1\rangle, \varphi)$.

***Epistemic Introspection*:**

$\mathcal{E}(t, \bigcirc\varphi) \subseteq \mathcal{E}(!t, \langle t, \bigcirc\varphi\rangle)$ for $\bigcirc = \mathrm{K}, \mathrm{U}$.

***Ideal Explanation*:**

If $w \in \mathcal{E}(\langle c, t_m, \ldots, t_1\rangle, \varphi)$ $(m \geqslant 2)$,
then $w \notin \mathcal{E}(\langle s, c, t_m, \ldots, t_1\rangle, \varphi)$ for any $s \in E$.

***Ideal Explanation Application I*:**

$$\mathcal{E}(\langle c, t_m, \ldots, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}(\langle c, s_n, \ldots, s_1\rangle, \varphi)$$
$$\subseteq \mathcal{E}(\langle c, t_m, \ldots, t_k \cdot c, t_n \cdot s_n, \ldots, t_1 \cdot s_1\rangle, \psi)\ (m > n \geqslant 2, m \geqslant k).$$

***Ideal Explanation Application II*:**

$$\mathcal{E}(\langle c, t_m, \ldots, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}(\langle c, s_n, \ldots, s_1\rangle, \varphi)$$
$$\subseteq \mathcal{E}(\langle c, s_n, \ldots, c \cdot s_k, t_m \cdot s_m, \ldots, t_1 \cdot s_1\rangle, \psi)\ (n > m \geqslant 2, n \geqslant k).$$

- $V : P \to 2^W$ is a valuation function.

In justification logic, operators such as $\cdot$, !, and c are conventionally used ([1, 15]). The set E is closed under the application operator $\cdot$, which combines two explanations into a single one, as well as under the positive introspection operator !. Additionally, a special symbol c is in E, which is a self-evident explanation. It is required to fulfill dual roles within the model: uniformly for all the self-evident formulas within the designated set Λ and uniformly all self-evident higher-order explanations. $c \cdot c = c$ is natural since c denotes any self-evident explanation. The sum operator $+$ is excluded because it would satisfy the condition $\mathcal{E}(t, \varphi) \cup \mathcal{E}(s, \varphi) \subseteq \mathcal{E}(t + s, \varphi)$. This is problematic in scenarios where different worlds may have different explanations $(t_1, \ldots, t_n)$ for the same formula φ, as $\mathrm{U}^{\mathrm{M}}\varphi$ might be incorrectly deduced from a uniform explanation formed by $t_1 + \ldots + t_n$.

The admissible explanation function $\mathcal{E}$ specifies the set of worlds for both first-level and higher-level explanations. If $w \in \mathcal{E}(t, \varphi)$, then t is a first-level explanation for φ in the world w. If $v \in \mathcal{E}(\langle t_n, \ldots, t_1\rangle, \psi)$, then v is a world where $\langle t_n, \ldots, t_1\rangle$ provides a higher-level explanation of ψ. Note that the notation $\mathcal{E}(\langle t_n, \ldots, t_1\rangle, \psi)$ here corresponds to the format $t_n : \cdots : t_1 : \psi$ in the Sect. 1.3.

The first two conditions are for $\mathcal{E}$ is obvious. The third condition says that higher-level explanations yield lower-level explanations, that is, $w \in \mathcal{E}(\langle t_{n+1}, t_n, \ldots, t_1\rangle, \varphi)$ implies w is a world where a lower-level explanation $\langle t_n, \ldots, t_1\rangle$ explains φ.

The reason the *epistemic introspection* condition of $\mathcal{E}$ is introduced is elaborated in [29]. Intuitively, an answer to a question "Why φ?" is an explanation of the fact φ. Thus the formula $\mathrm{U}^{\mathrm{M}}\mathrm{K}\varphi$, which requires an atomic explanation of $\mathrm{K}\varphi$ roughly, pertains to a why-question: why one knows φ. Typically, the person posing this question does not expect the agent to provide reasons for why her belief in φ is not subject to Gettier problems; instead, the agent should simply articulate her reasons for believing φ, i.e., her justification for φ (see [2]). In this context, justification essentially serves as an explanation. If $w \in \mathcal{E}(t, \mathrm{K}\varphi)$, then w is also a world where t justifies φ. Consequently, there is a conceptual link between "t explains $\mathrm{K}\varphi$" in this framework and "t is a justification for φ" in standard justification logics, in the sense that explanations can be justifications.

Justification logics commonly adhere to the logical principle: $t : \varphi \to !t : (t : \varphi)$. Fitting argues in [9] that we are typically able to substantiate the reasons underlying our knowledge in everyday life, and indeed, a purported reason holds no value without some justification for its validity. Therefore, this principle is essential in justification logic, asserting that $!t$ always serves as a justification for $t : \varphi$, or that $!t$ is an introspective act confirming $t : \varphi$. Consequently, a parallel logical principle can be derived here: if $w \in \mathcal{E}(t, \mathrm{K}\varphi)$ then $w \in \mathcal{E}(\langle !t, t\rangle, \mathrm{K}\varphi)$, and without obstacle to include the understanding modalities U for the same reason as that of Fitting. It is interesting to note that the *epistemic introspection* condition will give rise to the following nontrivial axioms about understanding: $\mathrm{U^M K}\varphi \to \mathrm{U^E K}\varphi$ and $\mathrm{U^M U}\varphi \to \mathrm{U^E U}\varphi$.

All in all, if there is an explanation t for epistemic claims, then an introspective second-level explanation $!t$ of t must always exist to facilitate everyday understanding. Conversely, when t serves as an explanation for the non-epistemic claim φ and is not a justification, it does not necessarily follow that t can be transformed into a second-level explanation for why t explains φ.

Finally, we incorporate the *ideal explanation* condition and the *ideal explanation application I* and *ideal explanation application II* conditions into $\mathcal{E}$ in order to capture the explanations that lead to ideal understanding, which will be further clarified according to the semantics below.

Definition 3.

$\mathcal{M}, w \vDash p$	$\Leftrightarrow$	$w \in V(p)$
$\mathcal{M}, w \vDash \neg\varphi$	$\Leftrightarrow$	$\mathcal{M}, w \nvDash \varphi$
$\mathcal{M}, w \vDash \varphi \wedge \psi$	$\Leftrightarrow$	$\mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$
$\mathcal{M}, w \vDash \mathrm{K}\varphi$	$\Leftrightarrow$	$\mathcal{M}, v \vDash \varphi$ for all v such that wRv
$\mathcal{M}, w \vDash \mathrm{U^M}\varphi$	$\Leftrightarrow$	(1) there exists $t \in E$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(t, \varphi)$ (2) for all $v \in W$ with wRv, $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \mathrm{U^E}\varphi$	$\Leftrightarrow$	(1) there exist $t_1, t_2 \in E$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(\langle t_2, t_1\rangle, \varphi)$; (2) for all $v \in W$ with wRv, $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \mathrm{U^D}\varphi$	$\Leftrightarrow$	(1) there exist $t_1, \dots, t_n \in E$ $(n > 2)$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(\langle t_n, \dots, t_1\rangle, \varphi)$; (2) for all $v \in W$ with wRv, $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \mathrm{U^I}\varphi$	$\Leftrightarrow$	(1) there exist $t_1, \dots, t_m \in E$ such that for all $v \in W$ with wRv, $v \in \mathcal{E}(\langle c, t_m, \dots, t_1\rangle, \varphi)$ $(m \geqslant 2)$; (2) for all $v \in W$ with wRv, $\mathcal{M}, v \vDash \varphi$

Note that if the principle that everything has an explanation underpins our framework, which may be represented by the condition on $\mathcal{E}$, expressed as $w \in \mathcal{E}(\langle t_k, \dots, t_1\rangle,$

$\varphi)$ for some $t_1, \ldots, t_k$ with $k \geqslant 1$ for any $w \in W$, then it is equivalent to write the truth condition of K-formula as follows:

$$\begin{aligned}\mathcal{M}, w \vDash \mathrm{K}\varphi \quad \Leftrightarrow \quad & (1) \text{ for all } v \in W \text{ with } wRv, \text{ there exist } t_i, \ldots, t_k \ (k \geqslant 1) \\ & \quad \text{such that } v \in \mathcal{E}(\langle t_k, \ldots, t_1 \rangle, \varphi), \\ & (2) \text{ for all } v \in W \text{ with } wRv,\ \mathcal{M}, v \vDash \varphi.\end{aligned}$$

That is, the structure of $\mathrm{K}\varphi$ can be displayed as $\mathrm{K}\exists t_1 \cdots \exists t_k(t_k : \cdots : t_1 : \varphi)$. In this case, the core difference between K-formulas and U-formulas in truth conditions lies in the nesting structure of quantifiers. For $\mathrm{K}\varphi$, the structure is $\forall\exists \cdots \exists$, while for $\mathrm{U}\varphi$, it is $\exists \cdots \exists\forall$, where $\forall$ checks all possible states, and $\exists \cdots \exists$ seeks explanations.

