Extending Dynamic Epistemic Logic to Express Misinformation

by

Julian Chow

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Philosophy

University of Alberta

Abstract

After a qualitative analysis of deception, I select the component of misinformation for encoding into a logic. I introduce dynamic epistemic logic (DEL) and the Cards scenario (Cards), then modify Cards to illustrate shortcomings of DEL in expressing misinformation. I develop two dynamic epistemic logics of misinformation as extensions of DEL. Each is distinguished by its fibbee protocol, a new sub-step of model update accommodating agents' internally consistent reaction to observing misinformation events. I prove that for two-valued logic both logics of misinformation are equivalent. Finally, I propose a dynamic epistemic logic of trust. I conclude with a discussion of the interdependence of misinformation and trust in the context of modal epistemic logics.

Preface

This thesis is an original work by Julian Chow. No part of this thesis has been previously published.

Acknowledgements

Above all I express my deepest gratitude to my supervisor, Katalin Bimbo, for her incredible kindness and patience in guiding me and permitting me much freedom throughout the tenure of this project. I thank Allen Hazen for organizing a continuing reading group which provided a fertile environment for the development of my own ideas. I am also grateful to David DeVidi for his encouragement throughout my program.

Funding for this project was graciously provided jointly by the Social Sciences and Humanities Research Council (SSHRC) through the SSHRC Joseph-Armand Bombardier Canada Graduate Scholarship - Master's (CGS-M) program and by the University of Alberta through the Walter H. Johns Graduate Fellowship.

Contents

List of Figures

1 The anatomy of deception: Intentionality, misinformation, and trust

In *An Essay in Modal Logic*, von Wright presented various interpretations of modal logic.[13] One such interpretation is the epistemic interpretation, which Hintikka felt was best matched with the modal logic S5.[7, 13]

Public announcement logic (PAL) followed much later, motivated by the prospect of an epistemic logic capable of expressing not only static snapshots of epistemic states but also the dynamism of epistemic interactions between knowers.[11] PAL was a predecessor of dynamic epistemic logic (DEL), which further generalized on the tools with which epistemic events are expressed.[6] DEL was a major advancement in the variety of events expressible using modal logic. However, even DEL has restrictions on the types of events that it can express.

PAL assumes that the informational content of any event must be true and received by all agents.[6, 11] DEL improves on PAL by dropping the latter assumption.[6] By doing this, DEL is able to capture events like secret communications. However, the first assumption remains.

My tasks in this investigation are to analyze the effects of dropping the assumption of event truth on DEL, demonstrate that DEL requires extension in order to successfully drop that assumption, and introduce just such an extension of DEL that thereby is able to capture events like misinformation. I'll also touch upon the logical relationship between trust and misinformation, and suggest how DEL might be further extended to capture trust.

For now let's loosely discuss the components of deception. We'll soon see that analysis of the components of deception helps us clarify precisely some required characteristics of a logic of misinformation.

1.1 What is deception?

In this section I argue that there are at least three components of deception. These are intentionality, misinformation, and trust. My arguments follow in part from practical considerations for formalizing deception using modal epistemic logic.

1.1.1 Actors and receivers

We distinguish between *acts of deception* and *being deceived.* In any example of deception there is some information involved, which may be true or may be false.[1] Let's consider each

[1]For this investigation we will exclusively use two-valued logic. Later we will see how this decision has interesting ramifications for our new logics.

of the points of view of the actor and the receiver using an example.

Suppose Bill informs Alice that whales are mammals. Alice didn't know beforehand that whales are mammals, and she's skeptical of what Bill has said.

Has Bill committed an act of deception? If we define deception from the point of view of Bill, the actor, then the act is not an act of deception because Bill was merely relaying true information. But if we define deception from the point of view of Alice, the receiver, then it could be argued that it qualifies as an act of deception. This is because Alice's response to Bill's act would involve revision of her knowledge base under the supposition that Bill has attempted, deliberately or not, to transmit to her false information.

Since our interest is in what should happen to agents when they realize they've been the target of an act of deception, I will define deception from the point of view of the receiver rather than that of the actor. That is, let us consider whether an act is an act of deception on the basis of whether its informational content is inconsistent with the knowledge base of the receiver.

