Using Justified True Beliefs to Explore Formal Ignorance.

Mirko Tagliaferri[1][0000−0003−3875−0512]

Università degli Studi di Urbino "Carlo Bo", Urbino, Italy
mirko.tagliaferri@uniurb.it

Abstract. The possibility of better understanding belief and knowledge modalities through justifications is not a novel one, however, the machinery of justifications has never been employed to explore the nature of ignorance from a formal perspective. By including justification terms into a modal logic for belief a major project (among others) can be pursued: different cognitive stances can can be formalized that imply ignorance, therefore highlighting even better the possible culprits of the emergence of the phenomenon of ignorance. This paper is a first step in the direction of tackling such problem.

Keywords: Doxastic Justification Logic · Formal Ignorance · Radical Ignorance.

1 Introduction and Motivations

In recent years, there has been a renaissance of interest in the formal treatment of ignorance [1, 7, 8, 10, 12]. This renewed interest is partly dependant on the necessity of understanding how ignorance can influence different reasoning dynamics. This is especially important when it is noticed that ignorance is a wide phenomenon that influences many different fields. Both human beings and artificial intelligent (AI) systems will often have to reason with incomplete sets of data, which means that they will be ignorant about various facts that might influence their decision making capabilities. Therefore, a good grasp of the characteristics and formal properties of ignorance could help to devise strategies that can be useful both in educating people and in programming AI systems. Moreover, as shown by Kit Fine [9], it is also important to better understand the dynamics that bridge simple and radical forms of ignorance. This is due to the fact that once an agent becomes second-order ignorant about a specific fact (s/he ignores that s/he ignores the fact), s/he will spiral into a black hole of ignorance, unable to escape from it by him/herself.

A common practice is to define ignorance in terms of (lack of) knowledge, which can then, in turn, be interpreted in various possible ways depending on the underlying logical languages that are employed.[1] This kind of approach has

[1] Obviously, other approaches are also possible, e.g., interpreting ignorance as a primitive notion [6]. Moreover, even following this simple approach of reducing ignorance

the advantage of being extremely simple and allows a first basic understanding of various characteristics of ignorance. However, this simplicity comes at a cost. The causes of ignorance are left obscure and connections between the phenomenon of ignoring and other cognitive phenomena are left unexplained.

A first step in the direction of exploring the connection between different cognitive phenomena and ignorance can be found in [1]. In such paper, the authors propose three formal representations of doxastic stances[2] and show how those stances can lead to ignorance, thus partially explaining the potential causes of ignorance. However, while their approach is successful in highlighting some of the connections between beliefs and ignorance, some of the cognitive phenomena they describe seem confusing[3] and thus fail to provide the whole picture.

In this paper, I propose to augment their formal machinery with evidence formulas, thus obtaining the language of evidence-based beliefs as presented in [4, 5], although with a slightly different interpretation of the operators. The idea is that such additional evidence could be interpreted as standing for justifications for specific facts. The inclusion of a justification element would then allow a definition of knowledge inside the language as justified true belief, following the spirit of the tripartite definition proposed by Plato in the *Theaetetus* [13]. Furthermore, having the extra element of evidence could clarify better some of the cognitive stance already analysed, while also allowing novel formal definitions for other stances relevant for ignorance that agents might have.

This possibility of better understanding belief and knowledge modalities through justifications is not a novel one (see, e.g., [2] for a good introduction), however, to my knowledge, the machinery of justifications has never been employed to explore the nature of ignorance from a formal perspective. By including justification terms into a modal logic for belief a major project (among others) can be pursued: different cognitive stances can can be formalized that imply ignorance, therefore highlighting even better the possible culprits of the emergence of the phenomenon of ignorance.

In order to achieve this goal, the paper will do three things: first (section 2), I will introduce the syntax and semantics of the language of evidence-based beliefs. I will also explain in which way my interpretation of the operators on formulas varies from the one proposed by the original authors of the language in [4, 5]. Even though practically identical from the formal point-of-view, I'll call such language in a different way, i.e., *JTB* (Justified True Belief). This is done to make clear the aim of this specific paper and to highlight the difference in interpretation of the operators. Then (section 3), the phenomenon of basic ignorance (i.e., ignoring a fact) will be analysed by providing formal definitions

---

to the lack of knowledge can produce different formalizations depending on how the authors interpret the phrase "lack of knowledge", e.g., as *not knowing that* or *not knowing whether*.

[2] Being agnostic: $\neg B(\phi) \wedge \neg B(\neg\phi)$, misbelieving: $B(\phi) \wedge \neg\phi$, and doubting: $B(\phi) \wedge \phi \wedge \neg K(\phi)$. Where $B(\phi)$ should be interpreted as ϕ is believed and $K(\phi)$ as ϕ is known.

[3] The phenomenon of doubting seems particularly obscure, since it is left unexplained why ϕ is not known even though it is believed and it is true.

of stances that imply ignorance. Intuitive examples of the presence of each stance will be provided. Finally (section 4), some concluding remarks will follow, and potential future direction of this work will be presented.

