A Hierarchical Characterization of Ignorance in Epistemic Logic

Alessandro Aldini, Pierluigi Graziani, Mirko Tagliaferri

University of Urbino Carlo Bo

Abstract

We study different forms of ignorance and their correlations in a bi-modal logical language expressing the two modalities of knowledge and belief. In particular, we are mainly interested in clarifying which definitions of ignorance and which circumstances trigger higher-order forms of ignorance, inducing ignorance about ignorance and so on. To this aim, three ground conditions concerning knowledge and belief are presented, which may be seen as a cause of ignorance and can help us to identify the conditions enabling the emergence of higher-order forms of ignorance.

Keywords: Ignorance, Epistemic Logic, Doxastic Logic.

1. Introduction

In 1962 Jakko Hintikka published a very influential book for discussions about knowledge (and its lack). The title of this seminal book is *Knowledge and beliefs: An introduction to the logic of the two notions* (Hintikka 1962). In this work, Hintikka provides a propositional axiomatization of the modal operators *K* (for knowledge) and *B* (for belief), as well as insights on the meaning of various forms of knowledge and lack of knowledge, like, e.g., *not knowing whether* ϕ (Hintikka 1962: 3), which is represented by Hintikka as $\sim K\phi \wedge \sim K\sim\phi$ (Hintikka 1962: 12).[1] This form of lack of knowledge, which we refer as *ignorance whether* and will denote in the following with $I(\phi)$, has indeed been widely investigated in several logical frameworks focused on ignorance, see, e.g., Fan et al. 2015, Steinsvold 2008, and van der Hoek and Lomuscio 2004.

Such a representation of ignorance might be considered stronger than another classical definition, referred as *unknown truth* (Fitch 1963). In this alternative view, being ignorant of ϕ simply stands for *not knowing that* ϕ, which is formally

[1] We use the symbol $\wedge$ for propositional conjunction, $\sim$ for propositional negation, $\vee$ for propositional disjunction, and $\rightarrow$ for propositional implication.

Argumenta 8, 2 (2023): 373-386
ISSN 2465-2334
DOI 10.14275/2465-2334/202316.ald

expressed as $\phi \wedge \sim K\phi$. We will refer to it as *ignorance of the fact* (or simply, *factive ignorance*), denoted $I_f(\phi)$, in order to emphasize that, with respect to $I(\phi)$, it represents a factive form of ignorance, satisfying the factivity axiom $I_f(\phi) \rightarrow \phi$, see, e.g., Kubyshkina and Petrolo 2021.

Effects of the (lack of) knowledge of ignorance are investigated both by Fitch himself concerning $I_f(\phi)$, and recently in the setting of $I(\phi)$ in Fine 2018. The two forms of ignorance are combined together in a disjunctive form of ignorance in Fan 2021, and are denoted by the symbol $\nabla(\phi)$, with the aim of studying their correlations. Other logical investigations about ignorance are provided in Fano and Graziani 2021, Goranko 2021, Halpern 1997, and Meyer and van der Hoek 1995.

With respect to the scenario surveyed above, Fine's work—very recently taken up by Fan and in Aldini et al. 2021—provides an exemplary taxonomy of various logical forms of ignorance expressed in terms of the *K* operator, of which we summarize the main relationships in Fig. 1, commented as follows. Basically, in the figure we can distinguish between first-order forms and higher-order forms of ignorance. In this paper, ignorance will always be indicated with a specific order, which denotes the depth of the ignoring phenomenon. First-order ignorance refers to the basic situation of being ignorant of something; second-order ignorance is about being ignorant of being ignorant, and so forth. As mentioned above, Fan's ignorance $\nabla(\phi)$ is the disjunction of $I(\phi)$ and $I_f(\phi)$, which are the two most prominent first-order definitions of Fig. 1. *Factive ignorance* triggers a loop of ignorance, as suggested by $I_f(\phi) \rightarrow \sim K(I_f(\phi))$, which is a result obtained in Fitch 1963. The main result given in Fine 2018 concerns the properties of second-order ignorance in the case of $I(\phi)$, which induces both first-order ignorance and third-order ignorance, thus triggering the loop towards higher orders. Moreover, Fine shows an alternative characterization of second-order *ignorance whether*, called *Rumsfeld ignorance*, which implies lack of knowledge on someone's ignorance whether ϕ, $I_R(\phi) := I(\phi) \wedge \sim K(I(\phi))$. This characterization is analogous to that of *factive ignorance* and captures the laymen interpretation of second-order ignorance. *Rumsfeld ignorance* states that someone is ignorant whether a specific fact holds without knowing to be so. As we will show later in the paper, given a suitable logical framework, Rumsfeld ignorance and second-order ignorance are equivalent.

An important achievement of our study is revisiting the forms of ignorance of Fig. 1 in a unique logical framework, by adding two novel contributions to such a picture:

(1) the study of the relation between $I(\phi)$ and $I_f(\phi)$ and their common roots;
(2) the study of conditions that trigger/block the loop of ignorance starting from $I(\phi)$.

The first point is important especially if we are able to identify ground conditions at the base of the notion of ignorance. The second point is particularly interesting as both *factive ignorance* and *Rumsfeld ignorance* enable higher orders of ignorance, while $I(\phi)$ does not, in principle. Moreover, Fine (2018) proved that once second-order ignorance is present, all higher-order levels of ignorance collapse, generating a so-called black hole of ignorance.

The formal framework we employ for our studies is a bi-modal logical language including the operators *K* and *B*. In this setting, we will propose three ground conditions on own belief and knowledge, which are related to $I(\phi)$ and

$I_f(\phi)$ and will help us to interpret the relation between the two forms of ignorance, thus achieving our first goal. On the other hand, these conditions will allow us to characterize also the passage from first-order *ignorance whether* to second-order *ignorance whether*, thus achieving our second goal.

The structure of the paper is organized as follows. In Section 2, an introduction to the logic of knowledge and belief is provided in which we specify the axioms and rules that we employ in the rest of the paper. In Section 3, we present the three ground conditions that are closely connected to *factive ignorance* and *ignorance whether*. Such conditions will help us to provide insights about the correlation between the two forms of ignorance and results about possible ways ignorance can emerge. In Section 4, we investigate the relationship between the various orders of ignorance and how they can emerge and propagate. Finally, in Section 5 concluding remarks follow.

This paper is a revised version and extends a previous paper by Aldini et al. 2021, which focused only on *ignorance whether*. Additional material concerns the discussion on the ignorance taxonomy and on I_f, and novel technical results about the three ground conditions (Section 3), $I_f(\phi)$ (Section 3.2) and its relation with $I(\phi)$ (Section 3.3), and the hierarchies of higher-orders of ignorance (Section 4.4).