1.1.2 Intentionality

Prioritizing the perspective of the receiver over that of the actor does not mean the deceptive intention of actors is irrelevant. Receivers' knowledge base includes not only propositional knowledge about facts of the world but also about other people. For example, Alice now thinks that Bill is either malicious or misinformed. Which one he is depends on his intention behind the act.

Without further information it is impossible for Alice to adjudicate. At the same time, notice that the difference between an act of deception and an act of mere misinformation is that there is no intent to deceive. From the perspective of the actor those who deliberately lie to those whom they think will believe them are being deceitful while those who fail in their good intentions to communicate truth are not.

This analysis of our example demonstrates that while intentionality has a meaningful impact on the ways agents react to each other's actions, it would technically be possible to set it aside from the perspective of the receiver with their implicit understanding of a disjunction regarding the knowledge base of the actor.

1.1.3 Trust

Epistemic interpretations of S5 have been criticized for assuming that agents possess inhumanly perfect introspective capabilities. This is because S5 is an extension of classical propositional logic, which also bears immediate and (mathematically) unerring conclusions on its input premises. Proofs in classical propositional logic follow deductive inference, in which all that can be proven from the premises are already implicitly understood from the combination of the logic's axioms (or rules, in other presentations) and those premises.[4] It's only due to limitations of human brains that deductive conclusions may not be immediately apparent to us from the combination of a set of axioms and given premises. This becomes even more problematic when such systems are epistemically interpreted. Then, the mathematical precision of S5 inherited from classical propositional logic implicitly adopts the assumption that epistemic agents adhere perfectly to deductive reasoning.

What this means for expressing deception is that deductive agents cannot be misinformed so long as their premises are true. Deductive reasoning permits only valid inference, and hence any falsehood discovered in the conclusion must be due to falsehood already in the premises. But a more interesting notion of misinformation is that where someone's premises are true and yet—namely as a result of seriously considering some information offered by another person—they arrive at a false conclusion.

We might say that a receiver "abandons" deductive reasoning in the special case where they trust the actor. Or we may say that with trust agents take it as granted that that information imparted by trusted agents is true. Had Alice trusted Bill, it would be valid for Alice that whales are mammals on the premises that that's what Bill said and Alice trusts Bill. Now if misinformation could be introduced into an act in some way, such as through an intentional act of transmitting misinformation, this trust means the act would be met with uptake such that the receiver results in holding some false proposition to be true.

This phenomenon, while related to intentionality in that an intentional deceiver deliberately exploits the trust of their interlocutor, should not itself be confused with intentionality. A receiver can become misinformed as a result of an act because they trust the actor, or in other words that the receiver accepts the informational content of the act without holding it to the high standard of validity. This is distinct from the receiver's awareness of whether the actor is acting deceptively, which as we have discussed would involve the receiver moreover revising their knowledge about the actor themselves.

While we will spend more time analyzing trust, we now turn our focus to misinformation.

1.1.4 Misinformation

We define misinformation as acts which informational content is known to be false. The wording here is important because the epistemic context under which the act is considered will affect whether it qualifies as an act of misinformation. Referring back to our simple example, from the perspective of a reasonably informed scientist Bill's act is not an act of misinformation even though from the perspective of Alice it is.

Notice that this definition of misinformation is independent of intentionality. Further, by borrowing from DEL it's possible to develop an extension which expresses misinformation without needing to manage mechanisms for trust.

1.2 What is a (modal epistemic) logic of deception?

So far I've broken down deception into three components: intentionality, trust, and misinformation. What would each of these look like in modal epistemic logic, in particular?

1.2.1 Intentionality

As discussed, in order for misinformation to qualify as deception there must be some intent on the part of the provider of that misinformation to mislead one or some number of people. If it were not for this, it would be merely misinformation instead.

But even if we mandate that deception must involve intention, that doesn't preclude the circumstance where intentional action on the part of some information issuer towards some target is observed by yet another party independent of the first two. This is especially relevant if it is supposed that this third party is observing the interaction discreetly, that is unbeknownst to the deceiver. This party can be led into a great variety of informational states by those observations, chiefly because the third party's state of knowledge will not inform the particular manifestation of deception by the first.

The complexity of situations involving third parties increases dramatically when we consider that not only do we need to keep track of what the actor thinks that the receivers consider to be the case but also what the receivers themselves consider, as well as what the receivers consider the actor's intentions are. So the simplicity of the two-agent scenario is misleading with regard to the corresponding simplicity of a logic which captures intentionality of deception.