We have got the *higher-level explanation factivity* condition in the models. Now, as in [29], we can show that the *first-level explanation factivity* defined in the following is not assumed in the model definition.

Definition 4. An **ELCU** model $\mathcal{M}$ has the property of first-level explanation factivity, if whenever $w \in \mathcal{E}(t, \varphi)$, then $\mathcal{M}, w \vDash \varphi$.

Given an **ELCU** model $\mathcal{M} = (W, E, R, \mathcal{E}, V)$, its first-level factive companion $\mathcal{M}^F = (W, E, R, \mathcal{E}^F, V)$ can be constructed as $\mathcal{E}^F(\langle t_n, \ldots, t_1 \rangle, \varphi) = \mathcal{E}(\langle t_n, \ldots, t_1 \rangle, \varphi) - \{w \mid \mathcal{M}, w \nvDash \varphi\}$ $(n \geqslant 1)$. It is obvious that the $\mathcal{M}^F$ constructed is indeed an **ELCU** model. The proposition below asserts that the **ELCU**-formulas are neutral with respect to the first-level explanation factivity.

Proposition 1. *For any* **ELCU** *formula φ, any $w \in W$, $\mathcal{M}, w \vDash \varphi$ iff $\mathcal{M}^F, w \vDash \varphi$.*

Proof. We do induction on the structure of the **ELCU**-formula. Boolean cases and the case of $\mathrm{K}\varphi$ are trivial. For U, we only check the case for U^{D} below:

- $\Longleftarrow$ Suppose $\mathcal{M}^F, w \vDash \mathrm{U}^{\mathrm{D}}\varphi$, then there exist $t_1, \ldots, t_n \in E$ $(n > 2)$ such that for all v with wRv, we have $\mathcal{M}^F, v \vDash \varphi$ and $v \in \mathcal{E}^F(\langle t_n, \ldots, t_1 \rangle, \varphi)$. Then by definition we get $v \in \mathcal{E}(\langle t_n, \ldots, t_1 \rangle, \varphi)$. Therefore by IH $\mathcal{M}, w \vDash \mathrm{U}^{\mathrm{D}}\varphi$.
- $\Longrightarrow$ The proof is similar as above.

□

3 Axiomatization

Now we develop the the proof system SCU for varying degrees of understanding.

It is worth noting that the axiom (KYU) expresses that "everyday understanding" is necessary for "minimal understanding" (ordinary knowing why) in epistemic situations, which corresponds to the *epistemic introspection* condition in the model.

System SCU

Axioms

(TAUT)	Classical Propositional Axioms	
(DISTK)	$\mathrm{K}(\varphi \to \psi) \to (\mathrm{K}\varphi \to \mathrm{K}\psi)$	
(T)	$\mathrm{K}\varphi \to \varphi$	
(4)	$\mathrm{K}\varphi \to \mathrm{KK}\varphi$	
(5)	$\neg\mathrm{K}\varphi \to \mathrm{K}\neg\mathrm{K}\varphi$	
(DISTU)	$\mathrm{U}(\varphi \to \psi) \to (\mathrm{U}\varphi \to \mathrm{U}\psi)$	(for $\mathrm{U} = \mathrm{U^M}, \mathrm{U^E}, \mathrm{U^D}, \mathrm{U^I}$)
(UYK)	$\mathrm{U^M}\varphi \to \mathrm{K}\varphi$	
(IYD)	$\mathrm{U^I}\varphi \to \mathrm{U^D}\varphi$	
(DYE)	$\mathrm{U^D}\varphi \to \mathrm{U^E}\varphi$	
(EYM)	$\mathrm{U^E}\varphi \to \mathrm{U^M}\varphi$	
(4*)	$\mathrm{U}\varphi \to \mathrm{KU}\varphi$	(for $\mathrm{U} = \mathrm{U^M}, \mathrm{U^E}, \mathrm{U^D}, \mathrm{U^I}$)
(KYU)	$\mathrm{U^M} \bigcirc \varphi \to \mathrm{U^E} \bigcirc \varphi$	(for $\bigcirc = \mathrm{K}, \mathrm{U}$)

Rules

(MP) Modus Ponens (N) $\vdash \varphi \Rightarrow \vdash \mathrm{K}\varphi$ (NE) $\varphi \in \Lambda \Rightarrow \vdash \mathrm{U^M}\varphi$

Recall that introspection of understanding is proposed by Medieval logicians (as mentioned in Footnote 1). Do we have the validity of $\mathrm{U}\varphi \to \mathrm{UU}\varphi$ for some specific degree of understanding over the **ELCU** models? Unfortunately, it is not valid in the current setting. However, we could discuss under what conditions it might be obtained. Once we accept $\mathrm{U^E}\varphi \to \mathrm{U^M U^E}\varphi$, with the help of $\mathrm{U^M U^E}\varphi \to \mathrm{U^E U^E}\varphi$–which is an instantiation of (KYU)–introspection of everyday understanding, $\mathrm{U^E}\varphi \to \mathrm{U^E U^E}\varphi$, will be valid.

Proposition 2. *The following is provable in* SCU*:* (5*) $\neg\mathrm{U}\varphi \to \mathrm{K}\neg\mathrm{U}\varphi$ *(for* $\mathrm{U} = \mathrm{U^M}, \mathrm{U^E}, \mathrm{U^D}, \mathrm{U^I}$*).*

Proof.

(1)	$\neg\mathrm{U}\varphi \to \neg\mathrm{KU}\varphi$	(T)
(2)	$\neg\mathrm{KU}\varphi \to \mathrm{K}\neg\mathrm{KU}\varphi$	(5)
(3)	$\mathrm{K}\neg\mathrm{KU}\varphi \to \mathrm{K}\neg\mathrm{U}\varphi$	(4*), normality of K
(4)	$\neg\mathrm{U}\varphi \to \mathrm{K}\neg\mathrm{U}\varphi$	(MP)

□

Theorem 1. SCU *is sound over* **ELCU** *models.*

Proof. We omit the cases of standard axioms and rules, as well most other cases without special tricks.

DISTU : For instance, suppose suppose $\mathcal{M}, w \vDash \mathrm{U^I}(\varphi \to \psi)$ and $\mathcal{M}, w \vDash \mathrm{U^I}\varphi$ for any **ELCU** model $\mathcal{M}$. Then there exist $t_1, \ldots, t_m, s_1, \ldots, s_n \in E$ $(m, n \geqslant 2)$ such that for all v with wRv, $v \in \mathcal{E}(\langle c, t_m, \ldots, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}(\langle c, s_n, \ldots, s_1\rangle, \varphi)$. Assume without loss of generality that $m > n$, we have $v \in \mathcal{E}(\langle c, t_m, \ldots, t_k \cdot c, t_n \cdot s_n, \ldots, t_1 \cdot s_1\rangle, \psi)$ $(m \geqslant k)$ by the *ideal explanation application I* condition of $\mathcal{E}$. Therefore, $\mathcal{M}, w \vDash \mathrm{U^I}\psi$. □

4 Completeness

Since the spectrum of understanding expressed in our framework is: $\mathrm{U}^{\mathrm{M}} < \mathrm{U}^{\mathrm{E}} < \mathrm{U}^{\mathrm{D}} < \mathrm{U}^{\mathrm{I}}$, thus, when constructing the canonical model, it is actually sufficient to temporarily include four levels of explanations. As discussed in the next Sect. 5, once we apply the **ELCU** models to the multi-agents situations, the restrictions on the four levels of explanations need to be further relaxed. The technical details of the proofs are inspired by [33] and [29].

Let Ω denote the set of all maximal SCU-consistent sets of formulas.

Definition 5 (Canonical Model). The canonical model $\mathcal{M}^c$ for SCU is a tuple $(W^c, E^c, R^c, \mathcal{E}^c, V^c)$ where:

- E^c is defined in BNF: $t ::= c \mid \varphi \mid (t \cdot t) \mid !t$, satisfying $c \cdot c = c$, where $\varphi \in \mathbf{ELCU}$.
- $W^c := \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid \langle \Gamma, F, G, H, L\rangle \in \Omega \times \mathcal{P}(E^c \times \mathbf{ELCU}) \times \mathcal{P}(E^{c2} \times \mathbf{ELCU}) \times \mathcal{P}(E^{c3} \times \mathbf{ELCU}) \times \mathcal{P}(\{c\} \times E^{c3} \times \mathbf{ELCU}), f : \{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\} \to E^c, g : \{\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Gamma\} \to E^{c2}, h : \{\mathrm{U}^{\mathrm{D}}\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Gamma\} \to E^{c3}, l : \{\mathrm{U}^{\mathrm{I}}\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Gamma\} \to \{c\} \times E^{c3}$ such that f, g, h and l satisfy the following conditions$\}$