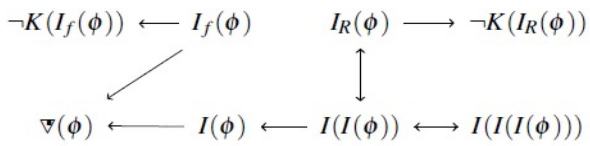

Figure 1: A taxonomy of ignorance

The proofs of the new results are in the Appendix.

2. Logic for Knowledge and Belief

The notions of knowledge and belief that we are going to formalize are interpreted in the language of propositional modal logic. We will give our presentation by abstracting as much as possible from the aspects related to semantics. Such an abstraction will allow us to generalize our results, which, as a side effect, will be easily understood by various communities interested in the topic of ignorance.

Definition 1.

Given a countable set *At* of atomic propositions, the bi-modal logical language $\mathfrak{L}$ is defined by the set of all formulas generated by the grammar:

$$\phi := p \mid \sim\phi \mid \phi \wedge \phi \mid K(\phi) \mid B(\phi) \text{ with p in } At$$

All the other Boolean connectives are defined in the standard way. The main modalities of the language are K and B, where the *knowledge formula* $K(\phi)$ should be read as "ϕ is known" and the *belief formula* $B(\phi)$ should be read as "ϕ is believed". *Ignorance whether* (denoted by I) is defined as follows: $I(\phi) := \sim K(\phi) \wedge \sim K(\sim\phi)$. *Factive ignorance* (denoted by I_f) is defined as follows: $I_f(\phi) := \phi \wedge \sim K(\phi)$. When clear from the context, $I(\phi)$ and $I_f(\phi)$ will be simply called *ignorance formulas*.

Moreover, it will be said that a formula ϕ is a *first-order ignorance formula*, if there is only one ignorance operator applied to it, the simplest case being, e.g., I(p). Analogously, a formula ϕ is a *second-order ignorance formula*, if there are at least, and no more than, two nested instances of ignorance operators applied to ϕ. A simple case is, e.g., $I(I$(p)). Higher-order instances of ignorance follow similarly.

Some classical properties of the two notions of knowledge and belief will be assumed:

Definition 2.

The axioms of knowledge in $\mathfrak{L}$ are:

K: $K(\phi\rightarrow\psi)\rightarrow(K(\phi)\rightarrow K(\psi))$;

T: $K(\phi)\rightarrow\phi$;

4: $K(\phi)\rightarrow K(K(\phi))$.

The axioms of belief in $\mathfrak{L}$ are:

B: $B(\phi\rightarrow\psi)\rightarrow(B(\phi)\rightarrow B(\psi))$;

D: $\sim(B(\phi)\wedge B(\sim\phi))$.

The interaction axiom of knowledge and belief in $\mathfrak{L}$ is:

Int$_1$: $K(\phi)\rightarrow B(\phi)$.

As far as the operator K is concerned, we first assume that it distributes over implications (logical consequence), usually termed axiom **K**. Then, we assume also that knowledge is factual (termed axiom **T**) and positively introspective (termed axiom **4**). The factuality (or truthfulness) of knowledge is straightforward: this comes mainly from philosophical reflections on the notion of knowledge, which is taken to be a rigorous cognitive phenomenon strongly tied with truth, i.e., only true things might be known. In fact, the strength of this axiom is what distinguishes proper knowledge from simple belief. Indeed, belief might be false, but knowledge never is. The positive introspection (knowing something is known) comes from the assumption that agents have a privileged access to their cognitive states. Note that assuming those properties defines knowledge in an analogous way to the box operator in **S4** systems of modal logic (Rendsvig and Symons 2019).

For the notion of belief, the operator B distributes over implications (termed axiom **B**) as in the case of knowledge. In addition, we also assume that beliefs are consistent (termed axiom **D**). Consistency of belief means that someone cannot believe that a fact is both true and false at the same time. Note that assuming the closure under logical consequence and consistency defines belief in an analogous way to the box operator in **KD** systems of modal logic (*ibid.*).

Finally, it is assumed that the two notions interact as follows: knowledge implies belief (termed axiom **Int$_1$**). This interaction axiom is commonly derived directly from the analysis of knowledge given in Plato's *Theatetus*: in such an analysis, knowledge is taken to be justified true belief. Unfortunately, the justification component is often neglected in formal frameworks, even though some attempts have been made to insert it, see, e.g., Artemov and Fitting 2020. The truth component is formalized through axiom **T**, while the belief component is given exactly by the interaction axiom **Int$_1$**. Such interaction axiom is studied by various authors—see, e.g., Meyer and van der Hoek 1995—and by itself, it is known that it does not cause any consistency issues. Since this is the only interaction axiom we assume in our setting, we are guaranteed that the system we are working with is indeed consistent.

These assumptions mean that our bi-modal logic is composed of an **S4** system for knowledge, plus a **KD** system for belief and the only interaction axiom $\textbf{Int}_1$.

In addition to the axioms discussed above, we consider also all classical propositional and modal deduction rules that are assumed in the systems we rely both on Meyer and van der Hoek 1995 and van Ditmarsch et al. 2015.

Another property that will be employed to study the effects of ignorance, but will not be assumed as an axiom, is that of *negative introspection*, often known as axiom **5** of epistemic logic. Negative introspection is similar in spirit to positive introspection: both axioms attribute to the agents a form of transparency towards their cognition. As said above, positive introspection allows an agent to know everything s/he knows; on the other hand, negative introspection states that an agent always knows what s/he does not know, i.e., $\sim K(\phi) \rightarrow K(\sim K(\phi))$. This property, while often assumed in epistemic logics employed in computer science (see, e.g., Halpern et al. 1995), might be too demanding, since it would imply that the agent is aware of all the facts s/he does not know.

Later on, in Section 4, we will show that assuming the presence (or lack) of negative introspection has a deep impact on the hierarchies of ignorance.

Now that all the formal details have been given, it is possible to move on to the reflections concerning the interplay among knowledge, belief, and ignorance.

3. Misbelieving, Doubting, and Being Agnostic

Understanding the origins of ignorance is not easy. The main issue is that ignorance is a *negative fact*, i.e., it is a lack of knowledge, and, therefore, it is difficult to identify a specific moment in time when ignorance is generated; it is either there the whole time or it is generated by an act of forgetting something. Therefore, this difficulty in identifying a precise moment in which ignorance is produced, makes it hard for researchers to focus on specific acts or behaviors that can improve our understanding of the phenomenon. For this reason, the formal research on the notion of ignorance helps to understand what are the constituents of such notion and thus which other phenomena are responsible for its emergence and/or existence. Specifically, three different, alternative conditions[2] will be explored (see, also, Aldini et al. 2021): *misbelieving*, *doubting*, and *agnosticism*.

