Enhancing Doxastic Logic with Dialetheism: A Dual Operator Approach

Stanisław Wasilewski

Advisor: Vera Stebletsova

Abstract

This paper explores doxastic logic, focusing specifically on the KD45 system, which provides a framework for understanding the beliefs of ideal agents who are logically omniscient. However, the characteristics of KD45 agents differ significantly from human reasoning, leading to challenges in modeling realistic belief systems. Humans are not all-knowing and often hold contradictory beliefs, necessitating modifications to the traditional KD45 framework to represent human cognitive processes more accurately.

Inspired by concepts such as dialetheism and paraconsistent logic, which allow for true contradictions and managing inconsistencies, this paper proposes an extension to KD45 doxastic logic by introducing a dual belief operator M, representing "belief as possibility." This operator coexists with the existing "belief as necessity" operator B, allowing for a more accurate and flexible model where strong beliefs and weak beliefs can coexist.

By dualizing the KD45 logic, this paper addresses the logical omniscience problem and offers a method to simulate more realistic agents that can manage conflicting beliefs without leading to logical explosions. This extended framework aims to bridge the gap between the idealized reasoning of KD45 agents and the error-prone, often contradictory, belief systems seen in human reasoning, thereby contributing to the development of realistic doxastic logic-based models.

1 Introduction

Doxastic logic, as a subset of modal logic, is designed to formalize and reason about the beliefs of agents. By extending the elements of propositional logic with specialized unary operators, doxastic logic enables us to articulate advanced belief systems. The central operator in this logic, typically denoted as B_α, stands for the belief of an agent α in a proposition.

The main focus of this paper is on KD45 doxastic logic, which introduces a structured framework for understanding and analyzing the belief systems of ideal agents — agents that are logically omniscient, consistent, and capable of deducing consequences from their beliefs. This form of doxastic logic makes several strong assumptions about agents, such as their inability to hold contradictory beliefs and their capability to know all logical truths and their implications. While this allows for a clean and mathematically rigorous model, it differs significantly from the actual cognitive processes of human beings, who often hold contradictory beliefs and are bound by limitations in memory and computational power.

Logical omniscience raises essential questions about the objectives of doxastic logic. If the goal is to model realistic human reasoning, logical omniscience appears as a significant limitation, as it fails to account for humans' cognitive inefficiencies and inconsistencies. On the other hand, if the aim is to construct an idealized version of reasoning free from such human limitations, logical omniscience may be seen as advantageous.

To address the gap between ideal logical agents and human-like reasoning, this paper explores modifications to the KD45 framework. Dialetheism is discussed - a philosophical perspective that accepts true contradictions. Moreover, the paraconsistent logic approach is examined, where agents can handle contradictory beliefs without leading to an explosion of inconsistencies. The concept of dualizing the belief operator in KD45 by introducing a weak belief operator M (which stands for 'belief as possibility') offers a pragmatic extension to the existing model. This dualization allows for a more accurate representation of beliefs, where strong beliefs (necessity) and weak beliefs (possibility) can coexist, thus providing a more flexible and realistic model of human-like reasoning.

In the following sections, I will detail the syntax and semantics of KD45 doxastic logic, analyze the problem of logical omniscience, and propose modifications for better alignment with human reasoning. I will explore paraconsistent logics and the introduction of dual belief operators as practical solutions, concluding with the implications these extensions have for creating more realistic agents while maintaining logical rigor.

2 Doxastic Logic

2.1 General Description

Doxastic logic is a logic that allows to express beliefs of individual agents. It is an augmentation of propositional logic with a unary operator $B\alpha$. One of the most popular doxastic logics, and the logic that I am concerned with the most, is logic KD45.

2.2 Syntax

The syntax of doxastic logic consists of the set of all propositional connectives based on negation and conjunction. Moreover, doxastic logic introduces a unary operator B_α, where α is an agent belonging to the defined set of all agents Agt. $B_\alpha\varphi$ can be therefore translated as "Agent α believes that φ is true."[Fag+04a]

The BNF notation of a well-formed doxastic logic formula can be defined as follows:

$$\phi :: \top | \perp | p | \neg\varphi | \varphi \wedge \varphi | B_\alpha\varphi$$

where p is an atomic formula and φ is a syntactic category of a well-formed formula.
Other propositional logic operators, such as disjunction, implication, or the Sheffer stroke are also a part of the syntax as long as they can be expressed using negations and conjunctions.