2 Justified True Belief Logic

2.1 Syntax

Definition 1 (*JTB*)**.** *Given a set* Φ *of atomic propositions, the language JTB of formulas* $\varphi \in JTB$ *is defined recursively as follows:*

$$\varphi \coloneqq p \in \Phi \mid \neg\varphi \mid \varphi \wedge \varphi \mid B(\varphi) \mid E_l(\varphi) \mid E_c(\varphi)$$

The other logical connectives are defined as usual from negation and conjunction. Intuitively, the modal formulas should be read in the following way: $B(\varphi)$ means that φ is believed; $E_l(\varphi)$ means that there is limited evidence for φ; finally, $E_c(\varphi)$ means that there is conclusive evidence for φ. The difference between limited and conclusive evidence is the following: someone has limited evidence for a fact whenever part of the evidence s/he possesses is consistent with the truth of the fact, but other pieces of evidence might defeat such truth (i.e., it might indicate that the fact is indeed false). On the other hand, conclusive evidence completely supports the truth of the fact, eliminating any potential doubt on such truth from an evidential standpoint. Obviously, someone possessing conclusive evidence implies that s/he also possesses limited evidence, but not viceversa.

Given the language JTB, it is possible to define a knowledge operator as justified true belief. Then, from this knowledge operator, it is possible to define an ignorance operator as (lack of) knowledge whether. Formally:

- $K(\varphi) \coloneqq B(\varphi) \wedge \varphi \wedge E_c(\varphi)$;
- $I(\varphi) \coloneqq \neg K(\varphi) \wedge \neg K(\neg\varphi)$.

Those two definitions will constitute the main elements of the analysis carried out in the later sections of this paper.

2.2 Semantics

In order to interpret the formulas of the language JTB, the following structure will be employed.

Definition 2 (Models). *A model* $\mathfrak{M}$*=*$(W, \pi, R_B, \mathcal{E})$ *is a structure consisting of a nonempty set* W *of possible worlds, a valuation function* $\pi : \Phi \to \mathcal{P}(W)$*, a binary relation* $R_B \subseteq W \times W$*, and an evidence function* $\mathcal{E} : W \to \mathcal{P}(\mathcal{P}(W))$*. A pair* $(\mathfrak{M}, w)$ *is called a pointed model.*

The set W is treated as in standard modal logic and contains elements (the possible worlds) that are maximally consistent descriptions of how the world could be. Those descriptions are maximal because all possible details are described, and they are consistent because no contradictory information is allowed in the same possible world.

The valuation function π determines the truth of all of the atomic propositions inside the structure.

The binary relation R_B is a doxastic accessibility relation and determines which worlds are doxastically accessible, i.e., which worlds are taken into consideration to determine the beliefs of a given agent. For the purposes of the present paper, it is assumed that the relation R_B is serial. As will be shown later, this ensures that beliefs are always consistent. It is possible to add properties to the binary relation R_B to obtain beliefs with various characteristics[4]; however, since only consistency will be important in the formalization of the cognitive stances introduced in section 3 and the derivations from those stances to basic ignorance, only the property of seriality will be assumed.

The last element of the structure is the evidence function $\mathcal{E}$ that indicates at each possible state of the model which pieces of evidence are available. In JTB, pieces of evidence are interpreted as set of possible worlds. This is a common practice in the formalization of uncertainty [11] and follows from the intuitive idea that a piece of evidence indicates to an agent a set of possibilities (possible worlds) from which the world s/he thinks s/he is in has to be chosen. For the purpose of this paper, only very few properties will be assumed about $\mathcal{E}$:

- $\emptyset \notin \mathcal{E}$;
- $W \in \mathcal{E}$.

The first assumption assures that the evidence sets are consistent, i.e., there is no direct piece of evidence supporting a contradiction. Even though contradictions may not be supported directly by evidence, it is still possible that two pieces of evidence contradict each other, i.e., the intersection of all the evidence possessed by an agent might still be the empty set. It would be possible to require that all the pieces of evidence possessed by an agent must be consistent with each other, Such assumption, however, might be too strong given that evidence is often gathered in different contexts and from different sources, thus allowing for the possibility of receiving conflicting evidence. The second assumption made simply assures that the whole space of possibilities is supported, i.e., an agent has always access to trivial evidence. No further assumptions will be made over the evidence function $\mathcal{E}$.

As said above, the set of evidence might contain elements that are conflicting, indicating that a fact is both true and false. This means that it might not always be possible to combine such evidence in order to obtain answers. Nonetheless, it is still possible to take consistent subsets of such evidence and combine it in order to obtain partial indications over the potential truth of the fact. Those subsets of evidence will be called *maximal consistent evidence sets* and will be

[4] See [14] for some examples.

employed to interpret the evidence operators of JTB. In the following definition I will use X_i to indicate subsets of W, i.e., $X_i \in \mathcal{P}(W)$, and use $\mathcal{X}_i$ to indicate sets of subsets of W, i.e., $\mathcal{X}_i \in \mathcal{P}(\mathcal{P}(W))$.