These issues deserve more time and space than can be afforded in this investigation. As such for our present purposes I'll set aside the task of implementing a logic of intentionality.

1.2.2 Misinformation handling

In epistemic S5, the condition of misinformation would be represented by a condition under which an agent believes both the affirmation and negation of a proposition. Of course, while misinformation is necessary in deception, what would be interesting for a logic of misinformation is less so that it expresses the condition of misinformation than that it expresses a convincing outcome after agents realize that they've arrived at the condition of misinformation.

Let's break misinformation down further into three requirements. These three requirements for a logic of misinformation are:

Firstly, a logic of misinformation has to have some means of agents' detecting a condition of misinformation. On the most simple proposal this would be continuing to simply detect the inconsistent knowledge base, that is the consideration base arrived at after an event update which finds that it now contains contradictory propositions. This has the downside of assuming that the only means by which an agent can arrive at the condition of misinformation is through events, but I think this is a simplification acceptable for a first pass at a logic of misinformation.

Secondly, a logic of misinformation has to be such that its mechanism of misinformation does not result in contradictory propositions in an agent's knowledge base. Following from the first requirement, while such a situation can be used to detect that misinformation transfer has happened, another interesting characteristic of misinformation is that when agents are aware that they've been on the receiving end of misinformation they don't continue to hold consideration in the purported information transmitted by the event.

Thirdly, a logic of misinformation should include some sort of assumed or prescriptive reaction to the detection of one's having arrived at a condition of misinformation. For example, should an agent simply ignore events of misinformation transfer? Perhaps an agent could take an event of misinformation as an opportunity to learn, by instead accepting the negation of its purported information into their knowledge base? I will explore both of these options.

The primary objective of this investigation is to present a logic of misinformation. As a smaller side project I will pitch a suggestion for formalizing trust which is also extended from dynamic epistemic logic.

1.2.3 Mutable trust

As previously established, in order for a condition of misinformation to exist there must be some preexisting mechanism of trust. In contrast, for deductive agents the only way to arrive at a condition of misinformation would be to begin from mistaken premises.

Trust can be formalized using either a static or a dynamic mechanism. I suppose that in general it is more interesting to observe the developing relationships of trust between agents over time in a logic of deception. After all, this sort of dynamicism is the essential contribution of PAL (and inherited by DEL) to epistemic S5.

Meanwhile a static trust mechanism could be presented as a special case of a mechanism of mutable trust where the rate of change of trust with respect to either truthful or discovered deceitful events is set to zero.

For present purposes I'll start with a static mechanism of trust. When modifying DEL to accommodate misinformation it turns out that static trust emerges automatically as a result of the design of DEL presupposing that the informational content of an event is true. Later I'll show that the endeavor reveals a deeper symbiosis between misinformation and trust. At least in modal epistemic logic, misinformation and trust are two sides of the same coin.

1.3 Assumptions and disclaimers

1.3.1 Atomic events

Traditionally, logic is understood as working with certain assumed atoms of a formal language. For example, in the case of classical propositional logic, these atoms are propositions. Generally speaking, the more that a logic is capable of expressing, the more complicated it will be to work with. For example, predicate logic can meaningfully express quantification whereas propositional logic cannot.[4]

The case is similar with PAL and DEL. In PAL, in order to capture relevant considerations in exchanges or transfers of information, it's necessary to include *public announcements* as atomic events, or in other words as single, indivisible units of examination. DEL breaks down the structure of events by introducing *event models*, which combine multiple atomic events into an epistemic structure at the multiple-agents level. Event models are a new type of model in contrast with and in addition to epistemic models.[6]

1.3.2 Static world

I'll also assume that matters of fact remain static. This is an assumption carried forward from as far back as epistemic S5.[2]

It would be an interesting project to see what happens to a logic of deception if facts about the world can change even while agents are interacting with one another. What would happen to the guiding definition of deception? How should agents verify that their consideration bases remain in correspondence with facts about the world?

1.3.3 Knowledge, belief, and consideration

Going forward I'll refer to "consideration" as opposed to "knowledge" or "belief", contrary to more common terminology in the modal logic literature. For example, of an agent g I will say that "g considers the proposition p to be the case" rather than "g knows that p is the case" or "g believes that p is the case". The reason for this is to avoid the thorny issues of whether knowledge and belief can be defined in terms of one another and the nature of the relationship (if any) between logics of knowledge and belief.