Intuitively, we say that an agent is subject to *misbelieving* with respect to ϕ whenever the agent believes that ϕ holds, while such fact does not actually hold.[3]

Definition 3.

Misbelieving (denoted by M) is defined as:

$$M(\phi) := B(\phi) \wedge \sim\phi$$

As stated by the following proposition, the agent cannot know that s/he is subject to misbelieving with respect to ϕ ($\vdash_{\mathfrak{L}}$ expresses that a formula holds in $\mathfrak{L}$).

[2] With the term "condition" we refer to a cognitive attitude an agent might be subject to. Those will be equivalent to logical formulas that explicitly indicate what is the state of affair for the agent described.

[3] See Fano and Graziani 2021 for a discussion about different aspects that relate misbelieving and ignoring.

Proposition 1.

$\vdash_{\mathfrak{L}} \sim K(M(\phi))$.

Intuitively, we say that an agent is subject to *doubting* with respect to ϕ whenever the agent believes ϕ, which actually holds, but s/he does not know that such fact truly holds, thus raising the doubt.

Definition 4.

Doubting (denoted by D) is defined as:

$$D(\phi) := B(\phi)\wedge\phi\wedge\sim K(\phi)^{4}$$

As stated by the following proposition, the agent cannot know that s/he is subject to doubting with respect to ϕ.

Proposition 2.

$\vdash_{\mathfrak{L}} \sim K(D(\phi))$.

As shown above, in general an agent is not in a position to know if s/he is misbelieving or s/he is doubting. Interestingly, if the agent is aware of the possessed belief about ϕ and of the lack of knowledge about ϕ, then s/he knows to be subject to either misbelieving or doubting, even though s/he cannot say which one. Formally, the hypothesis above, called *Socratic*, is expressed by the formula:

$$S(\phi) := K(B(\phi)) \wedge K(\sim K(\phi)).$$

Then, the intuition above is formalized by the following result.

Proposition 3.

$\vdash_{\mathfrak{L}} S(\phi)\rightarrow K(M(\phi)\vee D(\phi))$

Finally, intuitively, we say that an agent is subject to *agnosticism* with respect to ϕ whenever the agent neither believes that ϕ is true nor believes that ϕ is false.

Definition 5.

Agnosticism (denoted by A) is defined as:

$A(\phi) := \sim B(\phi)\wedge\sim B(\sim\phi)$

Notice that $A(\sim\phi) := \mathrm{A}(\phi)$ by definition, and that, differently from the previous conditions, agnosticism does not assume anything about the truth of ϕ. Moreover, being agnostic is not in contradiction with the rule of necessitation.

In fact, by this rule and axiom $\mathbf{Int}_1$, the agent believes the tautology stating that ϕ is either true or false. Simply put, the agent does not have any opinion leading to prefer one case over the other.

3.1 Relating *Ignorance Whether* and the Three Conditions

Quite reasonably from an intuitive standpoint, each of the three conditions imply first-order *ignorance whether* (for the complete proofs, see Aldini et al. 2021).

The formal statements referring to those facts are the following.

Proposition 4.

$\vdash_{\mathfrak{L}} M(\phi)\rightarrow I(\phi)$ and $\vdash_{\mathfrak{L}} M(\sim\phi)\rightarrow I(\phi)$

[4] Note that this is equivalent to stating that the agent believes ϕ, but is factively ignorant that ϕ, i.e., $D(\phi) := B(\phi)\wedge I_f(\phi)$. Moreover, it should be noted that following the traditional definition of knowledge as justified true belief, someone who is doubting would be equivalent to someone who has a true belief but lacks justification. For a full exploration of this connection, see Tagliaferri 2023.

Proposition 5.

$\vdash_{\mathfrak{L}} D(\phi)\rightarrow I(\phi)$ and $\vdash_{\mathfrak{L}} D(\sim\phi)\rightarrow I(\phi)$

Proposition 6.

$\vdash_{\mathfrak{L}} A(\phi)\rightarrow I(\phi)$

Interestingly, all those propositions also hold in systems weaker than the one presented in this paper. This is due to the fact that for their proofs, some of the axioms assumed in this paper are not necessary. In particular, for Proposition 4, axiom **4** and axiom **B** are not needed. For Proposition 5, only axiom **T** is required. Finally, for Proposition 6, only $\mathbf{Int}_1$ is required.

In addition to those proofs, it is also possible to prove that *ignorance whether* implies a disjunction of the three conditions, thus connecting the phenomena.

Theorem 1.

$\vdash_{\mathfrak{L}} I(\phi)\rightarrow(M(\phi)\vee M(\sim\phi)\vee D(\phi)\vee D(\sim\phi)\vee A(\phi))$

3.2 Relating *Factive Ignorance* and the Three Conditions

Similarly as in the case of $I(\phi)$ we have that both misbelieving on $\sim\phi$ and doubting on ϕ imply being factively ignorant of ϕ.

Proposition 7.

$\vdash_{\mathfrak{L}} M(\sim\phi)\rightarrow I_f(\phi)$ and $D(\phi)\rightarrow I_f(\phi)$

Notice that we have to take care of the factivity of $I_f(\phi)$, so that the two remaining cases $M(\phi)$ and $D(\sim\phi)$ turn out to be related to $I_f(\sim\phi)$ by an analogous proposition.

Unfortunately, being agnostic is not sufficient to establish *factive ignorance* of ϕ, but it is sufficient to state that the agent is factively ignorant of either ϕ or $\sim\phi$.

Proposition 8.

$\vdash_{\mathfrak{L}} A(\phi)\rightarrow(I_f(\phi)\vee I_f(\sim\phi))$

As in the case of $I(\phi)$, *factive ignorance* implies the same disjunction combining the three ground conditions stated in Theorem 1.[5] As a consequence, by the factivity of $I_f(\phi)$, the disjunction of Theorem 1 can be simplified to state that $I_f(\phi)\rightarrow M(\sim\phi)\vee D(\phi)\vee A(\phi)$, since both $M(\phi)$ and $D(\sim\phi)$ contradict the hypothesis of the truth of ϕ.

3.3 Relating *Ignorance Whether* and *Factive Ignorance*

In Fan 2021, *ignorance whether* and *factive ignorance* are considered to be two first-order forms of ignorance of which it is interesting to study a weak combination through disjunction.

Formally, in our language Fan (2021)'s operator is expressed as follows:

Definition 6.

Disjunctive ignorance (denoted by ∇) is defined as:

$\nabla(\phi) := I_f(\phi)\vee I(\phi)$

[5] We postpone the proof as a corollary of a result demonstrated in the next section stating that $I_f(\phi)$ implies $I(\phi)$.

Such an operator satisfies interesting properties, especially related to positive and negative introspection. Here, however, we take a step back and show that I_f and I are actually tied together. In fact, the former implies the latter.