Definition 3 (Maximal consistent evidence sets). *A set $\mathcal{X}_i$ is a maximal consistent evidence set iff*

1. *It has the finite intersection property (f.i.p.), i.e., $\bigcap_{X \in \mathcal{X}_i} X \neq \emptyset$;*
2. *It is a maximal set with such property, i.e., there is no set $\mathcal{X}_j$ such that $\mathcal{X}_i \subset \mathcal{X}_j$ and $\mathcal{X}_j$ has the finite intersection property.*

The formal structure provided allow an interpretation for the language JTB.

Definition 4 (Truth). *Given a model $\mathfrak{M}$, a possible world w and a formula φ of the language JTB, the satisfaction of a formula φ at a pointed model $(\mathfrak{M}, w)$, formally $(\mathfrak{M}, w) \models \varphi$, is defined recursively as follows:*

1. $(\mathfrak{M}, w) \models p$ *for* $p \in \Phi$ *iff* $w \in \pi(p)$;
2. $(\mathfrak{M}, w) \models \neg\varphi$ *iff* $(\mathfrak{M}, w) \not\models \neg\varphi$;
3. $(\mathfrak{M}, w) \models \varphi \wedge \psi$ *iff* $(\mathfrak{M}, w) \models \varphi$ *and* $(\mathfrak{M}, w) \models \psi$;
4. $(\mathfrak{M}, w) \models B(\varphi)$ *iff* $\forall v, wR_Bv, (\mathfrak{M}, w) \models \varphi$;
5. $(\mathfrak{M}, w) \models E_l(\varphi)$ *iff* $\exists \mathcal{X}_i \subseteq \mathcal{E}(w)$, *s.t.,* $\mathcal{X}_i$ *is a maximal consistent evidence set and* $\forall v \in \bigcap_{X \in \mathcal{X}_i} X, (\mathfrak{M}, w) \models \varphi$;
6. $(\mathfrak{M}, w) \models E_c(\varphi)$ *iff* $\forall \mathcal{X}_i \subseteq \mathcal{E}(w)$, *s.t.,* $\mathcal{X}_i$ *are maximal consistent evidence sets, it follows that* $\forall v \in \bigcap_{X \in \mathcal{X}_i} X, (\mathfrak{M}, w) \models \varphi$.

I will indicate with $\|\varphi\|_{\mathfrak{M}}$ the truth set of φ, i.e., $\|\varphi\|_{\mathfrak{M}} = \{w \mid (\mathfrak{M}, w) \models \varphi\}$ The first three conditions of definition 4 are the classical satisfaction relations of propositional logic. The fourth says that something is believed whenever it is true in all doxastically accessible worlds and it a common assumption in modal logic. The fifth and sixth conditions is where I diverge a little from the approach taken in [4, 5]. In such papers, the authors took the evidence operator as being true whenever there existed a single piece of evidence completely supporting the fact, i.e., if $\exists X \in \mathcal{E}, s.t., X \subseteq \|\varphi\|_{\mathfrak{M}}$. While I understand the benefits of doing so and do realize that sometimes this might indeed be the case, I dislike their approach in general. Always requiring just a single piece of evidence and discarding all the other available pieces of evidence (including those pieces of evidence that are consistent with the one chosen) does not seem to accurately represent real-case scenarios. In fact, they also seem to agree that taking all the consistent evidence and combine it is indeed a useful practice, but they then use it to define the belief operator, instead of a stronger evidence operator. While this aligns with their aims of bridging the gap between the formation of beliefs and availability of evidence, I have some reserves about the fact that the former should collapse on the latter. Having evidence (even consistent and conclusive evidence) does not always lead to the formation of an associated belief. For those reasons I choose to follow a different route.

I claim that there is limited evidence about a fact being true whenever there is at least a consistent set of pieces of evidence that indicate the truth of the given fact. Note that this condition is open to the possibility that the set of pieces of evidence is indeed a singleton set (containing only one piece of evidence). In such a case, my condition would collapse onto that given in [4, 5]. It is easy to see that someone could have limited evidence for contradictory facts, i.e., it is possible to satisfy both $E_l(\varphi)$ and $E_l(\neg\varphi)$ in a model.

On the other hand, there is conclusive evidence about a fact being true only when all pieces of evidence are taken into consideration with respect to their maximal consistent sets and the fact is true in all worlds that are compatible with such evidence. Note that the way conclusive evidence is evaluated is different from taking the intersection all the available evidence and then checking the truth of the fact. A simple example should help to clarify the difference in the two procedures. I'll present first the procedure using the simple intersection of evidence. Suppose you have two pieces of evidence, one indicating that φ is true and one indicating that it is false. Now, if the intersection is taken between the two pieces of evidence, then the empty set would obtain (the two pieces of evidence are disjoint). This would result in having conclusive evidence for both φ and its negation, which is absurd. Using the other procedure: being the two pieces of evidence disjoint, it would mean that there are two maximal consistent sets of evidence that should be taken into consideration, one constituted by the single piece of evidence supporting φ and one constituted by the single piece of evidence supporting $\neg\varphi$. At this point, in order to establish whether there is conclusive evidence for either φ or its negation, it would be necessary to check their truth in all worlds that are members of both the maximal consistent sets of evidence. Obviously, neither of the two would hold in all of those states, meaning that there is no conclusive evidence for one or the other, as intuitively expected. It is important to notice that the assumption $\emptyset \notin \mathcal{E}$ guarantees that there will always be at least a maximal consistent set of evidence pieces in $\mathcal{E}$ and, moreover, by the definition of maximal consistent set (definition 3, there will always be a possible world to check. This guarantees that there is no vacuous conclusive evidence for anything. Moreover, being tautologies true in all possible worlds, it is easy to see that there will always be conclusive evidence for those, no matter the evidence function.