As I proceed it will become evident that discussion of misinformation is not dependent on adherence to either of the concepts of knowledge or belief. It will be adequate to simply suppose that for each agent there is some subset of all possible propositions which are "considered" to be the case. For the present discussions it only matters that sometimes these considerations correspond with facts of the world and sometimes they do not.

As such, I'll henceforth refer to an agent's "consideration base" instead of their knowledge base.

1.3.4 Truth

As only a matter of practical convenience I define truth as correspondence with given propositional facts at the designated state, as per the simple correspondence theory of truth. For any given context, those facts are taken as assumptions independent of the epistemic condition of each agent. The simplification by defining truth in this way allows us to focus on misinformation.

1.3.5 Predicate logic

Predicate logic will not be used in the logic of misinformation to be developed here. While predication will be present in some of the formulas analyzed, those formulas will not be further decomposed into predicates and subjects. This is because our primary objective is to lay the groundwork for capturing misinformation, and dynamic epistemic logic serves as a sufficient foundation by itself.[2] Further, the context of misinformation our focus is more on the interactions between agents rather than the grammatical complexity of their utterances.

1.3.6 Event ambiguity

For the sake of this investigation it will be assumed that the epistemic content of any event is unambiguous to all of its observers. In other words, an agent's propensity to understand the same event in a different way than other agents might will not be encoded in this investigation.

[2]I realize that it has already been proven that modal logic is a strictly weaker logic than predicate logic, but this is an irrelevant detail for our purposes.

1.3.7 Common knowledge

A desirable trait of epistemic logics is the ability to express *common knowledge*, which is the agents' collective epistemic state where everybody knows that everybody knows some proposition. This is recursive, that is, it would be true also that everybody knows that everybody knows that everybody knows some proposition, and that everybody knows *that*, and so on.[6]

While there can be misinformation about common knowledge, in this investigation we will focus on the foundations of developing a logic of misinformation first. Therefore we will set aside common knowledge for our present purpose of encoding misinformation of propositions of finite length.

1.3.8 Self-deception

Self-deception is beyond the scope of this investigation because self-deception inherently involves examination of the internal (human) psychological structure of agents. Self-deception is the type of deception where the actor and the receiver are one and the same person. While dynamic epistemic logic addresses inter-agent interactions it does not address *intra*-agent psychologies.

Setting aside self-deception means we can treat agents as black boxes, and that they and their consideration bases stand in a one-to-one correspondence. This is advantageous for a foundational logic of misinformation because what an agent is understood to know can be unambiguously identified without having to invoke the jargon of conscious and unconscious consideration.

Furthermore, self-deception is not to be confused with introspection. The axioms of introspection are epistemic rather than psychological. They express statements in an agent's consideration base about their own consideration base, but are silent about psychological states or motivations.

1.3.9 Artificial intelligence

The artificial intelligence interest, if not applications, of this project lends itself well to speculation on how intelligent machines might fabricate impressions or veneers of fact in order to gain an epistemic advantage over other agents.[1]

In line with the preceding discussions on intentionality and psychological motivation, however, I defer a full investigation of the relevance of DEL and its extensions to artificial intelligence and the prospect of "socially intelligent" machines to future work with the time and greater depth merited by such an important area of application.[1] It would also be prudent to investigate the applicability of formal logic to current research in neural network-style (as opposed to expert system-style) artificial intelligence.

2 An introduction to dynamic epistemic logic

In this chapter I'll introduce dynamic epistemic logic (DEL) and the model $\langle \text{Cards}, JQK \rangle$ (Cards) as a working example. In the process of introducing DEL I'll also cover epistemic S5 and PAL, which are predecessors of DEL.[6]

Cards not only showcases the expressive power of DEL but also illustrates its limitations. This limitation applies to cases where the assumption of noncontradiction between facts and the content of unambiguous communication fails. In other words, DEL is limited in that it cannot express misinformation. Cards is introduced in van Ditmarsch *et al.*'s monograph on DEL, and the alternative events are introduced as end-of-chapter exercise questions.[6]

The major task of my project is to extend DEL to capture basic cases of misinformation, using the requirements for Cards as guidance for the additional capacities needed in a basic logic of misinformation. In this chapter I investigate what is needed, and in the next chapter I implement those discoveries in the development of a logic of misinformation.