Quantifiers are not a part of $KD45$ doxastic logic, although they can be expressed using the $B_\alpha\varphi$ operator: $B_\alpha\varphi$ can be regarded as a universal quantification across all worlds w accessible to an agent α. In first-order propositional logic, the existential quantifier $\exists x\varphi$ can be defined with the universal quantifier $\forall x\varphi$ and a negation: $\neg\forall x\neg\varphi$. In the case of doxastic logic, the formula $\neg B_\alpha\neg\varphi$ creates an existential quantification, stating that some world w exists such that $w \models \varphi$.

2.3 Semantics

The semantics of doxastic logic is expressed with Kripke models.[Fag+04a] A Kripke model M is a tuple:

$$M = \langle W, R_1, ..., R_n, \pi\rangle$$

where W is a non-empty set of worlds w_i, $R_j (j = 1, ..., n)$ is an insistinguishability relation in $W \times W$, and π is an interpretation, which is a truth assignment of atomic formulas in given worlds W.

Following properties of relations R_i can be distinguished:

- Serialiseness: R_i is serial when for every world $w \epsilon W$ there exists $w' \epsilon W$ such that wR_iw'
- Transitiveness: R_i is transitive when for all $\{w, w', w''\}\epsilon W$ such that wR_iw' wR_iw'' holds.
- Reflexiveness: R_i is reflexive when for all $w \epsilon W$ such that wR_iw holds.
- Symmetricity: R_i is symmetric when for all $\{w, w'\}\epsilon W$ such that wR_iw' $w'R_iw$ holds.
- Euclideaness: R_i is Euclidean when for for all $\{w, w', w''\}\epsilon W$ such that wR_iw' and wR_iw'', $w'R_iw''$ holds.

The indistinguishability relation is an equivalence relation, thus has to be reflexive, symmetric, and transitive.

It is impossible for an agent to have contradicting beliefs. Therefore the belief operator $B_\alpha\varphi$ holds if and only if the agent is connected only to worlds where φ is true. This can be expressed as follows:

$$M, w \models B_\alpha\varphi \text{ iff } (M, v) \models \varphi \text{ for all } v \text{ in } wRv$$

where M is the model, w and v are worlds and elements of the set of all worlds W, and R is the indistinguishability relation in $W \times W$.

2.4 Axioms

KD45 has axioms of propositional logic and various modal logics:

- **(P)** All propositional logic axioms and tautologies.
- **(D)** $B_\alpha\varphi \longrightarrow \neg B_\alpha \neg\varphi$
- **(4)** $BB_\alpha\varphi \longrightarrow B_\alpha\varphi$
- **(5)** $\neg B_\alpha\varphi \longrightarrow B_\alpha\neg B_\alpha\varphi$
- **(R1)** $\vdash \varphi, \vdash \varphi \longrightarrow \psi \implies \vdash \psi$
- **(R2)** $\vdash \varphi \implies B_\varphi$

  It is important to state that beliefs do not have to be true. Therefore axiom **T**: $B_\alpha\varphi \longrightarrow \varphi$ of epistemic logic does not hold for every α and ϕ in doxastic logic.

3 Logical Omniscience

3.1 Ideal Agents

In doxastic logic KD45, agents are assumed to be perfect logical reasoners who adhere to all axioms and respect inference rules.[Fag+04b] They are, therefore, ideal agents. Ideal agents, or omniscient agents, are usually defined by the following two criteria:

- An ideal agent knows all logical truths associated with the logical system in which they operate.
- An ideal agent knows all consequences of these truths and of the given information.

3.2 Logical Omniscience

The problem of Logical Omniscience depends on how we perceive the purpose of doxastic logic.

Assuming that doxastic logic is meant to model how humans reason with their beliefs, logical omniscience renders this representation to be highly not representative. Humans are often victims of restrictions associated with memory and computational capabilities. Moreover, humans are able to exist while having contradicting beliefs, which would be unheard of for an agent.

If we assume that doxastic logic is a tool that would allow us to not fall victim to these restraints, omniscience might be desirable, as it would be a solution to our flawed reasoning system.