Finally, it is also easy to see that the implication from conclusive evidence to limited evidence that was discussed in subsection 2.1 holds. This is due to the fact that if a fact is true in all world that are indicated by all maximal consistent evidence sets, then such fact must be true also in all worlds that are indicated by one of the maximal consistent evidence sets.

3 The Origins of Basic Ignorance

In this section, various stances will be explored and it will be shown that they all imply to basic ignorance (ignorance of a fact φ. The starting point of the reflections of this section are the results obtained in [1]. In such paper, the authors managed to show that three doxastic stances where sufficient and jointly necessary conditions for ignorance whether. While their results did help to shed some

Table 1. Cognitive stances that imply ignorance.

Disbelief	$\neg B(\varphi) \land \neg B(\neg\varphi) \land (E_c(\varphi) \lor E_c(\neg\varphi))$
Skepticism	$\neg B(\varphi) \land \neg B(\neg\varphi) \land (E_l(\varphi) \lor E_l(\neg\varphi)) \land (\neg E_c(\varphi) \lor \neg E_c(\neg\varphi))$
Unawareness	$\neg B(\varphi) \land \neg B(\neg\varphi) \land (\neg E_l(\varphi) \land \neg E_l(\neg\varphi))$
Mislead	$B(\varphi) \land \neg\varphi \land E_c(\varphi)$
Negative Belief Perseverance	$B(\varphi) \land \neg\varphi \land E_c(\neg\varphi)$
Credulity	$B(\varphi) \land \neg\varphi \land E_l(\varphi) \land \neg E_c(\varphi)$
Misbelief	$B(\varphi) \land \neg\varphi \land \neg E_l(\varphi)$
Positive Belief Perseverance	$B(\varphi) \land \varphi \land E_c(\neg\varphi)$
Doubt	$B(\varphi) \land \varphi \land E_l(\varphi) \land \neg E_c(\varphi)$
Intuition	$B(\varphi) \land \varphi \land \neg E_l(\varphi)$

lights on the origins of ignorance, some aspects where unclear. In this section, those aspects will be clarified and a more fine-grained analysis of the origins of basic ignorance will be pursued. Specifically, employing the evidence component of JTB, all doxastic stances (agnosticism, misbelieving, and doubting) will be further analysed and novel more interesting stances will be presented. Table 1 contains a summary of all the stances with their formalization.

Disbelief. *Disbelief* is a stance of mental rejection of a fact even in the face of conclusive evidence in its favour. An example of such a stance could be a parent that refuses to believe that her son committed a crime, even when she is presented with conclusive evidence that he did commit such crime. Another possibility of such a stance is when the evidence, even though conclusive, is considered forged by the agent, i.e., when the agent thinks that $E_c(\varphi) \land \neg\varphi$ holds.

Example 1 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_2)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg B(p)$
- $(\mathfrak{M}, w_1) \models \neg B(\neg p)$
- $(\mathfrak{M}, w_1) \models E_c(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $\neg B(p) \land \neg B(\neg p) \land (E_c(p) \lor E_c(\neg p))$ holds, showing that a *disbelief stance* is present.

I will now show that a disbelief stance leads to basic ignorance.

Theorem 1 (From Disbelief to Ignorance). *A disbelief stance implies basic ignorance. Formally:*

$$(\neg B(\varphi) \land \neg B(\neg\varphi) \land (E_c(\varphi) \lor E_c(\neg\varphi))) \to I(\varphi) \quad (1)$$

Proof. Unpack the definition of $I(\varphi)$ into $\neg K(\varphi) \wedge \neg K(\neg\varphi)$. To prove theorem 1, it is necessary to show that $\neg B(\varphi) \wedge \neg B(\neg\varphi) \wedge (E_c(\varphi) \vee E_c(\neg\varphi))$ implies both $\neg K(\varphi)$ and $\neg K(\neg\varphi)$. First, unpack the definition of $\neg K(\varphi)$ into $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$. The first element of the disjunction $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$ follows directly from the first element of the conjunction $\neg B(\varphi) \wedge \neg B(\neg\varphi) \wedge (E_c(\varphi) \vee E_c(\neg\varphi))$, proving $\neg K(\varphi)$. Now, unpack the definition of $\neg K(\neg\varphi)$ into $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$. The first element of the disjunction $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$ follows directly from the second element of the conjunction $\neg B(\varphi) \wedge \neg B(\neg\varphi) \wedge (E_c(\varphi) \vee E_c(\neg\varphi))$, proving $\neg K(\neg\varphi)$. □

Skepticism. *Skepticism* is the stance that is closest to the agnosticism effect given in [1]. Someone who is in a skeptic stance might have some evidence in favour of a specific fact, but still decides to suspend his/her judgement waiting for further evidence in favour or against the truth of the fact.