Theorem 2.

$\vdash_{\mathfrak{L}} I_f(\phi)\rightarrow I(\phi)$

As we mentioned, differently from $I(\phi)$, we have that $I_f(\phi)$ trivially satisfies the factivity axiom $I_f(\phi)\rightarrow\phi$. So, enforcing also the factivity of $I(\phi)$, which thus would be immediately reduced to $I_f(\phi)$, is a trivial way of collapsing $I(\phi)$ and $I_f(\phi)$. This is also emphasized by a result shown in the proof of Proposition 8.

Finally, as a corollary of Theorem 1 and Theorem 2, we have that $I_f(\phi)$ implies the disjunction of the three ground conditions expressed in Theorem 1.

4. Hierarchies of Ignorance

The taxonomy reported in Figure 1 emphasizes that the passage from first-order ignorance to second-order ignorance is critical. In fact, the peculiarity of second-order ignorance suggests that such a lack of knowledge is not mitigated by moving toward higher-orders of ignorance. Even more interestingly, under some forms of ignorance, this black-hole of higher-order levels of ignorance starts from the very beginning of first-order ignorance.

In the following, we recall some results concerning the hierarchies of ignorance and we fill the gap concerning the conditions that trigger/block the passage from first-order ignorance to second-order ignorance in the case of the operator I.

4.1 The Loop of *Factive Ignorance*

In Fitch 1963, it is demonstrated that $\sim K(I_f(\phi))$ holds. Here we recast this result and its consequences in our formal system, by showing that $I_f(\phi)\rightarrow I_f(I_f(\phi))$.[6] Notice that the major requirements of Fitch's proof are that the K operator is closed under conjunction elimination and that factivity holds (all requirements that are satisfied in **S4**, which is the system we are assuming for knowledge in this paper.

Theorem 3.

$\vdash_{\mathfrak{L}} I_f(\phi)\rightarrow I_f(I_f(\phi))$

We observe again that, by definition, we have that $I_f(I_f(\phi))$ implies $I_f(\phi)$. Hence, we conclude that all the orders of factive ignorance collapse and that factive ignorance, since the first-order, triggers a loop without remedy.

4.2 The Loop of Second-Order *Ignorance Whether*

Some relations concerning $I(I(\phi))$ have been well explored in Fine 2018. In particular, Fine shows that second-order ignorance and higher-orders of ignorance are tightly tied together in **S4** systems for knowledge. Once second-order ignorance is present, an agent is doomed to a collapse of higher-orders of ignorance, i.e., from second-order ignorance on, all order would become equivalent.

To prove such a result, Fine presents an alternative characterization of second-order ignorance, termed *Rumsfeld ignorance*. Intuitively, an agent is Rumsfeld ignorant when s/he is first-order ignorant of ϕ and does not know it.

[6] An elegant proof of this result is given in Fan 2022 that relies on neighborhood semantics.

Definition 7.
Rumsfeld ignorance (denoted by I_R) is defined as follows:
$I_R(\phi) := I(\phi) \wedge \sim K(I(\phi))$
where $I(\phi)$ is a first-order ignorance formula.

We point out that *Rumsfeld ignorance* can be represented by combining I_f and I, because by definition we have that $I_R(\phi) := I_f(I(\phi))$.

Then, we can rephrase the finding of Fine 2018 in our formal system as follows:

- from second-order to first-order: $\vdash_{\mathfrak{L}} I(I(\phi)) \rightarrow I(\phi)$.
- from second-order to *Rumsfeld*: $\vdash_{\mathfrak{L}} I(I(\phi)) \rightarrow I_R(\phi)$.
- from *Rumsfeld* to second-order: $\vdash_{\mathfrak{L}} I_R(\phi) \rightarrow I(I(\phi))$.
- lack of knowledge for *Rumsfeld*: $\vdash_{\mathfrak{L}} \sim K(I_R(\phi))$.

From these results, the lack of knowledge for second-order ignorance immediately derives, i.e., $\vdash_{\mathfrak{L}} I(I(\phi)) \rightarrow \sim K(I(I(\phi)))$, and, finally, the following theorem can be demonstrated.

Theorem 4.
In **S4** second-order ignorance implies higher-orders of ignorance. Specifically, second-order ignorance implies third-order ignorance. Third-order ignorance implies fourth-order ignorance and so forth.
Formally, $\vdash_{\mathfrak{L}} I^n(\phi) \rightarrow I^{n+1}(\phi)$, with $n \geq 2$

What Theorem 4 shows is that there is a deep connection between second-order ignorance and higher-order levels of ignorance. In fact, as soon as an agent is second-order ignorant, there is no possibility that s/he escapes the collapse of the various orders of ignorance on her/his own. Hence, it is evident why deep investigations on the relation between first-order ignorance and second-order ignorance are required. Once it is established what causes second-order ignorance in the presence of first-order ignorance, it might be possible to stop agents from having their levels of ignorance collapse one onto the other.

4.3 Relating Introspection and Hierarchies of *Ignorance Whether*

As shown in previous sections, once an agent is second-order ignorant, s/he is also subject to higher-orders of ignorance, which would cause the agent to not even try to make the appropriate steps to eliminate her/his ignorance. Moreover, it is common to have situations in which agents are (first-order) ignorant about specific facts, i.e., we cannot expect everybody to know everything. This situation suggests that it is important to avoid possible passages from first-order ignorance to second-order ignorance, so that the collapse of higher-order levels of ignorance is prevented, and agents are always in a position to recognize their ignorance and work towards solutions to it. Interestingly, a property of knowledge commonly studied in epistemic logic (i.e., negative introspection) is exactly what is needed to prevent this passage. Negative introspection is an incredibly powerful axiom that can block the passage from first-order ignorance to second-order ignorance.[7] This is exactly the meaning of the theorem that follows.[8]

Theorem 5.

$\vdash_{\mathfrak{L}+\mathbf{5}} \sim I(I(\phi))$.

It can therefore be safely claimed that negative introspection is an exceptionally effective measure to avoid the collapse of higher levels of ignorance.

Furthermore, it is possible to show that a direct negation of negative introspection is exactly what produces the passage from first-order ignorance to second-order ignorance, i.e., not having negative introspection is what causes the collapse of higher-orders of ignorance, as stated by the following theorem:

Theorem 6.

$\vdash_{\mathfrak{L}} I(\phi)\wedge(\sim K(\phi)\wedge\sim K(\sim K(\phi))) \rightarrow I(I(\phi))$.