2.1 Modal epistemic logic: Epistemic S5

Von Wright proposed using the modal operators of the logic S5 to represent epistemic states about particular propositions.[13] Hintikka is credited with the first major development of modal epistemic logic[3].[7] Kripke (1959) later introduced what is now known as the "possible worlds" interpretation of modal logic, which in the case of epistemic logic lends an intuitive understanding for epistemic structures.[9]

The possible worlds intepretation is particularly useful for understanding epistemic logic because modal operators can be conceptualized as agents' assessing that which must, could possibly, or cannot be the case on the basis of what they consider together with what they observe.

2.1.1 Definitions

The language of epistemic S5 is the set of *formulas* captured by the following Backus-Naur form (BNF)[4]:

$$\phi ::= p|\neg\phi|\phi \wedge \phi|K_g\phi$$

[3]Although Hintikka was against using S5 as an epistemic logic, much work in epistemic logic has now been done using S5. For our purposes here I continue in that tradition.

[4]Where ϕ is an arbitrary formula, and in practice parentheses should be used as needed to avoid ambiguity of the scopes of operators.

where $p \in$ ATM and ATM is the set of all atomic propositions (*atoms*), and $g \in$ AGT and AGT is the set of all *epistemic agents* or simply *agents*.

An *unpointed model* in epistemic S5 is a 3-tuple $\mathcal{M} = \langle S, R, V \rangle$, where

S is a non-empty set of states,
R is a mapping from agents to sets of pairs of states; i.e. $R : \text{AGT} \to \mathcal{P}(S^2)$, where
R_g is a set of pairs of states called the *epistemic accessibility relation for agent g*, and
V is a mapping from states to a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation function*; i.e. $V : S \to (ATM \to \{0,1\})$, where
V_s is a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation for state s*; i.e. $V_s : ATM \to \{0,1\}$, where
$V_s(p)$ is called the *truth assignment for atom p at state s*.

A *model* in epistemic S5 is a 2-tuple $\langle \mathcal{M}, s_0 \rangle$ where $\mathcal{M}$ is an unpointed model and $s_0 \in S$ is the *factual state*. (So to speak, an unpointed model is a model that lacks a factual or "point" state.)

The truth definitions for epistemic S5 are:

$\vDash_s^{\mathcal{M}} p$ iff $V_s(p) = 1$
$\vDash_s^{\mathcal{M}} \neg\phi$ iff $\nvDash_s^{\mathcal{M}} \phi$
$\vDash_s^{\mathcal{M}} \phi \wedge \psi$ iff $\vDash_s^{\mathcal{M}} \phi$ and $\vDash_s^{\mathcal{M}} \psi$
$\vDash_s^{\mathcal{M}} \phi \vee \psi$ iff $\vDash_s^{\mathcal{M}} \neg(\neg\phi \wedge \neg\psi)$
$\vDash_s^{\mathcal{M}} \phi \to \psi$ iff $\vDash_s^{\mathcal{M}} \neg\phi \vee \psi$
$\vDash_s^{\mathcal{M}} \phi \leftrightarrow \psi$ iff $\vDash_s^{\mathcal{M}} (\phi \to \psi) \wedge (\psi \to \phi)$
$\vDash_s^{\mathcal{M}} K_g\phi$ iff $\forall t \in S$ if $\langle s, t \rangle \in R_g$ then $\vDash_t^{\mathcal{M}} \phi$
$\vDash_s^{\mathcal{M}} \hat{K}_g\phi$ iff $\vDash_s^{\mathcal{M}} \neg K_g \neg\phi$

2.1.2 Shortcomings

In epistemic S5, what an agent considers is presented using the K operator. For example, that Alice considers it the case that p is presented as $K_{Alice}p$.

However, notice that there is no formalization of changes in agents' consideration bases. While it can express what an agent considers at any single slice of time, it is unable to express how agents' consideration bases should change in response to new information.

2.2 Public announcement logic (PAL)

After Hintikka, Plaza extended the logic to produce public announcement logic (PAL).[11] PAL is able to express changes in an individual agent's epistemic state, albeit only to a particular type of event: open disclosure of true information.[6]

2.2.1 Definitions

The language of public announcement logic (PAL) $\mathcal{L}_{\text{PAL}}$ is the set of formulas captured by the following BNF:

$$\phi ::= p|\neg\phi|\phi \wedge \phi|K_g\phi|[\phi]\phi$$

where $p \in$ ATM and ATM is the set of all atomic propositions (atoms), and $g \in$ AGT and AGT is the set of all epistemic agents.