Example 2 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_2), (w_1, w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg B(p)$
- $(\mathfrak{M}, w_1) \models \neg B(\neg p)$
- $(\mathfrak{M}, w_1) \models E_l(p)$
- $(\mathfrak{M}, w_1) \models \neg E_c(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $\neg B(p) \wedge \neg B(\neg p) \wedge (E_l(p) \vee E_l(\neg p)) \wedge (\neg E_c(p) \vee \neg E_c(\neg p))$ holds, showing that a *skeptic stance* is present.

Theorem 2 (From Skepticism to Ignorance). *A skeptic stance implies basic ignorance. Formally:*

$$(\neg B(\varphi) \wedge \neg B(\neg\varphi) \wedge (E_l(\varphi) \vee E_l(\neg\varphi)) \wedge (\neg E_c(\varphi) \vee \neg E_c(\neg\varphi))) \rightarrow I(\varphi) \tag{2}$$

Proof. The proof of theorem 2 is equivalent to that of theorem 1. □

Unawareness. *Unawareness* is the stance that describes an agent who does not have any information regarding a specific fact. This could happen for two reasons: either i) the agent never had the chance to gather evidence for or against the fact, thus is completely unaware of whether it might be true or false; or ii) the agent simply did not even consider to gather such evidence because s/he have never even entertained the idea of the fact.

Example 3 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3, w_4)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_2), (w_1, w_3), (w_1, w_4)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_2, w_4), (w_1, w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg B(p)$

- $(\mathfrak{M}, w_1) \models \neg B(\neg p)$
- $(\mathfrak{M}, w_1) \models \neg E_l(p)$
- $(\mathfrak{M}, w_1) \models \neg E_l(\neg p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $\neg B(p) \wedge \neg B(\neg p) \wedge (\neg E_l(p) \wedge \neg E_l(\neg p))$ holds, showing that an *unawareness stance* is present.

Theorem 3 (From Unawareness to Ignorance). *An unawareness stance implies basic ignorance. Formally:*

$$(\neg B(\varphi) \wedge \neg B(\neg\varphi) \wedge (E_l(\varphi) \vee E_l(\neg\varphi)) \wedge (\neg E_c(\varphi) \vee \neg E_c(\neg\varphi))) \rightarrow I(\varphi) \quad (3)$$

Proof. The proof of theorem 3 is equivalent to that of theorem 1. □

Note that the proofs of theorems 1, 2, and 3 are all equivalent. This is because the negative beliefs components of the respective stances is what implies ignorance, i.e., it is the fact that the agent does not believe either φ or $\neg\varphi$ that implies his/her ignorance of the fact. This class of stances represent a failure of the first component of the tripartite definition of knowledge.

The reader might think that the evidence component is useless in those cases, and s/he would not be completely wrong. In the cases in which beliefs are withheld, such withholding alone is sufficient to imply ignorance. However, the added component of evidence could highlight potential strategies to avoid such ignorance. For example, if it is known that an agent is in an unawareness stance, then it might be possible to eliminate his/her ignorance by providing him/her with evidence in favour or against the fact that is ignored. Differently, if it is known that the agent is in a disbelief stance, such strategy would be almost useless and more drastic measurements might be required. Obviously, those considerations apply only partially in the context of this paper, since the language introduced is static in nature (i.e., it does not contain elements that allow for a change of evidence and/or beliefs). Nonetheless, having an initial understanding of those phenomena could help in the future to design strategies inside potential extensions of *JTB* that allow such updates to happen.

The next two classes of stances that will follow have an important feature, i.e., they are equivalent *modulo* the truth of the fact that is evaluated. Assuming that agents do not have direct access to such truth, for them it is practically impossible to subjectively understand whether they are in the first class of stances (those in which what they believe is false) or the second (those in which what they believe is actually true). In the real world, there might be pragmatic consideration that could help an agent to distinguish the two classes of stances. Such considerations involve the way in which evidence is gathered and the expertise of the agent in the specific matter on which s/he is forming his/her beliefs. Since in *JTB* no reference is made to how the evidence is gathered, the language is not in a position to formalize such considerations. Again, the aim of this paper is to understand the potential origins of ignorance and not of solving the problem right away (e.g., by indicating which considerations should be made to enter stances that are less troublesome from the point-of-view of ignorance).

Mislead. *Mislead*[5] is the stance that describes an agent who has a false belief supported by conclusive evidence. Given that the fact is indeed false, it follows that the conclusive evidence possessed by the agent is misleading, possibly convincing the agent to believe the fact over false premises. This stance is the most worrisome among all the ones presented in this paper because the only difference between a mislead stance and knowledge is the truth of the fact itself. Given the assumption that the agent does not have direct access to such truth, it is subjectively impossible to distinguish between the two. This implies that mislead individuals might end up having troublesome higher-order beliefs about their cognitive state, i.e., they might believe that they know something even when they are actually ignorant. Therefore, whenever a mislead stance is recognized, particular attention must be paid in the treatment of the agent's ignorance.

Example 4 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_2, w_3\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_2, w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models E_c(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge \neg p \wedge E_c(p)$ holds, showing that a *mislead stance* is present.