However, even with this assumption, the capabilities of artificial agents are also limited in computational power and memory. The only difference that separates these agents from humans is the impossibility of a contradiction. Therefore, logical omniscience seems to not be a problem only in theoretical circumstances.

Different solutions have been proposed to tackle this problem. These solutions often involve restricting the agents on purpose, either in their memory or in computational power. There have been also other ideas, such as "impossible possible worlds" proposed by Hintikka in 1978, which would cause agents to not distinguish between contradicting worlds in certain situations. Despite all efforts, the problem of logical omniscience remains a topic of discussion within the scientific community to this day.

3.3 Realistic Agents

I will assume modeling human thinking as the goal of doxastic logic for the purpose of this paper. In that case, a non-omniscient agent would be desirable. The following question arises: What qualities, or rather flaws, should such a non-ideal agent possess?

As mentioned before, the main difference between an ideal doxastic logic agent and a human person is that a person can have contradictory beliefs. An example of this could be a person, whose belief set consists of the two following statements:

1. $\phi : p$ "It is wrong to ever lie to another human"

2. $\psi : q \rightarrow \neg p$ "It is good to lie to another human if it avoids hurting their feelings"

If propositional logic is applied to this belief system and we take ϕ and ψ to be true, as well as q to be true (in the event that someone's feelings might be hurt and it can be avoided by lying), then the following inference can be made:

$$\begin{array}{rl} & q \rightarrow \neg p \\ & q \\ & p \\ \hline \therefore & p \wedge \neg p \end{array}$$

This inference results in a contradiction. Ex contradictione quodlibet - from a contradiction anything follows, which renders the rest of the belief system unusable. This happens via the principle of explosion[Smi20]:

$$\begin{array}{rl} & p \\ & \neg p \\ & p \vee q \\ \hline \therefore & q \end{array}$$

If this would be inferred by an ideal agent, any further sentence can be proven to be true or false. In the case of humans, however, this phenomenon does not occur. A human person can either see the contradiction and fix their belief system by eliminating one of the contradictory sentences, or, what is more important, ignore the contradiction and continue existing with a non-coherent system of beliefs.

I believe that the ability to ignore contradictions while still using the belief system is the most important quality that a desirable non-ideal agent should possess.

4 Dialetheism

4.1 The Idea of Dialetheism

I will use the idea of Dialetheism to allow for the assumption that it is possible to have a belief system consisting of both ϕ and $\neg\phi$ for some ϕ. This idea, supported by philosophers such as Jc Beall, Graham Priest, and Richard Routley, states that it is possible for a statement to be true, when the negation of that statement is also true. A statement of such kind is then called a "dialetheia".[Whi04]

A known example of a dialetheia is as follows: "A person is standing in the doorframe of a room. This person is in the room and not in the room at the same time."

Dialetheism, besides allowing for the paradoxical nature of natural language, might also present a solution to more formal paradoxes, such as the liar's paradox; "This sentence is a lie". Under dialetheism, the statement in question could be true and false at the same time.

It could be argued that dialetheias arise from the ambiguity of natural language. Furthermore, from the standpoint of propositional logic, they might not really exist. However, I believe that this part of the philosophical debate around dialetheism does not matter when discussed in the current framework. Natural language allows for dialetheias. Therefore, for the purpose of this paper, it is irrelevant whether dialetheias are formally impossible. Rather, to achieve non-ideal agents, a discourse on how to simulate dialetheism would be more applicable.

4.2 Paraconsistent Logics

Paraconsistent Logic as a family of non-classical logics can be seen as an implementation of dialetheism in logic. Paraconsistent logic denies the principle of explosion and allows for inconsistency. There are many paraconsistent logics which could be potentially considered as an alternative basis for doxastic logic.

An attempt at a direct implementation of dialetheism to logic is dialetheic logic (dLP), proposed by H. Omori [Omo16]. in 2016. It is specifically designed to handle true contradictions and combines connectives from different traditions of paraconsistent logics. Dialetheic logic is based on the Logic of Paradox (LP) developed by Graham Priest. LP accepts that contradictions can exist and be true, providing a formal system where contradictions do not lead to explosion.