1. If $\langle t, \varphi \to \psi\rangle, \langle s, \varphi\rangle \in F$, then $\langle t \cdot s, \psi\rangle \in F$.
2. If $\langle t_2, t_1, \varphi \to \psi\rangle, \langle s_2, s_1, \varphi\rangle \in G$ then $\langle t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in G$.
3. If $\langle t_3, t_2, t_1, \varphi \to \psi\rangle, \langle s_3, s_2, s_1, \varphi\rangle \in H$ then $\langle t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in H$.
4. If $\langle c, t_3, t_2, t_1, \varphi \to \psi\rangle, \langle c, s_3, s_2, s_1, \varphi\rangle \in L$ then $\langle c, t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L$.
5. If $\langle c, t_3, t_2, t_1, \varphi \to \psi\rangle \in L, \langle c, s_2, s_1, \varphi\rangle \in H$ then $\langle c, t_3 \cdot c, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L$.
6. If $\langle c, t_2, t_1, \varphi \to \psi\rangle \in H, \langle c, s_3, s_2, s_1, \varphi\rangle \in L$ then $\langle c, c \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L$.
7. If $\varphi \in \Lambda$, then $\langle c, \varphi\rangle \in F$.
8. $\langle t_2, t_1, \varphi\rangle \in G$ implies $\langle t_1, \varphi\rangle \in F$
9. $\langle t_3, t_2, t_1, \varphi\rangle \in H$ implies $\langle t_2, t_1, \varphi\rangle \in G$.
10. $\langle c, t_3, t_2, t_1, \varphi\rangle \in L$ implies $\langle t_3, t_2, t_1, \varphi\rangle \in H$.
11. $\langle c, t_2, t_1, \varphi\rangle \in H$ implies $\langle c, c, t_2, t_1, \varphi\rangle \notin L$.
12. $\langle t, \bigcirc\varphi\rangle \in F$ implies $\langle !t, t, \bigcirc\varphi\rangle \in G$ for $\bigcirc = \mathrm{K}, \mathrm{U}$.
13. $\mathrm{U}^{\mathrm{M}}\varphi \in \Gamma$ implies $\langle f(\varphi), \varphi\rangle \in F$.
14. $\mathrm{U}^{\mathrm{E}}\varphi \in \Gamma$ implies $\langle g(\varphi), \varphi\rangle \in G$.
15. $\mathrm{U}^{\mathrm{D}}\varphi \in \Gamma$ implies $\langle h(\varphi), \varphi\rangle \in H$.
16. $\mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$ implies $\langle l(\varphi), \varphi\rangle \in L$.

- $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$ iff (1) $\{\varphi \mid \mathrm{K}\varphi \in \Gamma\} \subseteq \Delta$, and (2) $f = f', g = g', h = h', l = l'$.
- – $\mathcal{E}^c : E^{cn} \times$ **ELCU** $\to 2^{W^c}$ $(1 \leqslant n \leqslant 4)$ is defined by

$$\begin{cases} \mathcal{E}^c(t, \varphi) = \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid \langle t, \varphi\rangle \in F\} \\ \mathcal{E}^c(\langle t_2, t_1\rangle, \varphi) = \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid \langle t_2, t_1, \varphi\rangle \in G\} \\ \mathcal{E}^c(\langle t_3, t_2, t_1\rangle, \varphi) = \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid \langle t_3, t_2, t_1, \varphi\rangle \in H\} \\ \mathcal{E}^c(\langle t_4, t_3, t_2, t_1\rangle, \varphi) = \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid \langle t_4, t_3, t_2, t_1, \varphi\rangle \in L\} \end{cases}$$

  – $\mathcal{E}^c : E^{cn} \times$**ELCU** $\to 2^{W^c}$ $(n > 4)$ is defined as: $\mathcal{E}^c(\langle t_n, \ldots, t_1\rangle, \varphi) = \emptyset$.
- $V^c(p) = \{\langle \Gamma, F, G, H, L, f, g, h, l\rangle \mid p \in \Gamma\}$.

In the construction, for each world $\langle \Gamma, F, G, H, L, f, g, h, l\rangle \in W^c$, it contains information about the explanations leading to each degree of U formulas in Γ, respectively. More specifically, f is a witness function picking one t for each formula in $\{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\}$, with the information that t explains φ stored in the component F. Similarly, h, for instance, is a witness function that selects one $\langle t_3, t_2, t_1\rangle$ for each formula in $\{\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Gamma\}$, and the information that $\langle t_3, t_2, t_1\rangle$ explains φ is stored in H.

The following shows that W^c is indeed nonempty.

Definition 6. Given any $\Gamma \in \Omega$, construct $F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma$ as follows:

- $F^\Gamma_0 = \{\langle \varphi, \varphi\rangle \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\} \cup \{\langle c, \varphi\rangle \mid \varphi \in \Lambda\}$
- $G^\Gamma_0 = \{\langle \varphi \cdot \varphi, !\varphi, \varphi\rangle \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Gamma\}$
- $H^\Gamma_0 = \{\langle !!\varphi, \varphi \cdot \varphi, !\varphi, \varphi\rangle \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Gamma\}$
- $L^\Gamma_0 = \{\langle c, !!\varphi, \varphi \cdot \varphi, !\varphi, \varphi\rangle \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Gamma\}$
- $F^\Gamma_{n+1} = F^\Gamma_n \cup \{\langle t \cdot s, \psi\rangle \mid \langle t, \varphi \to \psi\rangle, \langle s, \varphi\rangle \in F^\Gamma_n \text{ for some } \varphi\} \cup \{\langle t_1, \varphi\rangle \mid \langle t_2, t_1, \varphi\rangle \in G^\Gamma_n\}$
- $G^\Gamma_{n+1} = G^\Gamma_n \cup \{\langle t_2 {\cdot} s_2, t_1 {\cdot} s_1, \psi\rangle \mid \langle t_2, t_1, \varphi \to \psi\rangle, \langle s_2, s_1, \varphi\rangle \in G^\Gamma_n \text{ for some } \varphi\} \cup \{\langle t_2, t_1, \varphi\rangle \mid \langle t_3, t_2, t_1, \varphi\rangle \in H^\Gamma_n\} \cup \{\langle !t, t, \bigcirc\varphi\rangle \mid \langle t, \bigcirc\varphi\rangle \in F^\Gamma_n \text{ for } \bigcirc = \mathrm{K}, \mathrm{U}\}$
- $H^\Gamma_{n+1} = H^\Gamma_n \cup \{\langle t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \mid \langle t_3, t_2, t_1, \varphi \to \psi\rangle, \langle s_3, s_2, s_1, \varphi\rangle \in H^\Gamma_n \text{ for some } \varphi\} \cup \{\langle t_3, t_2, t_1, \varphi\rangle \mid \langle c, t_3, t_2, t_1, \varphi\rangle \in L^\Gamma_n\}$
- $L^\Gamma_{n+1} = L^\Gamma_n \cup \{\langle c, t_3 {\cdot} s_3, t_2 {\cdot} s_2, t_1 {\cdot} s_1, \psi\rangle \mid \langle c, t_3, t_2, t_1, \varphi \to \psi\rangle, \langle c, s_3, s_2, s_1, \varphi\rangle \in L^\Gamma_n \text{ for some } \varphi\} \cup \{\langle c, t_3 \cdot c, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \mid \langle c, t_3, t_2, t_1, \varphi \to \psi\rangle \in L^\Gamma_n, \langle c, s_2, s_1, \varphi\rangle \in H^\Gamma_n \text{ for some } \varphi\} \cup \{\langle c, c {\cdot} s_3, t_2 {\cdot} s_2, t_1 {\cdot} s_1, \psi\rangle \mid \langle c, t_2, t_1, \varphi \to \psi\rangle \in H^\Gamma_n, \langle c, s_3, s_2, s_1, \varphi\rangle \in L^\Gamma_n \text{ for some } \varphi\}$
- $F^\Gamma = \bigcup_{n \in \mathbb{N}} F^\Gamma_n$
- $G^\Gamma = \bigcup_{n \in \mathbb{N}} G^\Gamma_n$
- $H^\Gamma = \bigcup_{n \in \mathbb{N}} H^\Gamma_n$
- $L^\Gamma = \bigcup_{n \in \mathbb{N}} L^\Gamma_n$

- $f^\Gamma : \{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\} \to E^c, f^\Gamma(\varphi) = \varphi.$
- $g^\Gamma : \{\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Gamma\} \to E^c \times E^c, g^\Gamma(\varphi) = \langle \varphi \cdot \varphi, !\varphi\rangle.$
- $h^\Gamma : \{\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Gamma\} \to E^c \times E^c \times E^c, h^\Gamma(\varphi) = \langle !!\varphi, \varphi \cdot \varphi, !\varphi\rangle.$
- $l^\Gamma : \{\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Gamma\} \to \{c \mid c \in E^c\} \times E^c \times E^c, l^\Gamma(\varphi) = \langle c, !!\varphi, \varphi \cdot \varphi, !\varphi\rangle.$

Proposition 3. *For any* $\Gamma \in \Omega$, $\langle \Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in W^c$.