Theorem 4 (From Mislead to Ignorance). *A mislead stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \neg\varphi \wedge E_c(\varphi)) \rightarrow I(\varphi) \tag{4}$$

Proof. Unpack the definition of $I(\varphi)$ into $\neg K(\varphi) \wedge \neg K(\neg\varphi)$. To prove theorem 4, it is necessary to show that $B(\varphi) \wedge \neg\varphi \wedge E_c(\varphi)$ implies both $\neg K(\varphi)$ and $\neg K(\neg\varphi)$. First, unpack the definition of $\neg K(\varphi)$ into $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$. The second element of the disjunction $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$ follows directly from the second element of the conjunction $B(\varphi) \wedge \neg\varphi \wedge E_c(\varphi)$, proving $\neg K(\varphi)$. Now, unpack the definition of $\neg K(\neg\varphi)$ into $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$. Due to the consistency of beliefs $B(\varphi) \rightarrow \neg B(\neg\varphi)$. From this fact and the first element of the conjunction $B(\varphi) \wedge \neg\varphi \wedge E_c(\varphi)$, it follows that $\neg B(\neg\varphi)$. This fact directly implies $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$, due to the first element of such disjunction. This proves $\neg K(\neg\varphi)$. $\square$

Negative Belief Perseverance. *Negative belief perseverance* is the stance that describes an agent whose false belief holds even in the light of evidence contradicting their beliefs. This stance is in place when there are phenomena

[5] From now on, the stances are presented employing only $B(\varphi)$. Obviously, the same considerations would be true employing, *mutando mutandis*, $B(\neg\varphi)$.

such as the backfire effect [15], where agents persevere on their beliefs (or even strengthen them) even after being exposed to evidence that point to their beliefs being false.

Example 5 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_2, w_3\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_1)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models E_c(\neg p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge \neg p \wedge E_c(\neg p)$ holds, showing that a *negative belief perseverance stance* is present.

Theorem 5 (From negative belief perseverance to Ignorance). *A negative perseverance stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \neg\varphi \wedge E_c(\neg\varphi)) \rightarrow I(\varphi) \tag{5}$$

Proof. The proof of theorem 5 is equivalent to that of theorem 4. □

Credulity. *Credulity* is the stance that describes an agent who holds a belief even though s/he only has limited evidence for the fact that s/he believes and such belief is indeed false. This happens in situations in which an agent forms beliefs even on grounds of limited evidence pieces. Note that this might be warranted in some situations, especially those where gaining conclusive evidence is hard and only the limited version of evidence is available. It is important to stress, however, that the agent must realize that even though s/he holds the beliefs, s/he is still ignorant about the fact (i.e., there must be a higher-level realization that ignorance might be present).

Example 6 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_2, w_3\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_1), (w_2, w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models \neg E_c(p)$
- $(\mathfrak{M}, w_1) \models E_l(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge \neg p \wedge E_l(p) \wedge \neg E_c(p)$ holds, showing that a *credulity stance* is present.

Theorem 6 (From Credulity to Ignorance). *A credulity stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \neg\varphi \wedge E_l(\varphi) \wedge \neg E_c(\varphi)) \rightarrow I(\varphi) \tag{6}$$

Proof. The proof of theorem 6 is equivalent to that of theorem 4. □

Misbelief. *Misbelief* is the stance that is closest to the misbelieving effect given in [1]. Someone is in a misbelieving stance if s/he holds a false beliefs which is based on no evidence whatsoever. This is typical of situations in which agents believe unjustified myths, either in the form of prejudices or simply due to irrational thinking. In those situations, agents often hold false beliefs based on various form of biases. This is also common in situations in which unconscious beliefs are held unknowingly from the agents.

Example 7 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3, w_4)$, $\pi(p) = \{w_2, w_3\}$, $R_B = \{(w_1, w_2), (w_1, w_3)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_2), (w_3, w_4)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models \neg p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models \neg E_l(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge \neg p \wedge \neg E_l(p)$ holds, showing that a *misbelief stance* is present.

Theorem 7 (From Misbelief to Ignorance). *A misbelief stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \neg\varphi \wedge \neg E_l(\varphi)) \rightarrow I(\varphi) \tag{7}$$

Proof. The proof of theorem 7 is equivalent to that of theorem 4. □

Note that the proofs of theorems 4, 5, 6 and 7 are all equivalent. This is because it is the fact that something false is believed that implies ignorance. Therefore, this class of stances represent a failure of the conjunction of the first two components of the tripartite definition of knowledge. As with the previous class, also in this case the evidence component is useful only as far as it explains the cognitive stance of the agent who is subject to the false belief and could therefore provide insights into how to treat his/her ignorance properly.

The last class of stances that will follow is what constitutes the major advancement from the work in [1]. In such work, the doubting effect what formalized as having a true belief of a fact that was not known. With the addition of evidence in the language, such lack of knowledge can be explained rather than being assumed. I would like to stress again that all the stances that will follow might apply, *mutando mutandis*, to the same scenarios that were just introduced. Again, the difference in those scenarios is just the truth of the fact examined, which is often not a directly accessible feature in the real world. However, the examples proposed will show exemplary cases in which it is likely that those stances are present instead of the ones just introduced. True, in practical terms it would be difficult to prove that one stance is present instead of the other, but it is hoped that, in the future, further studies on how evidence is gathered could help in discerning the stances.