The theorems above seem to show a clear picture: if we include the axiom of negative introspection to our system, we avoid higher-order levels of ignorance altogether; if we explicitly negate the axiom in our system, we are doomed to the collapse of all levels of ignorance to first-order ignorance. Those results seem to suggest that if we wish to avoid higher-order levels of ignorance we would simply need to include negative introspection as an axiom to our formal system. While this is true, the requirements placed on an agent in order to have negative introspection might be too demanding. We cannot expect all agents to be fully introspective and always know when they are ignoring something; while desirable and effective, this is unrealistic. Thus, what is needed, is something that avoids the passage from first-order ignorance to second-order ignorance, but that it is not too demanding to become unrealistic. In short, we need something that stays in the middle between having negative introspection and explicitly negating it. The next sections of this paper will explore potential candidates that could achieve this desirable result.

4.4 Relating Introspection and the Three Conditions

In the previous section, we have shown that negative introspection bridges basic and higher-order levels of ignorance in the case of $I(\phi)$. Hence, it is worth studying whether negative introspection can be expressed through more specific conditions tailored to the cases of misbelieving, doubting, and agnosticism. Along the same path followed for the study of first-order ignorance, we will identify conditions that, combined with the three conditions, imply negative introspection (resp., its negation), thus blocking (resp., triggering) the passage from first-order ignorance to second-order ignorance. Notice that when trying to prove that negative introspection follows from a given condition, we have to prove it generally, while, when trying to prove that the negation of negative introspection follows from a given condition, we only need to show an instance of such negation of negative

[7] This relationship between second-order ignorance and the axiom **5** of modal logic was already noticed by Fine 2018. In fact, Fine observes that Rumsfeld ignorance, which is tied with second-order ignorance, could be considered such a counter-example.

[8] See Aldini et al. 2021 for complete proofs of all theorems presented in this section. Notice, moreover, that since our formal system coincides with **S4** for the representation of knowledge, the following theorem would hold also in **S5**, i.e., our starting system plus negative introspection.

introspection. This is necessary if we wish to tie those results to potential applications of the Theorems 5 and 6.

We start by considering the relation with agnosticism. The following result shows that knowing to be agnostic is sufficient to imply negative introspection and thus to block second-order ignorance.

Proposition 9.

$\vdash_{\mathfrak{L}} K(A(\phi)) \rightarrow (\sim K(\phi) \rightarrow K(\sim K(\phi))) \wedge (\sim K(\sim\phi) \rightarrow K(\sim K(\sim\phi)))$

Intuitively, while the negative introspection condition, per se, can be considered too strict, being agnostic and knowing to be agnostic is a much more reasonable condition. Essentially, by the proposition above, we conclude that the latter causes first-order ignorance but inhibits second-order ignorance, thus avoiding the ignorance loop.

As an orthogonal result with respect to that expressed above, we now present requirements that, in combination with misbelieving/doubting, represent sufficient conditions for negative introspection. As we will show, these requirements are a dual version of the knowledge of agnosticism. Hence, in combination with misbelieving/doubting, they block the passage from first-order ignorance to second-order ignorance.

Proposition 10.

$\vdash_{\mathfrak{L}} (M(\phi) \vee D(\phi)) \wedge K(B(\phi)) \wedge K(\sim K(\phi)) \rightarrow$
$\rightarrow (\sim K(\phi) \rightarrow K(\sim K(\phi))) \wedge (\sim K(\sim\phi) \rightarrow K(\sim K(\sim\phi)))$

We now comment on the interpretation of the requirement $K(B(\phi)) \wedge K(\sim K(\phi))$. The first conjunct expresses the condition of knowing to believe ϕ, i.e., knowing to be under either misbelieving or doubting, since one of the two conditions is assumed. In fact, we recall that both conditions assume $B(\phi)$ and that, by Propositions 1 and 2, the agent is not aware of which condition s/he is subject to. Hence, $K(B(\phi))$ is analogous to the knowledge of agnosticism used in Proposition 9.

The second conjunct, $K(\sim K(\phi))$, combined with the previous one, defines the *Socratic* formula $S(\phi)$ (hence let us use the expression *Socratic agent*). The formula $K(\sim K(\phi))$ is often interpreted in the literature with an alethic reading, where $\sim K(\phi)$ is read as *it is possible to know* $\sim\phi$. In such reading, the formula $K(\sim K(\phi))$ can be interpreted to stand for the cognitive attitude of knowing that it is possible to know $\sim\phi$, i.e., knowing that $\sim\phi$ is a possibility. It therefore becomes reasonable that a Socratic agent avoids second-order ignorance, since such agent would have (and knowing it) a certain belief, i.e., $B(\phi)$, but would be open to the possibility of knowing the opposite of what s/he believes, which is represented by $K(\sim K(\phi))$.

By Proposition 10 and Theorem 5, we derive that the Socratic agent is first-order ignorant but not second-order ignorant, thus avoiding the ignorance loop. Therefore, assuming to know that something is believed, but remaining open to the possibility of knowing the opposite of that something, which are quite reasonable conditions in the presence of misbelieving or doubting, is sufficient to block the loop of ignorance.

While the previous result shows conditions inhibiting the ignorance loop in case of misbelieving/doubting, we now examine conditions enabling the ignorance loop for such conditions. We start by presenting a requirement that, combined with misbelieving, imply the negation of negative introspection and, therefore, the loop of ignorance.[9]

Proposition 11.

$\vdash_{\mathfrak{L}} (M(\phi)\wedge(B(\phi)\rightarrow B(K(\phi))))\rightarrow(\sim K(\phi)\wedge\sim K(\sim K(\phi)))$

The requirement $B(\phi)\rightarrow B(K(\phi))$ represents a very strong condition about an agent's own beliefs and, similarly to $\mathbf{Int}_1$, is one of the interaction principles that has received attention in the literature (Rendsvig and Symons 2019). In particular, if we decided to assume it as an axiom together with negative introspection, in our system the two notions of knowledge and belief would collapse (Rendsvig and Symons 2019), thus making inconsistent the theory we presented about the three conditions inducing ignorance. Anyway, notice that Proposition 11 describes a scenario in which $B(\phi)\rightarrow B(K(\phi))$ negates negative introspection.

Just as in the case of misbelieving, we now consider the case of doubting. In particular, even in such a case, we obtain the same result inducing the negation of negative introspection and, therefore, the loop of ignorance.

Proposition 12.

$\vdash_{\mathfrak{L}} (D(\phi)\wedge(B(\phi)\rightarrow B(K(\phi))))\rightarrow(\sim K(\phi)\wedge\sim K(\sim K(\phi)))$

5. Conclusion

In this paper, we investigated some properties and correlations of two forms of ignorance emerged in the literature, which helped us to provide novel insights in the taxonomy of Figure 1. This was done in the formal framework of a standard **S4** system for knowledge with the addition of a standard **KD** system for belief, which we tied together with the standard interaction axiom $\mathbf{Int}_1$.