An unpointed model in PAL is defined identically as in epistemic S5 as a 3-tuple $\mathcal{M} = \langle S, R, V \rangle$, where

S is a non-empty set of states,
R is a mapping from agents to sets of pairs of states; i.e. $R : \text{AGT} \to \mathcal{P}(S^2)$, where
R_g is a set of pairs of states called the *accessibility relation for agent g*,
and
V is a mapping from states to a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation function*; i.e. $V : S \to (ATM \to \{0,1\})$, where
V_s is a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation for state s*; i.e. $V_s : ATM \to \{0,1\}$, where
$V_s(p)$ is a member of the set $\{0,1\}$ called the *truth assignment for atom p at state s.*

A model in PAL is again a 2-tuple $\langle \mathcal{M}, s_0 \rangle$ where $\mathcal{M}$ is an unpointed model and $s_0 \in S$ is the factual state.

The truth definitions for PAL are:

$$
\begin{array}{lll}
\vDash_s^{\mathcal{M}} p & \text{iff} & V_s(p) = 1 \\
\vDash_s^{\mathcal{M}} \neg\phi & \text{iff} & \nvDash_s^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \phi \wedge \psi & \text{iff} & \vDash_s^{\mathcal{M}} \phi \text{ and } \vDash_s^{\mathcal{M}} \psi \\
\vDash_s^{\mathcal{M}} \phi \vee \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg(\neg\phi \wedge \neg\psi) \\
\vDash_s^{\mathcal{M}} \phi \rightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg\phi \vee \psi \\
\vDash_s^{\mathcal{M}} \phi \leftrightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi) \\
\vDash_s^{\mathcal{M}} K_g\phi & \text{iff} & \forall t \in S \text{ if } \langle s, t\rangle \in R_g \text{ then } \vDash_t^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \hat{K}_g\phi & \text{iff} & \vDash_s^{\mathcal{M}} \neg K_g \neg\phi \\
\vDash_s^{\mathcal{M}} [\psi]\phi & \text{iff} & \vDash_s^{\langle \mathcal{M}, s\rangle \otimes \psi} \phi
\end{array}
$$

where $\otimes$ is the *public announcement update operator* such that an *updated model* $\langle \mathcal{M}, s\rangle \otimes \psi$ is a 2-tuple $\langle \mathcal{M}', s\rangle$, where $\mathcal{M}'$ is a 3-tuple $\langle S', R', V'\rangle$, where

$S' = \{s \in S | \vDash_s^{\mathcal{M}} \psi\}$,
$R' = \{\langle g, R'_g\rangle | R'_g = \{\langle s, t\rangle \in R_g | s, t \in S'\}\}$,
and
$V' = \{\langle s, V_s\rangle | s \in S'\}$.

Notes: If updating $\langle \mathcal{M}, s\rangle$ with ψ would cause it to be the case that $s \notin S'$, then it's illegal to update $\langle \mathcal{M}, s\rangle$ with ψ.
An updated model is a special case of model.
An updated model $\langle \mathcal{M}, s\rangle \otimes \psi$ still has factual state s, albeit $s \in S'$.

Using PAL, we're able to express not only the static epistemic states of each agent between every update but also the effect of each communication on their epistemic states.

2.2.2 Shortcomings

I emphasize that PAL is limited in the types of events it can express. PAL encodes epistemic update only for events of common information disclosure. However, PAL is unable to express events where some strict subset of agents is the audience for information disclosure.

This limitation is better demonstrated through an example. I do so in the following sections.

2.3 The Cards epistemic model

Consider the following scenario, which we'll call "Cards"[5]:

Alice, Bill, and Carol are playing a guessing game using playing cards. The game uses only the three face cards, namely Jack, Queen, and King. The three cards are shuffled and placed face-down. Then, each person takes one and only one card without showing the other two which card was taken. The objective of the game is for each person to deduce which card the other two players have.

Our interest in Cards is with respect the effect that certain actions of the players have on what each player deduces about which card is held by whom. For example, suppose Bill is dealt the Queen. Initially he does not know whether Carol has the Jack or the King. (But he does know she doesn't have the Queen, because he has the only Queen.) But if Alice shows everyone that she has the Jack, then Bill immediately is able to deduce that Carol holds the King.