Proof. We show that the conditions $1-16$ in the definition of W^c are all satisfied. Merely selected conditions are discussed below:

- For the condition 1, suppose $\langle t_2, t_1, \varphi \to \psi\rangle, \langle s_2, s_1, \varphi\rangle \in G^\Gamma$. Then there exist $k, l \in \mathbb{N}$ such that $\langle t_2, t_1, \varphi \to \psi\rangle \in G^\Gamma_k, \langle s_2, s_1, \varphi\rangle \in G^\Gamma_l$. Assume without loss of generality that $k > l$. Then we get $\langle t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in G^\Gamma_{k+1}$ by the construction. Therefore $\langle t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in G^\Gamma$.
- For the condition 6, suppose $\langle c, t_2, t_1, \varphi \to \psi\rangle \in H^\Gamma$ and $\langle c, s_3, s_2, s_1, \varphi\rangle \in L^\Gamma$, then there are $k, l \in \mathbb{N}$ such that $\langle c, t_2, t_1, \varphi \to \psi\rangle \in H^\Gamma_k$, $\langle c, s_3, s_2, s_1, \varphi\rangle \in L^\Gamma_l$. Assume without loss of generality that $k > l$. Then we get $\langle c, c \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L^\Gamma_{k+1}$ by the construction. Hence $\langle c, c \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L^\Gamma$.
- For the condition 12, suppose $\langle t, \bigcirc\varphi\rangle \in F^\Gamma$. Then we have $\langle t, \bigcirc\varphi\rangle \in F^\Gamma_k$ for some $k \in \mathbb{N}$, which implies $\langle !t, t, \bigcirc\varphi\rangle \in G^\Gamma_{k+1}$ by the construction of G^Γ.
- For the condition 14, suppose $\mathrm{U}^{\mathrm{E}}\varphi \in \Gamma$. Then we get $\langle \varphi \cdot \varphi, !\varphi, \varphi\rangle \in G^\Gamma$ by the constructions of G^Γ_0 and G^Γ. Moreover, we have $\langle g^\Gamma(\varphi), \varphi\rangle \in G^\Gamma$ by the construction of g^Γ.
- For the condition 16, suppose $\mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$. Then we get $\langle c, !!\varphi, \varphi \cdot \varphi, !\varphi, \varphi\rangle \in L^\Gamma$ by the constructions of L^Γ_0 and L^Γ. Moreover, we have $\langle l^\Gamma(\varphi), \varphi\rangle \in L^\Gamma$ by the construction of l^Γ. □

For the construction of R^c in the canonical model in Definition 5, we claim:

Proposition 4. *R^c is an equivalence relation.*

Proof. It is trivial by the construction of R^c and axioms (T), (4) and (5). □

Regarding $\mathcal{E}^c$ in canonical models, we check the following:

Proposition 5. *$\mathcal{E}^c$ satisfies all the conditions in* **ELCU** *model definition.*

Proof. We only check some of these cases:

Explanation application: Suppose, for instance, $\langle \Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in \mathcal{E}^c(t, \varphi \to \psi) \cap \mathcal{E}^c(s, \varphi)$. By the construction of $\mathcal{E}^c$, we have both $\langle t, \varphi \to \psi\rangle$ and $\langle s, \varphi\rangle$ are in F. Then by condition 1 of W^c, we have $\langle t \cdot s, \psi\rangle \in F$, which means $\langle \Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in \mathcal{E}^c(t \cdot s, \psi)$.

Epistemic introspection: Obvious by condition 12.

Ideal Explanation: Suppose, for instance, $\langle\Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in \mathcal{E}(\langle c, t_2, t_1\rangle, \varphi)$. Then, by the construction of $\mathcal{E}^c$, we have $\mathcal{E}(\langle s, c, t_2, t_1\rangle, \varphi) = \emptyset$ for any $s \in E^c$. Thus $\langle\Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \notin \mathcal{E}(\langle s, c, t_2, t_1\rangle, \varphi)$ for any $s \in E^c$.

Ideal Explanation Application I: Suppose $\langle\Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in \mathcal{E}(\langle c, t_3, t_2, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}(\langle c, s_2, s_1\rangle, \varphi)$. Then, by the construction of $\mathcal{E}^c$, we have $\langle c, t_3, t_2, t_1, \varphi \to \psi\rangle \in L$ and $\langle c, s_2, s_1, \varphi \to \psi\rangle \in H$, which imply $\langle c, t_3 \cdot c, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L$ by the condition 5 of W^c. Therefore, $\langle\Gamma, F^\Gamma, G^\Gamma, H^\Gamma, L^\Gamma, f^\Gamma, g^\Gamma, h^\Gamma, l^\Gamma\rangle \in \mathcal{E}(\langle c, t_3 \cdot c, t_2 \cdot s_2, t_1 \cdot s_1\rangle, \psi)$ by the definition. □

Hence the canonical model is well-defined, based on Proposition 3, 4 and 5.

Proposition 6. *The canonical model $\mathcal{M}^c$ is well-defined.*

Now we prove the existence lemmas for K, $\mathrm{U^M}$, $\mathrm{U^E}$,$\mathrm{U^D}$and $\mathrm{U^I}$ respectively.

Lemma 1 (K-Existence Lemma). *For any $\langle\Gamma, F, G, H, L, f, g, h, l\rangle \in W^c$, if $\neg\mathrm{K}\varphi \in \Gamma$, then there exists a $\langle\Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ such that $\langle\Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle\Delta, F', G', H', L', f', g', h', l'\rangle$, and $\neg\varphi \in \Delta$.*

Proof. (Sketch) Suppose $\neg\mathrm{K}\varphi \in \Gamma$. Let $\Delta^- = \{\psi \mid \mathrm{K}\psi \in \Gamma\} \cup \{\neg\varphi\}$. First, Δ^- is consistent. The proof is routine by (DISTK) and (N). Next we extend Δ^- into a MCS Δ. Finally, we construct $F', G', H', L', f', g', h', l'$ to form a world in W^c. We can simply let $F' = F$, $G' = G$, $H' = H$, $L' = L$ and $f' = f, g' = g, h' = h, l' = l$. □

To semantically refute $\mathrm{U^M}\psi$ while preserving $\mathrm{K}\psi$, we could construct an accessible world where the first-level explanation for ψ differs from that of the current world. In [33], all original first-level explanations for φ are replaced with different ones during the construction. However, we will simplify the demonstration by deleting all those explanations for ψ when constructing a canonical world that refutes $\mathrm{U^M}\psi$, as in [29].

Lemma 2 ($\mathrm{U^M}$-Existence Lemma). *For any $\langle\Gamma, F, G, H, L, f, g, h, l\rangle \in W^c$ where $\mathrm{K}\psi \in \Gamma$, if $\mathrm{U^M}\varphi \notin \Gamma$, then for any $\langle t, \psi\rangle \in F$, there exists a $\langle\Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ such that $\langle t, \psi\rangle \notin F'$ and $\langle\Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle\Delta, F', G', H', L', f', g', h', l'\rangle$.*

Proof. Suppose $\mathrm{U^M}\psi \notin \Gamma$, we construct $\langle\Delta, F', G', H', L', f', g', h', l'\rangle$ as follows:

- $\Delta = \Gamma$
- $F' = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F$ and $\mathrm{U^M}\varphi \in \Gamma\}$
- $G' = \{\langle s_2, s_1, \varphi\rangle \mid \langle s_2, s_1, \varphi\rangle \in G$ and $\mathrm{U^M}\varphi \in \Gamma\}$
- $H' = \{\langle s_3, s_2, s_1, \varphi\rangle \mid \langle s_3, s_2, s_1, \varphi\rangle \in H$ and $\mathrm{U^M}\varphi \in \Gamma\}$

- $L' = \{\langle c, s_3, s_2, s_1, \varphi\rangle \mid \langle c, s_3, s_2, s_1, \varphi\rangle \in L$ and $\mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\}$
- $f' : \{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Delta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Delta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Delta\} \to E^c$ is defined as: $h'(\varphi) = h(\varphi)$
- $l' : \{\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Delta\} \to E^c$ is defined as: $l'(\varphi) = l(\varphi)$

Note that $F' \subseteq F, G' \subseteq G, H' \subseteq H, L' \subseteq L$. The main idea behind the constructions of F' , G', H' and L' is to "carefully" delete all first-level explanations for $\{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \notin \Gamma\}$. Given $\mathrm{U}^{\mathrm{M}}\psi \notin \Gamma$, it is clear that $\langle t, \psi\rangle \notin F'$ for any $\langle t, \psi\rangle \in F$ by the construction. In order to complete this proof, firstly, we show that $\langle \Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ by checking the conditions $1 - 16$ in the definition of W^c. Only several cases are written below:

- For the condition 4, suppose $\langle c, t_3, t_2, t_1, \varphi \to \psi\rangle, \langle c, s_3, s_2, s_1, \varphi\rangle \in L' \subseteq L$, then $\langle c\cdot c, t_3\cdot s_3, t_2\cdot s_2, t_1\cdot s_1, \varphi\rangle = \langle c, t_3\cdot s_3, t_2\cdot s_2, t_1\cdot s_1, \varphi\rangle \in L, \mathrm{U}^{\mathrm{M}}(\varphi \to \psi)$ and $\mathrm{U}^{\mathrm{M}}\varphi \in \Gamma$. Moreover due to the axiom (DISTU) and the property of MCS, we have $\mathrm{U}^{\mathrm{M}}\psi \in \Gamma$. Hence $\langle c, t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L'$.
- For condition 7, suppose $\varphi \in \Lambda$, then $\mathrm{U}^{\mathrm{M}}\varphi \in \Gamma$ by (NE) and the property of MCS, which implies $\langle c, \varphi\rangle \in F'$.
- For condition 11, suppose $\langle c, t_2, t_1, \varphi\rangle \in H' \subseteq H$, then $\langle c, c, t_2, t_1, \varphi\rangle \notin L$, which implies $\langle c, c, t_2, t_1, \varphi\rangle \notin L'$.
- For condition 12, suppose $\langle t, \bigcirc\varphi\rangle \in F' \subseteq F$. Then we get $\langle !t, t, \bigcirc\varphi\rangle \in G$ and $\mathrm{U}^{\mathrm{M}} \bigcirc \varphi \in \Gamma$, which implies $\langle !t, t, \bigcirc\varphi\rangle \in G'$.
- For condition 16, suppose $\mathrm{U}^{\mathrm{I}}\varphi \in \Delta$. Then we get $\mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$ by $\Gamma = \Delta$, thus $\langle l(\varphi), \varphi\rangle \in L$. By (IYD), (DYE), (EYM) and the property of MCS, we have $\mathrm{U}^{\mathrm{M}}\varphi \in \Delta$, so $\langle l'(\varphi), \varphi\rangle = \langle l(\varphi), \varphi\rangle \in L'$.

Secondly, $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$ holds. We just need to check the following conditions:

- Since $\Delta = \Gamma$, obviously we have $\{\varphi \mid \mathrm{K}\varphi \in \Gamma\} \subseteq \Delta$.
- Since $\Delta = \Gamma$, it is clear that $dom(f) = dom(f')$, $dom(g) = dom(g')$, $dom(h) = dom(h)$ and $dom(l) = dom(l')$. Then for any $\varphi \in \{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Delta\}$, by definition of f', we have $f(\varphi) = f'(\varphi)$. Similarly, for any $\varphi \in \{\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Delta\}$, we have $g(\varphi) = g'(\varphi)$; for any $\varphi \in \{\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Delta\}$, we have $h(\varphi) = h'(\varphi)$; for any $\varphi \in \{\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Delta\}$, we have $l(\varphi) = l'(\varphi)$. Hence $f = f'$, $g = g'$,$h = h'$ and $l = l'$. □

Similarly, to refute $\mathrm{U}^{\mathrm{E}}\varphi$ while preserving $\mathrm{U}^{\mathrm{M}}\varphi$, we construct an accessible world where any second-level explanation for φ differs from that in the current world.

Lemma 3 (U^{E}-Existence Lemma). *For any $\langle \Gamma, F, G, H, L, f, g, h, l\rangle \in W^c$ where $\mathrm{U}^{\mathrm{M}}\psi \in \Gamma$, if $\mathrm{U}^{\mathrm{E}}\psi \notin \Gamma$, then for any $\langle s, t, \psi\rangle \in G$, there exists a $\langle \Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ such that $\langle s, t, \psi\rangle \notin G'$ and $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$.*

Proof. Suppose $\mathrm{U^E}\psi \notin \Gamma$ we construct $\langle \Delta, F', G', H', L', f', g', h', l' \rangle$ as follows:

- $\Delta = \Gamma$
- $F' = \{\langle s, \varphi \rangle \mid \langle s, \varphi \rangle \in F$ and $\mathrm{U^M}\varphi \in \Gamma\}$
- $G' = \{\langle s_2, s_1, \varphi \rangle \mid \langle s_2, s_1, \varphi \rangle \in G$ and $\mathrm{U^E}\varphi \in \Gamma\}$
- $H' = \{\langle s_3, s_2, s_1, \varphi \rangle \mid \langle s_3, s_2, s_1, \varphi \rangle \in H$ and $\mathrm{U^E}\varphi \in \Gamma\}$
- $L' = \{\langle c, s_3, s_2, s_1, \varphi \rangle \mid \langle c, s_3, s_2, s_1, \varphi \rangle \in L$ and $\mathrm{U^E}\varphi \in \Gamma\}$
- $f' : \{\varphi \mid \mathrm{U^M}\varphi \in \Delta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathrm{U^E}\varphi \in \Delta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{\varphi \mid \mathrm{U^D}\varphi \in \Delta\} \to E^c$ is defined as: $h'(\varphi) = h(\varphi)$
- $l' : \{\varphi \mid \mathrm{U^I}\varphi \in \Delta\} \to E^c$ is defined as: $l'(\varphi) = l(\varphi)$

We omit the remaining proof since it is very similar to the corresponding part in the proof of Lemma 4 below. □

Lemma 4 ($\mathrm{U^D}$-Existence Lemma). *For any $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \in W^c$ where $\mathrm{U^E}\psi \in \Gamma$, if $\mathrm{U^D}\psi \notin \Gamma$, then for any $\langle r, s, t, \psi \rangle \in H$, there exists a $\langle \Delta, F', G', H', L', f', g', h', l' \rangle \in W^c$ such that $\langle r, s, t, \psi \rangle \notin H'$ and $\langle \Gamma, F, G, H, L, f, g, h, l \rangle R^c \langle \Delta, F', G', H', L', f', g', h', l' \rangle$.*

Proof. Suppose $\mathrm{U^D}\psi \notin \Gamma$. we construct $\langle \Delta, F', G', H', L', f', g', h', l' \rangle$ as follows:

- $\Delta = \Gamma$
- $F' = \{\langle s, \varphi \rangle \mid \langle s, \varphi \rangle \in F$ and $\mathrm{U^M}\varphi \in \Gamma\}$
- $G' = \{\langle s_2, s_1, \varphi \rangle \mid \langle s_2, s_1, \varphi \rangle \in G$ and $\mathrm{U^E}\varphi \in \Gamma\}$
- $H' = \{\langle s_3, s_2, s_1, \varphi \rangle \mid \langle s_3, s_2, s_1, \varphi \rangle \in H$ and $\mathrm{U^D}\varphi \in \Gamma\}$
- $L' = \{\langle c, s_3, s_2, s_1, \varphi \rangle \mid \langle c, s_3, s_2, s_1, \varphi \rangle \in L$ and $\mathrm{U^D}\varphi \in \Gamma\}$
- $f' : \{\varphi \mid \mathrm{U^M}\varphi \in \Delta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathrm{U^E}\varphi \in \Delta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{\varphi \mid \mathrm{U^D}\varphi \in \Delta\} \to E^c$ is defined as: $h'(\varphi) = h(\varphi)$
- $l' : \{\varphi \mid \mathrm{U^I}\varphi \in \Delta\} \to E^c$ is defined as: $l'(\varphi) = l(\varphi)$

Obviously, $F' \subseteq F, G' \subseteq G, H' \subseteq H, L' \subseteq L$. Again, the main idea behind the constructions is to "carefully" delete all third-level explanations for $\{\psi \mid \mathrm{U^D}\psi \notin \Gamma\}$. Clearly $\langle t_3, t_2, t_1, \psi \rangle \notin H'$ for any $\langle t_3, t_2, t_1, \psi \rangle \in H$ by the construction, given $\mathrm{U^D}\psi \notin \Gamma$. In order to complete this proof, firstly, we show that $\langle \Delta, F', G', H', L', f', g', h', l' \rangle \in W^c$ by verifying the conditions $1 - 16$ in the definition of W^c. Only selected cases are detailed below:

- For the condition 3, suppose $\langle t_3, t_2, t_1, \varphi \to \psi \rangle, \langle s_2, s_2, s_1, \varphi \rangle \in H' \subseteq H$, then $\langle t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \varphi \rangle \in H$. Moreover due to the axiom (DISTU) and the fact that $\mathrm{U^D}(\varphi \to \psi), \mathrm{U^D}\varphi \in \Gamma$, we have $\mathrm{U^D}\psi \in \Gamma$. Hence $\langle t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi \rangle \in H'$.
- For condition 9, suppose $\langle t_3, t_2, t_1, \varphi \rangle \in H' \subseteq H$, then $\langle t_2, t_1, \varphi \rangle \in G$ and $\mathrm{U^D}\varphi \in \Gamma$, and thus $\mathrm{U^E}\varphi \in \Gamma$ by (DYE), which imply $\langle t_2, t_1, \varphi \rangle \in G'$.

- For condition 16, suppose $\mathrm{U}^{\mathrm{I}}\varphi \in \Delta$. Then we get $\mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$ by $\Gamma = \Delta$, thus $\langle l(\varphi), \varphi\rangle \in L$. By (IYD) and the property of MCS, we have $\mathrm{U}^{\mathrm{D}}\varphi \in \Delta$, so $\langle l'(\varphi), \varphi\rangle = \langle l(\varphi), \varphi\rangle \in L'$.