On one hand, we explored three ground conditions that contribute to the emergence of ignorance. Together, those conditions are both sufficient and necessary for *ignorance whether* to emerge, while they are necessary for *factive ignorance*.

On the other hand, we clarified the conditions that make the loop of higher-order levels of ignorance (un)avoidable. This is particularly meaningful, as it should be noticed that a system that is unaware of being ignorant, will not be in a position to question such ignorance and, thus, will be unable to produce plans to achieve extra information and fill the gap.

An important future work starting from the theorems proved in this paper will be the study of the form of ignorance that appears in the consequences of Gödel's theorems (Alexander 2014). Alexander (2014) proved that a machine can know its own code exactly but cannot know at the same time its correctness (despite actually being sound). Even though in Aldini et al. 2016 and Aldini et al. 2015 a machine is defined that for (at least) a specific case, knows its own code, and knows to be sound, the ignorance present in Alexander's disjunction still

[9] Differently from the previous propositions, in this case we do not need to prove the negation of negative introspection for both the formula and its negation, since having one portion of it is sufficient to generate second-order ignorance loops.

needs to be deeply explored. So we also expect from this work a contribution to what we would like to dub *ignoring machines*.[10]

References

Aldini, A., Fano, V., & Graziani, P. 2016, "Theory of Knowing Machines: Revisiting Gödel and the Mechanistic Thesis", in Gadducci, F. & Tavosanis, M. (eds.), *History and Philosophy of Computing*, *HaPoC 2015*, *IFIP Advances in Information and Communication Technology*, 487, 57-70.

Aldini, A., Graziani, P., & Tagliaferri, M. 2021, "Reasoning about Ignorance and Beliefs", in Cleophas, L. & Massink, M. (eds.), *Software Engineering and Formal Methods*, *SEFM 2020 Collocated Workshops*, *LNCS 12524*, 214-30.

Aldini, A., Fano, V., & Graziani, P. 2015, "Do the Self-Knowing Machines Dream of Knowing Their Factivity?", *Workshop on Artificial Intelligence and Cognition (AIC)*, 1510, 125-32.

Alexander, S.A. 2014, "A Machine that Knows Its Own Code", *Studia Logica*, 102, 567-76.

Artemov, S. & Fitting, M. 2020, "Justification Logic", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), https://plato.stanford.edu/archives/spr2021/entries/logic-justification/

Fan, J. 2021, "A Logic for Disjunctive Ignorance", *Journal of Philosophical Logic*, 50, 6, 1293-1312.

Fan, J. 2022, "Unknown Truths and False Beliefs: Completeness and Expressivity Results", *Studia Logica*, 110, 1-45.

Fan, J., Wang, Y., & Ditmarsch, H.V. 2015, "Contingency and Knowing Whether", *The Review of Symbolic Logic*, 8, 1, 75-107.

Fano, V. & Graziani, P. 2021, "A Working Hypothesis for the Logic of Radical Ignorance", *Synthese*, 199, 1-2, 601-16.

Fine, K. 2018, "Ignorance of Ignorance", *Synthese*, 195, 9, 4031-45.

Fitch, F. 1963, "A Logical Analysis of Some Value Concepts", *Journal of Symbolic Logic*, 28, 2, 135-42.

Goranko, V. 2021, "On Relative Ignorance filosofiska notiser", *Årgång*, 8, 119-40.

Halpern, J. 1997, "A Theory of Knowledge and Ignorance for Many Agents", *Journal of Logic and Computation*, 7, 1, 79-108.

Halpern, J., Moses, Y., Fagin, R., & Vardi, M. 1995. *Reasoning about Knowledge*, Cambridge, MA: The MIT Press.

Hintikka, J. 1962, *Knowledge and Beliefs: An Introduction to the Logic of the Two Notions*, Ithaca: Cornell University Press.

Kubyshkina, E. & Petrolo, M. 2021, "A Logic for Factive Ignorance", *Synthese*, 198, 5917-28.

Meyer, J.J. & van der Hoek, W.V.D. 1995, *Epistemic Logic for AI and Computer Science*, Cambridge: Cambridge University Press.

[10] This work was supported by the Italian Ministry of Education, University and Research through the PRIN 2017 project "The Manifest Image and the Scientific Image" prot. 2017ZNWW7F_004.

Plato 2004, *Theaetetus* (transl. by J. McDowell), Oxford: Oxford University Press.

Rendsvig, R. & Symons, J. 2019, "Epistemic Logic", *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), https://plato.stanford.edu/archives/sum2021/entries/logic-epistemic/

Steinsvold, C. 2008, "A Note on Logics of Ignorance and Borders", *Notre Dame Journal of Formal Logic*, 49, 4, 385-92.

Tagliaferri, M. 2023, "Using Justified True Beliefs to Explore Formal Ignorance", *Software Engineering and Formal Methods*, *SEFM 2022 Collocated Workshops*, *LNCS 13765*, 400-17.

van der Hoek, W. & Lomuscio, A. 2004, "A Logic for Ignorance", in Leite, J., Omicini, A., Sterling, L., & Torroni, P. (eds.), *Declarative Agent Languages and Technologies,* 97-108.

van Ditmarsch, H., van der Hoek, W., Halpern, J., & Barteld, K. (eds.) 2015, *Handbook of Epistemic Logic*, Milton Keynes: College Publications.

Appendix

Proof of results

Proof. [Proposition 1.]
By contradiction:

(1)	$K(B(\phi) \wedge \neg\phi)$	by contradictory hypothesis
(2)	$K(B(\phi)) \wedge K(\neg\phi)$	by K-distribution over $\wedge$ from (1)
(3)	$K(B(\phi))$	by $\wedge$-elimination from (2)
(4)	$B(\phi)$	by axiom **T** from (3)
(5)	$K(\neg\phi)$	by $\wedge$-elimination from (2)
(6)	$B(\neg\phi)$	by axiom $\mathbf{Int_1}$ from (5)
(7)	$B(\phi) \wedge B(\neg\phi)$	by $\wedge$-introduction from (4) and (6)
(8)	$\neg(B(\phi) \wedge B(\neg\phi))$	axiom **D**
(9)	Contradiction	from (7) and (8)

Proof. [Proposition 2.]
By contradiction:

(1)	$K(B(\phi) \wedge \phi \wedge \neg K(\phi))$	by contradictory hypothesis
(2)	$K(B(\phi)) \wedge K(\phi) \wedge K(\neg K(\phi))$	by K-distribution over $\wedge$ from (1)
(3)	$K(\phi)$	by $\wedge$-elimination from (2)
(4)	$K(\neg K(\phi))$	by $\wedge$-elimination from (2)
(5)	$\neg K(\phi)$	by axiom **T** from (4)
(6)	Contradiction	from (3) and (5)