In the following sections we'll analyze how certain actions can or cannot be expressed as events in an epistemic model using PAL.

2.3.1 PAL can express public display

Suppose that Alice receives the Jack, Bill receives the Queen, and Carol receives the King. Symbolically we'll refer to this situation as JQK, where each letter refers to a particular card and the position of each letter refers to the person who has that card. So JQK means that Alice has the Jack, Bill has the Queen, and Carol has the King; for example if the situation was QJK then that would mean that Alice has the Queen, Bill has the Jack, and Carol has the King.

Now suppose that Alice shows everybody that she has a Jack. The outcome is that both Bill and Carol learn which card each person has while Alice remains uncertain which of two possible deals is actually the case. PAL is able to express this situation[6]. The initial situation is expressed by the following model:

[5] This scenario is found in chapter 5 of van Ditmarsch *et al.*'s monograph on dynamic epistemic logic.[6] I've merely replaced the numbering 0,1,2 of the cards with the more familiar names of Jack, Queen, and King (J,Q,K). Van Ditmarsch *et al.*'s pedagogical presentation of Cards covers only for the public announcement event. The end-of-chapter exercises guide the reader to consider cases of private disclosure, but no answers are given within the text. In this investigation we use these exercises as a starting point. Our first contribution will be to identify content ignorance from event ignorance. It is shown that this distinction leads to a design methodology for a logic of misinformation extended from DEL.

[6] Logically, the event of Alice showing everybody her Jack is just the same as if she had (truthfully) announced to everybody that she has a Jack.

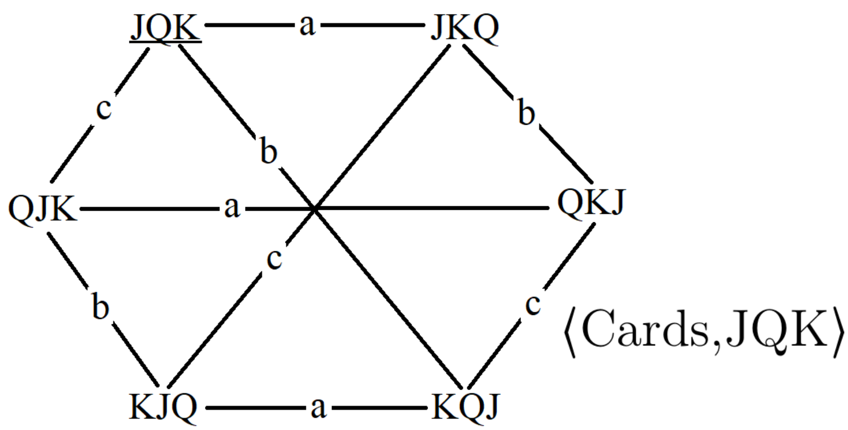


Figure 1: The ⟨Cards,JQK⟩ epistemic model

The factual state, JQK, is underlined. States are named according to the deal of the three cards, where the letters correspond, in order, with Alice's card, Bill's card, and Carol's card respectively. Accessibilities are presented as lines between states and are labelled according to the relevant agent. For example, in this model it is the case that $K_a(Q_b \vee K_b)$, which says that Alice knows that Bill either has the Queen or the King, where Q_b for example means that Bill (agent "b" in the formula) has the Queen ("Q" in the formula).

As Alice shows everybody that she has a Jack, the situation undergoes model update from ⟨Cards,JQK⟩ to ⟨Cards′,JQK⟩ = ⟨Cards,JQK⟩ $\otimes J_a$. The public announcement event is that Alice has the Jack, expressed as J_a. Now after Alice shows everybody that she has a Jack, the situation is expressed by the following updated model:

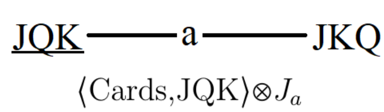


Figure 2: Cards after update with public announcement J_a

2.3.2 PAL cannot express private disclosure

But suppose that instead of showing everybody, Alice shows only Bill which card she has. PAL is unable to express this event because events in PAL are assumed to be observed by all agents involved. Formally, this assumption also means that updates in PAL affect all agents in the same way.

Public announcement events exclude those events which appear different to different agents. That agents may receive different information from a particular event is beyond the expressive capacity of PAL.