Then it is straightforward to check that $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$ holds. □

Lemma 5 (U^{I}-Existence Lemma). *For any $\langle \Gamma, F, G, H, L, f, g, h, l\rangle \in W^c$ where $\mathrm{U}^{\mathrm{D}}\psi \in \Gamma$, if $\mathrm{U}^{\mathrm{I}}\psi \notin \Gamma$, then for any $\langle c, r, s, t, \psi\rangle \in H$, there exists a $\langle \Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ such that $\langle c, r, s, t, \psi\rangle \notin H'$ and $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$.*

Proof. If $\mathrm{U}^{\mathrm{I}}\psi \notin \Gamma$, then we construct $\langle \Delta, F', G', H', L', f', g', h', l'\rangle$ as follows:

- $\Delta = \Gamma$
- $F' = \{\langle s, \varphi\rangle \mid \langle s, \varphi\rangle \in F \text{ and } \mathrm{U}^{\mathrm{M}}\varphi \in \Gamma\}$
- $G' = \{\langle s_2, s_1, \varphi\rangle \mid \langle s_2, s_1, \varphi\rangle \in G \text{ and } \mathrm{U}^{\mathrm{E}}\varphi \in \Gamma\}$
- $H' = \{\langle s_3, s_2, s_1, \varphi\rangle \mid \langle s_3, s_2, s_1, \varphi\rangle \in H \text{ and } \mathrm{U}^{\mathrm{D}}\varphi \in \Gamma\}$
- $L' = \{\langle c, s_3, s_2, s_1, \varphi\rangle \mid \langle c, s_3, s_2, s_1, \varphi\rangle \in L \text{ and } \mathrm{U}^{\mathrm{I}}\varphi \in \Gamma\}$
- $f' : \{\varphi \mid \mathrm{U}^{\mathrm{M}}\varphi \in \Delta\} \to E^c$ is defined as: $f'(\varphi) = f(\varphi)$
- $g' : \{\varphi \mid \mathrm{U}^{\mathrm{E}}\varphi \in \Delta\} \to E^c$ is defined as: $g'(\varphi) = g(\varphi)$
- $h' : \{\varphi \mid \mathrm{U}^{\mathrm{D}}\varphi \in \Delta\} \to E^c$ is defined as: $h'(\varphi) = h(\varphi)$
- $l' : \{\varphi \mid \mathrm{U}^{\mathrm{I}}\varphi \in \Delta\} \to E^c$ is defined as: $l'(\varphi) = l(\varphi)$

We have $F' \subseteq F, G' \subseteq G, H' \subseteq H, L' \subseteq L$ by the construction. By "carefully" deleting all fourth-level explanations for $\{\psi \mid \mathrm{U}^{\mathrm{I}}\psi \notin \Gamma\}$, we have that $\langle c, t_3, t_2, t_1, \psi\rangle \notin L'$ for any $\langle c, t_3, t_2, t_1, \psi\rangle \in L$, given $\mathrm{U}^{\mathrm{I}}\psi \notin \Gamma$. To complete this proof, firstly, we show that $\langle \Delta, F', G', H', L', f', g', h', l'\rangle \in W^c$ and omit most cases below:

- For the condition 4, Suppose $\langle c, t_3, t_2, t_1, \varphi \to \psi\rangle, \langle c, s_3, s_2, s_1, \varphi\rangle \in L' \subseteq L$, then $\langle c, t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \varphi\rangle \in L$. Moreover due to the axiom (DISTU) and the fact that $\mathrm{U}^{\mathrm{I}}(\varphi \to \psi), \mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$, we have $\mathrm{U}^{\mathrm{I}}\psi \in \Gamma$. Hence $\langle c, t_3 \cdot s_3, t_2 \cdot s_2, t_1 \cdot s_1, \psi\rangle \in L'$.
- For condition 10, suppose $\langle c, t_3, t_2, t_1, \varphi\rangle \in L' \subseteq L$, then $\langle t_3, t_2, t_1, \varphi\rangle \in H$ and $\mathrm{U}^{\mathrm{I}}\varphi \in \Gamma$, and thus $\mathrm{U}^{\mathrm{D}}\varphi \in \Gamma$, which imply $\langle t_3, t_2, t_1, \varphi\rangle \in H'$.

Secondly, $\langle \Gamma, F, G, H, L, f, g, h, l\rangle R^c \langle \Delta, F', G', H', L', f', g', h', l'\rangle$ is clearly holds. □

Finally, it is time to prove the following:

Lemma 6 (Truth Lemma). *For any $\varphi \in$ **ELCU**, $\langle \Gamma, F, G, H, L, f, g, h, l\rangle \vDash \varphi$ iff $\varphi \in \Gamma$.*

Proof. The proof is by induction on the structure of φ. The atomic case and boolean cases are routine. For the case of $\varphi = \mathrm{K}\psi$, it is clear by Lemma 1. For the case of

$\varphi = \mathrm{U^M}\psi$, the proof is not hard with the help of Lemma 1 and 2. For the case of $\mathrm{U^E}\psi$, we omit the proof since it is very similar to the proofs of the cases below, by applying Lemma 3.

For the case of $\mathrm{U^D}\psi$,

- $\Longleftarrow$: Suppose $\mathrm{U^D}\psi \in \Gamma$. Then for any $\langle \Delta, F', G', H', L', f', g', h', l' \rangle$ such that $\langle \Gamma, F, G, H, L, f, g, h, l \rangle R^c \langle \Delta, F', G', H', L', f', g', h', l' \rangle$, we get $\mathrm{U^D}\psi \in \Delta$, which implies $\psi \in \Delta$ by (4*), (DYE) ,(EYM), (UYK), (T), and the property of MCS. Thus $\langle \Delta, F', G', H', L', f', g', h', l' \rangle \vDash \psi$ by IH. Furthermore, we have $\langle h(\psi), \psi \rangle \in H$, $\langle h'(\psi), \psi \rangle \in H'$ and $h = h'$ by the definition of $\mathcal{M}^c$, which means that there exist $h(\psi) = h'(\psi) \in E^{c3}$ and $\langle \Delta, F', G', H', L', f', g', h', l' \rangle \in \mathcal{E}^c(h'(\psi), \psi)$. Hence $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \vDash \mathrm{U^D}\psi$.
- $\Longrightarrow$: Suppose $\mathrm{U^D}\psi \notin \Gamma$. Then we have the following four cases:
  - $\mathrm{K}\psi \notin \Gamma$. By Lemma 1, we have $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{K}\psi$, thus $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^D}\psi$.
  - $\mathrm{K}\psi \in \Gamma$ but $\mathrm{U^M}\psi \notin \Gamma$. By Lemma 2, we have $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^M}\psi$, thus $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^D}\psi$.
  - $\mathrm{U^M}\psi \in \Gamma$ but $\mathrm{U^E}\psi \notin \Gamma$. By Lemma 3, we have $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^E}\psi$, thus $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^D}\psi$.
  - $\mathrm{U^E}\psi \in \Gamma$. If $\langle r, s, t, \psi \rangle \notin H$ for any $r, s, t \in E^c$, then by the semantics, $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^D}\psi$. If there exist t_1 , t_2 and $t_3$with $\langle t_3, t_2, t_1, \psi \rangle \in H$, then we complete it with the help of Lemma 4.

For the case of $\mathrm{U^I}\psi$,

- $\Longleftarrow$: It is similar to the case of $\mathrm{U^D}\psi$ above.
- $\Longrightarrow$: Suppose $\mathrm{U^I}\psi \notin \Gamma$. Then only one case is checked according to the proof for the case of $\mathrm{U^D}\psi$.
  - $\mathrm{U^D}\psi \in \Gamma$. If $\langle c, r, s, t, \psi \rangle \notin H$ for any $r, s, t \in E^c$, then clearly $\langle \Gamma, F, G, H, L, f, g, h, l \rangle \nvDash \mathrm{U^I}\psi$. If there exist t_1 , t_2, $t_3$with $\langle c, t_3, t_2, t_1, \psi \rangle \in H$, then it is done by Lemma 5. □

Theorem 2. *The system* SCU *is strongly complete over* **ELCU** *models.*

Proof. Given a SCU-consistent set Σ^-, it can be extended to a MCS $\Sigma \in \Omega$. Then there exist $F^\Sigma, G^\Sigma, H^\Sigma, L^\Sigma, f^\Sigma, g^\Sigma, h^\Sigma, l^\Sigma$ such that $\langle \Sigma, F^\Sigma, G^\Sigma, H^\Sigma, L^\Sigma, f^\Sigma, g^\Sigma, h^\Sigma, l^\Sigma \rangle \in W^c$ by Proposition 3. Due to the Truth Lemma 6, we have a canonical model $\mathcal{M}^c$ satisfying Σ and hence Σ^-. □

5 Discussions

In this section, we explore the applications of the current logical framework to some multi-agent scenarios.

5.1 Comparative understanding

Given the framework described above, one might be tempted to extend it to multi-agent scenarios, which would enable the expression of comparative statements about the understanding among different agents, e.g., Alice understands why the sky is blue better than Bob does. In such an extended framework, we could introduce new modalities to compare the depth of understanding between various agents.