Proof. [Proposition 3.]
As follows:

(1)	$K(B(\phi))$	by $\wedge$-elimination from assumption
(2)	$K(\neg K(\phi))$	by $\wedge$-elimination from assumption
(3)	$B(\phi)$	by axiom T from (1)
(4)	$B(\phi) \wedge K(\neg K(\phi))$	by $\wedge$-introduction from (2) and (3)
(5)	$\neg K(\phi)$	by axiom **T** from (2)
(6)	$B(\phi) \wedge \neg K(\phi)$	by $\wedge$-introduction from (3) and (5)
(7)	$\phi \vee \neg\phi$	propositional tautology
(8)	$(B(\phi) \wedge \neg K(\phi)) \wedge (\phi \vee \neg\phi)$	by $\wedge$-introduction from (6) and (7)

(9)	$(B(\phi) \wedge \neg K(\phi) \wedge \neg\phi) \vee$ $(B(\phi) \wedge \neg K(\phi) \wedge \phi)$	by $\wedge$-distribution over $\vee$ from (8)
(10)	$(B(\phi) \wedge \neg K(\phi) \wedge \neg\phi) \rightarrow (B(\phi) \wedge \neg\phi)$	propositional tautology
(11)	$(B(\phi) \wedge \neg\phi) \vee (B(\phi) \wedge \neg K(\phi) \wedge \phi)$	by substitution from (9) and (10)
(12)	$(B(\phi) \wedge K(\neg K(\phi))) \rightarrow$ $((B(\phi) \wedge \neg\phi) \vee (B(\phi) \wedge \neg K(\phi) \wedge \phi))$	by $\rightarrow$-introduction from (4) and (11)
(13)	$K(K(\neg K(\phi)))$	by axiom **4** from (2)
(14)	$K(B(\phi) \wedge K(\neg K(\phi)))$	by K-distribution over $\wedge$ from (1) and (13)
(15)	$K((B(\phi) \wedge K(\neg K(\phi))) \rightarrow$ $((B(\phi) \wedge \neg\phi) \vee (B(\phi) \wedge \neg K(\phi) \wedge \phi)))$	by *necessitation* from (12)
(16)	$K((B(\phi) \wedge \neg\phi) \vee (B(\phi) \wedge \phi \wedge \neg K(\phi))$	by axiom K from (14) and (15)

Proof. [Proposition 7.]

The result about doubting holds trivially by definition. Here we show the result about misbelieving:

(1)	ϕ	by $\wedge$-elimination from assumption
(2)	$B(\neg\phi)$	by $\wedge$-elimination from assumption
(3)	$\neg B(\phi)$	by axiom **D** from (2)
(4)	$K(\phi) \rightarrow B(\phi)$	axiom $\mathbf{Int_1}$
(5)	$\neg B(\phi) \rightarrow \neg K(\phi)$	from (4) by contraposition
(6)	$\neg K(\phi)$	from (3) and (5) by *Modus Ponens*
(7)	$\phi \wedge \neg K(\phi)$	by $\wedge$-introduction from (1) and (6)

Proof. [Proposition 8.]

By Prop. 6, we have that $A(\phi)$ implies $\neg K(\phi) \wedge \neg K(\neg\phi)$, which in turn implies $(\neg K(\phi) \wedge \neg K(\neg\phi)) \wedge (\phi \vee \neg\phi)$ by propositional tautology. By distributivity, we derive $(\phi \wedge \neg K(\phi) \wedge \neg K(\neg\phi)) \vee (\neg\phi \wedge \neg K(\phi) \wedge \neg K(\neg\phi))$, from which the result follows immediately.

Proof. [Theorem 2.]

As follows:

(1)	ϕ	by $\wedge$-elimination from assumption
(2)	$\neg K(\phi)$	by $\wedge$-elimination from assumption
(3)	$K(\neg\phi) \rightarrow \neg\phi$	Axiom T
(4)	$\phi \rightarrow \neg K(\neg\phi)$	from (3) by contraposition
(5)	$\neg K(\neg\phi)$	from (1) and (4) by *Modus Ponens*
(6)	$\neg K(\phi) \wedge \neg K(\neg\phi)$	from (2) and (5) by $\wedge$-introduction
(7)	$I(\phi)$	by definition of (6)

Proof. [Theorem 3.]

Observed that $I_f(I_f(\phi))$ corresponds by definition to $I_f(\phi) \wedge \neg K(I_f(\phi))$, the proof will be given by contradiction, by assuming both $I_f(\phi)$ and $K(I_f(\phi))$:

(1)	$K(I_f(\phi))$	Assumption
(2)	$K(\phi \wedge \neg K(\phi))$	by definition of (1)
(3)	$K(\phi \wedge \neg K(\phi)) \rightarrow (\phi \wedge \neg K(\phi))$	Axiom T
(4)	$\phi \wedge \neg K(\phi)$	from (2) and (3) by *Modus Ponens*
(5)	$\neg K(\phi)$	from (4) by $\wedge$-elimination
(6)	$K((\phi \wedge \neg K(\phi)) \rightarrow \phi) \rightarrow$ $(K(\phi \wedge \neg K(\phi)) \rightarrow K(\phi))$	Axiom K
(7)	$K(\neg\phi \vee K(\phi) \vee \phi) \rightarrow$ $(K(\phi \wedge \neg K(\phi)) \rightarrow K(\phi))$	from (6) by substitution and DeMorgan
(8)	$(\neg\phi \vee \phi)$	propositional tautology
(9)	$(\neg\phi \vee K(\phi) \vee \phi)$	from (8) by $\vee$-introduction
(10)	$K(\neg\phi \vee K(\phi) \vee \phi)$	from (9) by necessitation
(11)	$K(\phi \wedge \neg K(\phi)) \rightarrow K(\phi)$	from (7) and (10) by *Modus Ponens*
(12)	$\neg K(\phi) \rightarrow \neg K(\phi \wedge \neg K(\phi))$	from (11) by contraposition
(13)	$\neg K(\phi \wedge \neg K(\phi))$	from (5) and (12) by *Modus Ponens*
(11)	Contradiction	from (2) and (13)

Proof. [Proposition 9.]

By axiom T, the hypothesis $K(A(\phi))$ implies $A(\phi)$, which, by Prop. 6, implies $I(\phi)$, and, in particular, $\neg K(\phi)$ and $\neg K(\neg\phi)$ (the two antecedents on the conditionals that comprise the consequent of Prop. 9). What we must show is that, starting with such hypothesis, both $K(\neg K(\phi))$ (the consequent of the first conjunct of the consequent of Prop. 9) and $K(\neg K(\neg\phi))$ (the consequent of the second conjunct of the consequent of Prop. 9) hold.