Positive Belief Perseverance. *Positive belief perseverance* is the stance that obtains when someone holds on to his/her beliefs even when presented with false conclusive evidence. While this stance is desirable when observed from the outside, i.e., the agent is able to resist the false conclusive evidence and is indeed correct in doing so (because what s/he believes is indeed true), from a subjective perspective, it is as troublesome as the negative belief perseverance stance. For instance, if a scientist has a firm belief in his theory (which turns out to be true), s/he might reject the conclusive evidence s/he is presented against it, even though good scientific practices would require him/her to at least consider it while forming his/her beliefs. Now, even though in the future it might turn out that s/he was correct in resisting such conclusive evidence (because it was false), this would still not justify his/her behaviour when the evidence was received.

Example 8 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_1), (w_1, w_2)\}$, and $\mathcal{E}(w_1) = \{(w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models E_c(\neg p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge p \wedge E_c(\neg p)$ holds, showing that a *positive belief perseverance stance* is present.

Theorem 8 (From positive belief perseverance to Ignorance). *A positive belief perseverance stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \varphi \wedge E_c(\neg\varphi)) \rightarrow I(\varphi) \tag{8}$$

Proof. Unpack the definition of $I(\varphi)$ into $\neg K(\varphi) \wedge \neg K(\neg\varphi)$. To prove theorem 8, it is necessary to show that $B(\varphi) \wedge \varphi \wedge E_c(\neg\varphi)$ implies both $\neg K(\varphi)$ and $\neg K(\neg\varphi)$. First, unpack the definition of $\neg K(\neg\varphi)$ into $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$. The second element of the disjunction $\neg B(\neg\varphi) \vee \varphi \vee \neg E_c(\neg\varphi)$ follows directly from the second element of the conjunction $B(\varphi) \wedge \varphi \wedge E_c(\neg\varphi)$, proving $\neg K(\neg\varphi)$. Now, unpack the definition of $\neg K(\varphi)$ into $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$. Due to the consistency of conclusive evidence (such property is easily provable using definition 4 for E_c) $E_c(\neg\varphi) \rightarrow \neg E_c(\varphi)$. From this fact and the third element of the conjunction $B(\varphi) \wedge \varphi \wedge E_c(\neg\varphi)$, it follows, by *Modus Ponens*, that $\neg E_c(\varphi)$. This fact directly implies $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$, due to the third element of such disjunction. This proves $\neg K(\varphi)$. □

Doubt. *Doubt* is possibly the stance that is closest to the doubting effect presented in [1]. A doubt stance is present when an agent believes something which is true but only on the ground of limited evidence. Obviously, such limited evidence is not sufficient for knowledge to be present, but could constitute a good starting point for the agent to indeed form such knowledge. This is common of many scientific practices where conclusive evidence is sought, but it is still lacking. In fact, it could be claimed that science in itself is the practice of looking for conclusive evidence for hypothesis that are currently based only on limited forms of evidence. Thus, it could be fairly safe to assume that a doubting stance is present in each scientist that is performing his/her work properly.

Example 9 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_1), (w_1, w_2)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_2), (w_3)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models E_l(p)$
- $(\mathfrak{M}, w_1) \models \neg E_c(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge p \wedge E_l(p) \wedge \neg E_c(p)$ holds, showing that a *doubt stance* is present.

Theorem 9 (From Doubt to Ignorance). *A doubt stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \varphi \wedge E_l(\varphi) \wedge \neg E_c(\varphi)) \rightarrow I(\varphi) \tag{9}$$

Proof. Unpack the definition of $I(\varphi)$ into $\neg K(\varphi) \wedge \neg K(\neg\varphi)$. To prove theorem 9, it is necessary to show that $B(\varphi) \wedge \varphi \wedge E_l(\varphi) \wedge \neg E_c(\varphi)$ implies both $\neg K(\varphi)$ and $\neg K(\neg\varphi)$. To prove $\neg K(\neg\varphi)$ a procedure similar to that employed for theorem 8 can be employed. Now, unpack the definition of $\neg K(\varphi)$ into $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$. The third element of the disjunction $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$ is directly implied by the fourth element of the conjunction $B(\varphi) \wedge \varphi \wedge E_l(\varphi) \wedge \neg E_c(\varphi)$, proving $\neg K(\varphi)$. □

Intuition. *Intuition* is a stance that is present whenever an agent holds a belief that is true without any form of evidence whatsoever. This kind of stance is typical of early stages of research in which a scientist might form a belief in the truth of a fact (which is indeed true) based on intuition alone and then proceeds to seek evidence to corroborate or falsify such belief.