Below, focusing on comparing different people's understanding, we define:

Definition 7 (Multi-agents epistemic language of comparative understanding)**.** Fix nonempty set P of propositional letters and nonempty set I of agent names, the language **MELCU** is defined as (where $p \in P$, and $i, j \in I$):

$$\varphi ::= \ p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \mathrm{K}_i\varphi \mid \mathrm{U}_i\varphi \mid \mathrm{U}_{i>j}\varphi$$

$\mathrm{U}_i\varphi$ is read as agent i understands why φ. The comparative understanding between two agents denoted by $\mathrm{U}_{i>j}\varphi$ indicates that agent i understands why φ better than agent j does.

Then modify the previous definition of **ELCU** model in Definition 2 to obtain a new multi-agent model for the the language **MELCU**.

Definition 8. An **MELCU** model $\mathcal{M}^*$ is a tuple $(W, E, \{R_i \mid i \in I\}, \{\mathcal{E}_i \mid i \in I\}, V)$ where:

- W is a non-empty set of possible worlds.
- E is a non-empty set of explanations equipped with operators $\cdot$, ! and c such that:
  1. If $t, s \in E$, then $t \cdot s \in E$,
  2. If $t \in E$, then $!t \in E$,
  3. A special symbol c is in E.
- $R_i \subseteq W \times W$ is an equivalence relation for each $i \in I$.
- $\mathcal{E}_i : E^n \times \textbf{MELCU} \to 2^W$ $(n \geqslant 1)$ is an admissible explanation function satisfying the following conditions:

  Explanation Application: $\mathcal{E}_i(\langle t_n, \ldots, t_1\rangle, \varphi \to \psi) \cap \mathcal{E}_i(\langle s_n, \ldots, s_1\rangle, \varphi) \subseteq \mathcal{E}_i(\langle t_n \cdot s_n, \ldots, t_1 \cdot s_1\rangle, \psi)$.

  Constant Specification: If $\varphi \in \Lambda$, then $\mathcal{E}_i(c, \varphi) = W$.

  Higher-level Explanation Factivity: $\mathcal{E}_i(\langle t_{n+1}, t_n, \ldots, t_1\rangle, \varphi) \subseteq \mathcal{E}_i(\langle t_n, \ldots, t_1\rangle, \varphi)$.

  Epistemic Introspection: $\mathcal{E}_i(t, \bigcirc_i\varphi) \subseteq \mathcal{E}_i(\langle !t, t\rangle, \bigcirc_i\varphi)$ for $\bigcirc = \mathrm{K}, \mathrm{U}$.
- $V : P \to 2^W$ is a valuation function.

Those conditions for ideal explanation are temporarily omitted. As discussed in Sect. 2, explanations for epistemic claims are treated as justifications. Therefore, in the model, we need an admissible explanation function $\mathcal{E}_i$ relative to the set of individuals I. To consider the *epistemic introspection* condition, it is more reasonable to say that w is a world where t is a justification of i for $\mathrm{K}_i\varphi$. Adapting admissible explanation function to be relative to I will also make the model more general, facilitating the discussion of meta-understanding statements, such as Alice understanding Bob's understanding. We will address this issue in the next subsection.

Below we omit the straightforward cases in the definition of the semantics.

Definition 9.

$\mathcal{M}^*, w \Vdash \mathrm{U}_i\varphi$	$\Leftrightarrow$	(1) there exist $t_1, \ldots, t_n \in E$ $(n \geqslant 2)$ such that for all $v \in W$ with wR_iv, $v \in \mathcal{E}_i(\langle t_n, \ldots, t_1\rangle, \varphi)$; (2) for all $v \in W$ with wR_iv, $\mathcal{M}^*, v \Vdash \varphi$.
$\mathcal{M}^*, w \Vdash \mathrm{U}_{i>j}\varphi$	$\Leftrightarrow$	(1) there are $t_1, \ldots, t_n, \ldots, t_m \in E$ $(m > n \geqslant 2)$ such that for all $v \in W$ with wR_iv, $v \in \mathcal{E}_i(\langle t_m, \ldots, t_1\rangle, \varphi)$, for all $u \in W$ with wR_ju, $u \in \mathcal{E}_j(\langle t_n, \ldots, t_1\rangle, \varphi)$ and there is a u' with wR_ju', $u' \notin \mathcal{E}_j(\langle t_m, \ldots, t_1\rangle, \varphi)$; (2) for all $v \in W$ with wR_iv or wR_jv, $\mathcal{M}^*, v \Vdash \varphi$.

While $\mathrm{U}_i\varphi$ involves grasping a demanding explanation, $\mathrm{U}_{i>j}\varphi$ says that agent i has a deeper explanation than what agent j possesses. The formula $\mathrm{U}_{i>j}\varphi$ is roughly $\exists t_1 \cdots \exists t_n \cdots \exists t_m(\mathrm{K}_i(t_m : \cdots : t_1 : \varphi) \wedge \mathrm{K}_j(t_n : \cdots : t_1 : \varphi) \wedge \neg\mathrm{K}_j(t_m : \cdots : t_1 : \varphi))$ $(m > n \geqslant 2)$. As mentioned earlier in Sect. 1.3, it is accepted in philosophical literature that the depth of explanation can influence levels of understanding. When i and j can both offer explanations for the φ-phenomenon, it indicates they have some level of understanding of why it occurs, though the degree of their understanding may vary. Accordingly, we have the following valid formulas concerning $\mathrm{U}_{i>j}\varphi$:

- $\vdash \neg\mathrm{U}_{i>i}\varphi$.
- $\vdash \mathrm{U}_{i>j}\varphi \rightarrow \mathrm{U}_i\varphi \wedge \mathrm{U}_j\varphi$.

However, the formula $\mathrm{U}_{i>j}\varphi \wedge \mathrm{U}_{j>i}\varphi$ can be satisfied in some pointed **MELCU** models, that is, there exist two alternative explanations $t_1, \ldots, t_n, \ldots, t_m \in E$ $(m > n \geqslant 2)$ and $s_1, \ldots, s_{n'}, \ldots, s_{m'} \in E$ $(m' > n' \geqslant 2)$ such that $\exists t_1 \cdots \exists t_n \cdots \exists t_m(\mathrm{K}_i (t_m : \cdots : t_1 : \varphi) \wedge \mathrm{K}_j(t_n : \cdots : t_1 : \varphi) \wedge \neg\mathrm{K}_j(t_m : \cdots : t_1 : \varphi))$ and $\exists s_1 \cdots \exists s_{n'} \cdots \exists s_{m'}(\mathrm{K}_j(s_{m'} : \cdots : s_1 : \varphi) \wedge \mathrm{K}_i(s_{n'} : \cdots : s_1 : \varphi) \wedge \neg\mathrm{K}_i(s_{m'} : \cdots : s_1 : \varphi))$ hold simultaneously.

The formula appears to be counterintuitive. When comparing different individuals' understanding, we typically refer to an objective standard, like scientific explanations. The problem is that the framework's comparative measure of explanatory power is limited to explanatory depth. Once we can further compare two alternative explanations, such as stating that one is closer to the correct scientific explanation than the other, we may get a total order among different explanations. In this context, given a φ-phenomenon to be explained,

- An explanation $t_n : \cdots : t_1 : \varphi$ is *explanatory stronger* than $s_m : \cdots : s_1 : \varphi$ iff either $t_n : \cdots : t_1 : \varphi$ is deeper than $s_m : \cdots : s_1 : \varphi$, or they are *alternative explanations* of φ and $t_n : \cdots : t_1 : \varphi$ is *more scientific* than $s_m : \cdots : s_1 : \varphi$.

Accordingly, the formula $\mathrm{U}_{i>j}\varphi$ will say that for every explanations j possessing for φ, there is an explanatory stronger explanation than it that i has.

5.2 Meta-understanding

When the framework is applied to multi-agent situations, it sparks more interesting discussions about understanding. Consider the formula $\mathrm{U}_i\mathrm{U}_j\varphi$, it is meant to reflect one agent's (i.e., i's) understanding of another agent j's understanding, which could be termed as *meta-understanding*. The $\mathrm{U}_i\mathrm{U}_j\varphi$ indicates i's meta-understanding from j's perspective.

What does it mean to say that, for instance, Alice has an understanding of Bob's understanding of why the sky is blue? Basically, it involves several things. First, Alice needs to have her own understanding of why the sky is blue. Second, she needs to understand Bob's explanation or perspective on why the sky is blue. This goes beyond simply knowing that Bob understands the phenomenon; it involves grasping the explanations that lead to his understanding. This requires a meta-cognitive layer where Alice reflects on Bob's reasoning process. Finally, Alice may also need to identify any differences between her understanding and Bob's, and understand why these differences might exist.