This is done as follows:

(1)	$K(\neg B(\phi) \wedge \neg B(\neg\phi))$	Assumption
(2)	$K(\neg B(\phi)) \wedge K(\neg B(\neg\phi))$	from (1) by K-distribution over $\wedge$
(3)	$K(\neg B(\phi))$	from (2) by $\wedge$-elimination
(4)	$K(\neg B(\neg\phi))$	from (2) by $\wedge$-elimination
(5)	$K(\phi) \rightarrow B(\phi)$	Axiom *Int₁*
(6)	$\neg B(\phi) \rightarrow \neg K(\phi)$	from (5) by contraposition
(7)	$K(\neg\phi) \rightarrow B(\neg\phi)$	Axiom *Int₁*
(8)	$\neg B(\neg\phi) \rightarrow \neg K(\neg\phi)$	from (7) by contraposition
(9)	$K(\neg B(\phi) \rightarrow \neg K(\phi))$	from (6) by necessitation
(10)	$K(\neg B(\neg\phi) \rightarrow \neg K(\neg\phi))$	from (8) by necessitation
(11)	$K(\neg B(\phi) \rightarrow \neg K(\phi)) \rightarrow$ $(K(\neg B(\phi)) \rightarrow K(\neg K(\phi)))$	Axiom K
(12)	$K(\neg B(\neg\phi) \rightarrow \neg K(\neg\phi)) \rightarrow$ $(K(\neg B(\neg\phi)) \rightarrow K(\neg K(\neg\phi)))$	Axiom K
(13)	$(K(\neg B(\phi)) \rightarrow K(\neg K(\phi)))$	from (9) and (11) by *Modus Ponens*
(14)	$K(\neg K(\phi))$	from (3) and (13) by *Modus Ponens*

(15)	$(K(\neg B(\neg\phi)) \rightarrow K(\neg K(\neg\phi)))$	from (10) and (12) by *Modus Ponens*
(16)	$K(\neg K(\neg\phi))$	from (4) and (15) by *Modus Ponens*

Proof. [Proposition 10.]

First, the hypothesis $(M(\phi) \vee D(\phi))$ implies $I(\phi)$ by Prop. 4 and Prop. 5. Hence, we have both $\neg K(\phi)$ and $\neg K(\neg\phi)$, which are the two antecedents on the conditionals that comprise the consequent of Prop. 10. Hence, it is sufficient to show that the conjunction $K(B(\phi)) \wedge K(\neg K(\phi))$ implies both $K(\neg K(\phi))$ and $K(\neg K(\neg\phi))$ in order to prove the proposition:

(1)	$K(B(\phi)) \wedge K(\neg K(\phi))$	Assumption
(2)	$B(\phi) \rightarrow \neg B(\neg\phi)$	Axiom D
(3)	$K(B(\phi) \rightarrow \neg B(\neg\phi))$	from (2) by necessitation
(4)	$K(B(\phi) \rightarrow \neg B(\neg\phi)) \rightarrow$ $(K(B(\phi)) \rightarrow K(\neg B(\neg\phi)))$	Axiom K
(5)	$K(B(\phi))$	from (1) by $\wedge$-elimination
(6)	$(K(B(\phi)) \rightarrow K(\neg B(\neg\phi)))$	from (3) and (4) by *Modus Ponens*
(7)	$K(\neg B(\neg\phi))$	from (5) and (6) by *Modus Ponens*
(8)	$K(\neg\phi) \rightarrow B(\neg\phi)$	Axiom Int_1
(9)	$\neg B(\neg\phi) \rightarrow \neg K(\neg\phi)$	from (8) by contraposition
(10)	$K(\neg B(\neg\phi) \rightarrow \neg K(\neg\phi))$	from (9) by necessitation
(11)	$K(\neg B(\neg\phi) \rightarrow \neg K(\neg\phi)) \rightarrow$ $(K(\neg B(\neg\phi)) \rightarrow K(\neg K(\neg\phi)))$	Axiom K
(12)	$(K(\neg B(\neg\phi)) \rightarrow K(\neg K(\neg\phi)))$	from (10) and (11) by *Modus Ponens*
(13)	$K(\neg K(\neg\phi))$	from (7) and (12) by *Modus Ponens*
(14)	$K(\neg K(\phi))$	from (1) by $\wedge$-elimination

Proof. [Proposition 11.]

As follows:

(1)	$B(\phi) \wedge \neg\phi$	Assumption
(2)	$B(\phi) \rightarrow B(K(\phi))$	Assumption
(3)	$B(\phi)$	from (1) by $\wedge$-elimination
(4)	$B(K(\phi))$	from (2) and (3) by *Modus Ponens*
(5)	$\neg B(K(\phi)) \vee \neg B(\neg K(\phi))$	Axiom D
(6)	$\neg B(\neg K(\phi))$	from (4) and (5) by disjunctive syllogism
(7)	$K(\neg K(\phi)) \rightarrow B(\neg K(\phi))$	Axiom Int_1
(8)	$\neg B(\neg K(\phi)) \rightarrow \neg K(\neg K(\phi))$	from (7) by contraposition
(9)	$\neg K(\neg K(\phi))$	from (6) and (8) by *Modus Ponens*
(10)	$\neg\phi$	from (1) by $\wedge$-elimination
(11)	$K(\phi) \rightarrow \phi$	Axiom T
(12)	$\neg\phi \rightarrow \neg K(\phi)$	from (11) by contraposition
(13)	$\neg K(\phi)$	from (10) and (13) by *Modus Ponens*
(14)	$\neg K(\phi) \wedge \neg K(\neg K(\phi))$	from (9) and (13) by $\wedge$-introduction

Proof. [Proposition 12.]

As follows:

(1)	$B(\phi) \wedge \phi \wedge \neg K(\phi)$	Assumption
(2)	$B(\phi) \rightarrow B(K(\phi))$	Assumption
(3)	$B(\phi)$	from (1) by $\wedge$-elimination
(4)	$B(K(\phi))$	from (2) and (3) by *Modus Ponens*
(5)	$\neg B(K(\phi)) \vee \neg B(\neg K(\phi))$	by axiom D
(6)	$\neg B(\neg K(\phi))$	from (4) and (5) by disjunctive syllogism
(7)	$K(\neg K(\phi)) \rightarrow B(\neg K(\phi))$	Axiom Int_1
(8)	$\neg B(\neg K(\phi)) \rightarrow \neg K(\neg K(\phi))$	from (7) by contraposition
(9)	$\neg K(\neg K(\phi))$	from (6) and (8) by *Modus Ponens*
(10)	$\neg K(\phi)$	from (1) by $\wedge$-elimination
(11)	$\neg K(\phi) \wedge \neg K(\neg K(\phi))$	from (9) and (10) by $\wedge$-introduction