Example 10 (Formal model example). Take the following model $\mathfrak{M} = (W, \pi, R_B, \mathcal{E})$ where $W = (w_1, w_2, w_3, w_4)$, $\pi(p) = \{w_1, w_2\}$, $R_B = \{(w_1, w_1), (w_1, w_2)\}$, and $\mathcal{E}(w_1) = \{(w_1, w_3), (w_2, w_4)\}$. It is easy to check that:

- $(\mathfrak{M}, w_1) \models p$
- $(\mathfrak{M}, w_1) \models B(p)$
- $(\mathfrak{M}, w_1) \models \neg E_l(p)$

From the above satisfiability relations, it follows that in the pointed model $(\mathfrak{M}, w_1)$, the formula $B(p) \wedge p \wedge \neg E_l(p)$ holds, showing that an *intuition stance* is present.

Theorem 10 (From Intuition to Ignorance). *An intuition stance implies basic ignorance. Formally:*

$$(B(\varphi) \wedge \varphi \wedge \neg E_l(\varphi)) \rightarrow I(\varphi) \tag{10}$$

Proof. Unpack the definition of $I(\varphi)$ into $\neg K(\varphi) \wedge \neg K(\neg\varphi)$. To prove theorem 10, it is necessary to show that $B(\varphi) \wedge \varphi \wedge \neg E_l(\varphi)$ implies both $\neg K(\varphi)$ and $\neg K(\neg\varphi)$. To prove $\neg K(\neg\varphi)$ a procedure similar to that employed for theorem 8 can be employed. Now, unpack the definition of $\neg K(\varphi)$ into $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$. Due to the implication from $\neg E_l(\varphi)$ to $\neg E_c(\varphi)$ (such property is easily provable using contrapposition and definition 4 for E_l and E_c), it follows from the third element of the conjunction $B(\varphi) \wedge \varphi \wedge \neg E_l(\varphi)$ by *Modus Ponens* that $\neg E_c(\varphi)$. This fact directly implies $\neg B(\varphi) \vee \neg\varphi \vee \neg E_c(\varphi)$, due to the third element of such disjunction. This proves $\neg K(\varphi)$. $\square$

4 Conclusion and Future Works

In this paper, a new interpretation of an existing language for evidence-based beliefs [4, 5] has been presented. This new interpretation has then been employed to define knowledge as justified true belief. Such formal language has then been employed to describe various cognitive stances that lead to ignorance. Those stances are believed to improve the understanding already given in [1] about the relationship between doxastic cognitive stances and ignorance. Each stance has been described and examples have been given from potential scenarios in the real world where the stance might be present. Then, it has been shown how each of the stances imply ignorance. In the future, two main venues of research might be pursued: i) the language of *JTB* could be used to explore cognitive stances that inhibit or produce higher-order levels of ignorance (e.g., ignoring to ignore); moreover, ii) the language could be augmented with dynamic operators (in the spirit of [5]) to analyse the effects of different actions on the ignorance of the agents described by the language.

References

1. A. Aldini, P. Graziani, M. Tagliaferri, "Reasoning About Ignorance and Beliefs", in: L. Cleophas, M. Massink (eds.), Software Engineering and Formal Methods. SEFM 2020 Collocated Workshops, pp. 214–230, 2021.
2. S. Artemov, M. Fitting, "Justification Logic", The Stanford Encyclopedia of Philosophy (Spring 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2021/entries/logic-justification/.
3. S. Artemov, M. Fitting, Justification Logic: Reasoning with Reason, Cambridge University Press, 2019.
4. J. van Benthem, D. Fernandez-Duque, E. Pacuit, "Evidence Logic: A New Look at Neighborhood Structures", in: M. Kracht, M. de Rijke, H. Wansing, M. Zakharyaschev (eds.), Advances in Modal Logic, pp. 97–118 2012.

5. J. van Benthem, E. Pacuit, “Dynamic Logics of Evidence-Based Beliefs”, Studia Logica, Vol. 99 (61), 2011.
6. S. Bonzio, V. Fano, P. Graziani, M. Pra Baldi, “A Logical Modeling of Severe Ignorance”, Journal of Philosophical Logic, 2022.
7. J. Fan, “A Logic for Disjunctive Ignorance”, Journal of Philosophical Logic, Vol. 50 (6), pp. 1293–1312, 2021.
8. V. Fano, P. Graziani, “A working hypothesis for the logic of radical ignorance”, Synthese, 2020.
9. K. Fine, “Ignorance of ignorance”, Synthese, Vol. 195 (9), pp. 4031–4045, 2018.
10. V. Goranko, “On relative ignorance”, Filosofiska Notiser, Vol. 8 (1), pp. 119–140, 2021.
11. J.Y. Halpern, Reasoning About Uncertainty, The MIT Press, 2005.
12. E. Kubyshkina, M. Petrolo, “A logic for factive ignorance”, Synthese, Vol. 198, pp. 5917–5928, 2021.
13. Plato, Theaetetus, L. Campbell (transl.), Clarendon Press, 1883.
14. R. Rasmus, J. Symons, “Epistemic Logic”, The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.), URL = ¡https://plato.stanford.edu/archives/sum2021/entries/logic-epistemic/¿.
15. B. Swire-Thompson, J. DeGutis, D. Lazer, “Searching for the Backfire Effect: Measurement and Design Considerations”, Journal of Applied Research in Memory and Cognition, Vol. 9 (3), pp. 286–299, 2020.