Based on the analysis of meta-understanding, the formula $\mathrm{K}_i\mathrm{U}_{i>j}\varphi \to \mathrm{U}_i\mathrm{U}_j\varphi$ should be valid, while the formula $\mathrm{K}_j\mathrm{U}_{i>j}\varphi \to \mathrm{U}_j\mathrm{U}_i\varphi$ should not. In order to embody such information in the model, we define $\mathcal{M}^{**}$ be the **MELCU*** model with meta-understanding, based on the above. It's similar to the previously defined $\mathcal{M}^*$ in Definition 8, just with an additional condition added to $\mathcal{E}_i$:

Meta-understanding: For any w, if $v \in \mathcal{E}_i(\langle t_m, \ldots, t_n, \ldots, t_1\rangle, \varphi)$ for each v with wR_iv and $u \in \mathcal{E}_j(\langle t_n, \ldots, t_1\rangle, \varphi)$ $(m > n \geqslant 2)$ for each u with wR_ju, then $w \in \mathcal{E}_i(\langle t_n, \ldots, t_1\rangle, \mathrm{U}_j\varphi)$.

That is, if in a world w, agent i grasps an explanation $\langle t_m, \ldots, t_1\rangle$ of φ, and j grasps a lower-level explanation $\langle t_n, \ldots, t_1\rangle$ of φ, then w is a world where $\langle t_n, \ldots, t_1\rangle$ is traceable by agent i to understand why j understands why φ. We do not use $\langle t_m, \ldots, t_1\rangle$ or $\langle t_m, \ldots, t_{n+1}\rangle$ as an explanation of $\mathrm{U}_j\varphi$, since $\langle t_m, \ldots, t_{n+1}\rangle$ may not be involved in j's perspective.

Hence, it is straightforward to check the following validity:

Proposition 7. $\mathrm{K}_i\mathrm{U}_{i>j}\varphi \to \mathrm{U}_i\mathrm{U}_j\varphi$ *is valid over* **MELCU*** *models.*

Proof. For each **MELCU*** model $\mathcal{M}^{**}$, suppose $\mathcal{M}^{**}, w \Vdash \mathrm{K}_i\mathrm{U}_{i>j}\varphi$. Then for all v with wR_iv, there exist $t_1, \ldots, t_n, \ldots, t_m \in E$ $(m > n \geqslant 2)$ such that for all

u with vR_iu, $u \in \mathcal{E}_i(\langle t_m, \ldots, t_1\rangle, \varphi)$, for all v' with vR_jv', $v' \in \mathcal{E}_j(\langle t_n, \ldots, t_1\rangle, \varphi)$ and there is a u' with vR_ju', $u' \notin \mathcal{E}_j(\langle t_m, \ldots, t_1\rangle, \varphi)$. By the *meta-understanding* condition of $\mathcal{E}_i$, $v \in \mathcal{E}_i(\langle t_n, \ldots, t_1\rangle, \mathrm{U}_j\varphi)$. Moreover, $\mathcal{M}^{**}, v \Vdash \mathrm{U}_j\varphi$, which implies $\mathcal{M}^{**}, w \Vdash \mathrm{U}_i\mathrm{U}_j\varphi$. □

We will leave the full logic over **MELCU** or **MELCU*** models to a future occasion.

6 Conclusions and Future Work

This paper explores the logical structure of varying degrees of understanding. We establish a partial order among different explanations within our models, which facilitates comparative understanding. Our framework spans a spectrum that includes minimal understanding, everyday understanding, demanding understanding, and ideal understanding. We achieve a complete axiomatization of our logic, and discuss its application in multi-agent systems.

Possible future studies include axiomatizations for multi-agent scenarios that incorporate more comparative measures of explanations and meta-understanding. Additionally, our framework prepares us for future extensions involving the dynamics of explanations, such as reasoning about multi-agent communication and explanation acquisition, as studied in [18]. Furthermore, it is controversial whether groups can properly be said to possess understanding, the interesting notion of *group understanding* discussed in [7] can be explored within this framework.

References

[1] S. Artemov and M. Fitting, 2019, *Justification Logic: Reasoning with Reasons*, Cambridge University Press.

[2] J. Avigad, 2008, “Understanding proofs”, in P. Mancosu (ed.), *The Philosophy of Mathematical Practice*, pp. 317–353, Oxford: Oxford University Press.

[3] C. Baumberger, 2014, “Types of understanding: their nature and their relation to knowledge”, *Conceptus*, **40(98)**: 67–88.

[4] C. Baumberger, C. Beisbart and G. Brun, 2017, “What is understanding? an overview of recent debates in epistemology and philosophy of science”, in S. G. C. Baumberger and S. Ammon (eds.), *Explaining Understanding: New Perspectives from Epistemolgy and Philosophy of Science*, pp. 1–34, Routledge.

[5] I. Boh, 1993, *Epistemic Logic in the Later Middle Ages*, Routledge.

[6] I. Boh, 2000, “Four phases of medieval epistemic logic”, *Theoria*, **66(2)**: 129–144.

[7] K. Boyd, 2019, “Group understanding”, *Synthese*, **198(7)**: 6837–6858.

[8] P. M. Dung, 1995, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games", *Artificial Intelligence*, **77(2)**: 321–357.

[9] M. Fitting, 2004, "A logic of explicit knowledge", *Logica Yearbook*: 11–22.

[10] E. C. Gordon, 2012, "Is there propositional understanding?", *Logos & Episteme*, **3(2)**: 181–192.

[11] C. Hempel, 1965, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, The Free Press.

[12] C. G. Hempel and P. Oppenheim, 1948, "Studies in the logic of explanation", *Philosophy of Science*, **15(2)**: 135–175.

[13] K. Khalifa, 2013, "The role of explanation in understanding", *The British Journal for the Philosophy of Science*, **64(1)**: 161–187.

[14] K. Khalifa, 2017, *Understanding, Explanation, and Scientific Knowledge*, Cambridge, UK: Cambridge University Press.

[15] R. Kuznets and T. Studer, 2019, *Logics of proofs and justifications*, College Publications.

[16] I. Lawler, 2019, "Understanding why, knowing why, and cognitive achievements", *Synthese*, **196(11)**: 4583–4603.

[17] P. Lipton, 2009, "Understanding without explanation", in H. W. de Regt, S. Leonelli and K. Eigner (eds.), *Scientific Understanding: Philosophical Perspectives*, pp. 43–63, University of Pittsburgh Press.

[18] J. Luo, T. Studer and M. Dastani, 2023, "Providing personalized explanations: a conversational approach", in A. Herzig, J. Luo and P. Pardo (eds.), *Logic and Argumentation*, pp. 121–137, Cham: Springer Nature Switzerland.

[19] D. Pritchard, 2008, "Knowing the answer, understanding and epistemic value", *Grazer Philosophische Studien*, **77(1)**: 325–339.

[20] D. Pritchard, 2014, "Knowledge and understanding", *Virtue Epistemology Naturalized*, pp. 315–327, Springer.

[21] P. Railton, 1981, "Probability, explanation, and information", *Synthese*, **48(2)**: 233–256.

[22] L. D. Ross, 2020, "Is understanding reducible?", *Inquiry*, **63(2)**: 117–135.

[23] W. C. Salmon, 1985, *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

[24] I. Sedlár and J. Halas, 2015, "Modal logics of abstract explanation frameworks", *Abstract in proceedings of CLMPS 15*.

[25] D. Šešelja and C. Straßer, 2013, "Abstract argumentation and explanation applied to scientific debates", *Synthese*, **190(12)**: 2195–2217.

[26] P. Sliwa, 2015, "Iv—understanding and knowing", *Proceedings of the Aristotelian Society*, **Vol. 115**, pp. 57–74.

[27] M. Strevens, 2013, "No understanding without explanation", *Studies in History and Philosophy of Science Part A*, **44(3)**: 510–515.

[28] P. Thagard, 2007, "Coherence, truth, and the development of scientific knowledge", *Philosophy of Science*, **74(1)**: 28–47.

[29] Y. Wei, 2024, "A logical framework for understanding why", in A. Pavlova, M. Y. Pedersen and R. Bernardi (eds.), *Selected Reflections in Language, Logic, and Information*, pp. 203–220, Cham: Springer Nature Switzerland.

[30] D. A. Wilkenfeld, 2014, "Functional explaining: a new approach to the philosophy of explanation", *Synthese*, **191(14)**: 3367–3391.

[31] J. Woodward and L. Ross, 2021, "Scientific Explanation", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University.

[32] J. F. Woodward, 2003, *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press.

[33] C. Xu, Y. Wang and T. Studer, 2021, "A logic of knowing why", *Synthese*, **198(2)**: 1259–1285.

理解与解释：一种知识逻辑的视角

魏宇

摘　要

知识逻辑学家很少关注“理解”这一概念，这与科学哲学和知识论领域的研究现状形成了鲜明对比。本文提出了一种类似知识逻辑的框架来刻画“理解”。由于“解释”帮助理解，该模型包含不同程度的解释概念，并在这些解释之间建立起一种偏序关系。受哲学讨论的启发，本文在语形上包含了一系列理解的模态，从最低限度的理解到日常的理解、高要求的理解、以及理想的理解。文中给出了一个可靠完全的公理系统刻画这些理解概念，并讨论了这样的逻辑在多主体情境中的应用，如探讨不同主体之间的理解比较以及主体之间的元理解等。

魏宇　华东师范大学哲学系
ywei@philo.ecnu.edu.cn