To allow for true contradictions, dLP introduces two distinct negation symbols: $\sim$ and $\circ$. A well-formed dLP formula can be defined as follows:

$$\phi :: \top | \perp |p| \sim \varphi | \circ \varphi | \varphi \wedge \varphi | \varphi \vee \varphi | \varphi \rightarrow \varphi \ | \exists x A | \forall x A$$

where p is an atomic formula and φ is a syntactic category of a well-formed formula.

While dialetheic logic could potentially be a solution to the problem of omniscience in doxastic logic, it also poses challenges. The necessary restrictions to keep the dLP consistent limit the inference rules of this system.

Assuming that the goal of doxastic logic is to model human thinking, one could require individual agents to differ in how far they are from being omniscient, effectively varying their mental capabilities. When applied to doxastic logic, dLP could render agents more non-ideal than necessary, not offering any flexibility in their inference capabilities.

5 Necessity and Possibility

Although it is intuitive to represent the idea of dialetheism with dialetheic logic, or any other paraconsistent logic, it might be pragmatically more feasible to emulate this idea without leaving the system of normal, propositional logic. A more viable approach would be to allow for paraconsistent beliefs in an agent, while not allowing an explosion of the belief system to occur. In other words, two conflicting beliefs, although being believed by the same agent, should not be a part of one world in a Kripke model.

5.1 Dualization of KD45

The basic interpretation of a belief in KD45 is as follows: *When agent α considers φ to be true, α assumes that φ is true in all possible worlds.* This interpretation can be called *Belief as necessity*, as it is necessary for an agent to believe a belief φ in every possible world accessible to agent α.

An alternative to that, proposed by W. van der Hoek and J.-J. Ch. Meyer [HM89] is the *belief as possibility*, which interprets a belief as true in at least one possible world. The most important feature of that new understanding of belief is that it can function as a dual operator M of belief as necessity B. M, the dual operator of B would be defined as follows: $M = \neg K \neg$. Thus, if $B\neg\varphi$ is not true, that is, if there is at least one possible world where φ is true, $M\varphi$ is true. The literal meaning of $M\varphi$ would be: *It is possible that φ is true.* [HM89]

Both these operators can be present in the same coherent system, allowing for two different kinds of beliefs within the same agent. This effectively classifies beliefs into two categories: strong beliefs denoted by B, and weak beliefs, denoted by M.

Dualization of KD45 requires extensions to already included axioms. Each axiom in 2.4 is extended for the weak belief operator M in the following way:

- **(P)** Is not changed as it does not mention the belief operator B.
- **(D)** $M_\alpha\psi \longrightarrow (M_\alpha\varphi \vee M_\alpha(\neg\varphi \wedge \psi))$
- **(4)** $MM_\alpha\varphi \longrightarrow M_\alpha\varphi$
- **(5)** $\neg M_\alpha\varphi \longrightarrow M_\alpha\neg M_\alpha\varphi$
- **(R1)** Is not changed as it does not mention the belief operator B
- **(R2)** $\vdash \varphi \implies \vdash \neg M \neg\varphi$

Axiom D is not modified in the same way as other axioms that mention B, as $M_\alpha\varphi \longrightarrow \neg M_\alpha\neg\varphi$ does not align with the idea of belief as possibility: If $M_\alpha\neg\varphi$, there might be worlds accessible to α, where φ is true, and therefore $M_\alpha\varphi$ might also hold.

It is worth noting that doxastic logic KD45 with operator M resembles modal logic K. B and M correspond to $\Box$ and $\diamond$, respectively denoting necessity and possibility.

5.2 Escaping the principle of explosion

Doxastic logic KD45 extended by the operator M satisfies the requirement of allowing for inconsistent beliefs while preventing the principle of explosion. An agent α can have two beliefs: $M_\alpha\varphi$ and $M_\alpha\neg\varphi$ at the same time. As long as φ and $\neg\varphi$ are in different worlds, there is never a situation, where $\varphi\wedge\neg\varphi$ is believed to be true, whether by M or by B.

It could be argued that the extension of $KD45$ by M does not allow for inconsistencies, but rather for beliefs that would be inconsistent if they were believed in the same world. Thus, this new system would not be suitable to be used to represent truly realistic agents. I believe, however, that it is not feasible, or even possible, to design a system that would be a perfect model of a human reasoner. Simulating the phenomenon of inconsistent beliefs while still allowing for inferencing capabilities using a completely deterministic system is a problem of resemblance, not perfection.

5.3 Two belief operators in one system

One of the problems of paraconsistent logic was a lack of flexibility when it comes to balancing the inference power of agents with their capability to possess contradicting beliefs, as mentioned in 4.2 The extended version of KD45 grants the designer of an agent this helpful feature. While assigning an agent a set of beliefs, one can choose an appropriate belief operator B or M. Using B is most likely to yield the best results while simulating robust reasoners capable of solving various puzzles. On the other hand, if one aims to to simulate a human who might be confused about a certain topic, M might be more appropriate. Conversely, if the purpose of the agent requires the best of both worlds, constructing a belief system that utilizes both B and M allows for a non-omniscient agent that would be powerful enough for this goal.

6 Conclusion

In conclusion, the exploration of doxastic logic, with a particular focus on the KD45 system, has highlighted significant challenges posed by logical omniscience. This concept asserts that agents know all logical truths and the following consequences. This characteristic greatly diverges from how actual human reasoning works. Humans are not all-knowing and often hold inconsistent or contradictory beliefs, further distancing real-world reasoning from the ideals of KD45 logic.

To address the discrepancy between ideal agents in doxastic logic and the more fallible reasoning processes of humans, concepts of dialetheism and paraconsistent logics were explored. Dialetheism adopts the notion that certain contradictions can coexist without resulting in logical explosion, providing a framework that allows for more realistic modeling of human beliefs. However, the restrictive nature of paraconsistent systems such as dLP limits their pragmatical value in simulating nuanced belief systems that balance consistency and reasoning capabilities.

A potential solution was considered via the introduction of the dual operator M of B in KD45 logic. This dualization allows for the *belief as possibility* interpretation, presenting an alternative way to handle agent beliefs. By implementing both necessity (B) and possibility (M) beliefs, we are able to build a more flexible model that can represent both strong and weak beliefs within the same system. This dual approach respects the prevention of the principle of explosion, permitting the coexistence of beliefs like $M_\alpha\varphi$ and $M_\alpha\neg\varphi$ without falling prey to inconsistency.

Ultimately, while the perfect emulation of human reasoning within a deterministic logical system remains an elusive objective, the extended KD45 logic with the dual operators B and M offers a promising approach to narrowing the gap. It provides a more realistic portrayal of agents who can manage contradictory beliefs pragmatically, thus allowing us to simulate more lifelike reasoning processes. As the landscape of artificial intelligence and cognitive modeling continues to evolve, the pursuit of ever more accurate and realistic representations of human thought will undoubtedly benefit from these kinds of logical extensions and the understanding they bring to the field.

References

[HM89] W. van der Hoek and J.-J. Ch. Meyer. "POSSIBLE LOGICS FOR BELIEF". In: *Logique et Analyse* 32.127/128 (1989), pp. 177–194. ISSN: 00245836, 22955836. URL: http://www.jstor.org/stable/44084271 (visited on 06/12/2024).

[Fag+04a] Ronald Fagin et al. *Reasoning About Knowledge*. The MIT Press, Jan. 2004. ISBN: 9780262256094. DOI: 10.7551/mitpress/5803.001.0001. URL: https://doi.org/10.7551/mitpress/5803.001.0001.

[Fag+04b] Ronald Fagin et al. *Reasoning About Knowledge*. The MIT Press, Jan. 2004, pp. 333–373. ISBN: 9780262256094. DOI: 10.7551/mitpress/5803.001.0001. URL: https://doi.org/10.7551/mitpress/5803.001.0001.

[Whi04] Bruno Whittle. "Dialetheism, Logical Consequence and Hierarchy". In: *Analysis* 64.4 (2004), pp. 318–326. DOI: 10.1093/analys/64.4.318.

[Omo16] H. Omori. "From Paraconsistent Logic to Dialetheic Logic. In: Andreas, H., Verdée, P. (eds) Logical Studies of Paraconsistent Reasoning in Science and Mathematics." In: *TUG-Boat* 45 (2016), pp. 111–134.

[Smi20] Peter Smith. *An Introduction to Formal Logic*. 2nd ed. Cambridge Introductions to Philosophy. Cambridge University Press, 2020.