Specifically, PAL is only able to work with events where *all* agents are *truthfully* told that something is the case. The assumptions that events affect all agents equally and that the information content of all events are true limit PAL from being able to capture events that are more complex than public announcements.

In comparison, DEL drops the first assumption that events affect all agents equally. In the following section we will see how forgoing this assumption allowed DEL to express private disclosure in Cards.

2.4 Dynamic epistemic logic (DEL)

Dynamic epistemic logic (DEL) extends PAL by dropping the assumption that events are public. As a result DEL is able to express private events such as Alice showing Bill and only Bill the card that she received.

We should also be cognizant that it makes a difference to the epistemic outcome whether Carol is aware of *the occurrence of the event of* Alice showing Bill her card, independent of her not being aware of which card Alice showed to Bill. That is, it makes an epistemic difference whether the occurrence of Alice's event, in addition to the information conveyed in Alice's event, is private. Which card Alice has is kept secret from Carol, but that Alice showed Bill which card she has itself may also be kept secret from Carol.

For now we'll work only with the case where Carol is aware that Alice showed Bill her card, and it is only that Carol is unaware of exactly which card Alice showed to Bill. In the next chapter it will be shown that DEL is able to express Alice's private disclosure even if its occurrence in addition to its information is private to Carol.

2.4.1 Definitions

The language of dynamic epistemic logic (DEL) $\mathcal{L}_{\text{DEL}}$ is the set of formulas captured by the following BNF:

$$\phi ::= p | \neg\phi | \phi \wedge \phi | K_g \phi | [\langle \mathcal{E}, e \rangle] \phi$$

where $p \in \text{ATM}$ and ATM is the set of all atomic propositions (atoms), $g \in \text{AGT}$ and AGT is the set of all epistemic agents, and $\langle \mathcal{E}, e \rangle$ is an event model (see below).

An *unpointed epistemic model* in DEL $\mathcal{M}$ is defined as a 3-tuple $\langle S, R, V\rangle$, where

S is a non-empty set of states,
R is a mapping from agents to sets of pairs of states; i.e. $R : \mathrm{AGT} \to \mathcal{P}(S^2)$, where
R_g is a set of pairs of states called the *epistemic accessibility relation for agent* g,
and
V is a mapping from states to a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation function*; i.e. $V : S \to (ATM \to \{0,1\})$, where
V_s is a mapping from atoms to a member of the set $\{0,1\}$ called the *valuation for state* s; i.e. $V_s : ATM \to \{0,1\}$, where
$V_s(p)$ is a member of the set $\{0,1\}$ called the *truth assignment for atom* p *at state* s.

An *epistemic model* in DEL is a 2-tuple $\langle \mathcal{M}, s\rangle$ where $s \in S$ is the actual state.

We also first need to define what an *event model* in DEL is before we can move on to its truth definitions.

An *unpointed event model* in DEL $\mathcal{E}$ is defined as a 3-tuple $\langle E, Q, P\rangle$, where

E is a non-empty set of *events*,
Q is a mapping from agents to sets of pairs of events; i.e. $Q : \mathrm{AGT} \to \mathcal{P}(E^2)$, where
Q_g is a set of pairs of events called the *event accessibility relation for agent* g,
P is a mapping from events to formulas called the *precondition function*;
i.e. $P : E \to \mathcal{L}_{\mathrm{DEL}}$, where
$P(e)$ is a formula called the *precondition for event* e.

An *event model* in DEL is a 2-tuple $\langle \mathcal{E}, e_0\rangle$, where $e_0 \in E$ is the factual event.

Now the truth definitions for DEL follow:

$$
\begin{array}{lll}
\vDash_s^{\mathcal{M}} p & \text{iff} & V_s(p) = 1 \\
\vDash_s^{\mathcal{M}} \neg\phi & \text{iff} & \nvDash_s^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \phi \wedge \psi & \text{iff} & \vDash_s^{\mathcal{M}} \phi \text{ and } \vDash_s^{\mathcal{M}} \psi \\
\vDash_s^{\mathcal{M}} \phi \vee \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg(\neg\phi \wedge \neg\psi) \\
\vDash_s^{\mathcal{M}} \phi \rightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg\phi \vee \psi \\
\vDash_s^{\mathcal{M}} \phi \leftrightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi) \\
\vDash_s^{\mathcal{M}} K_g\phi & \text{iff} & \forall t \in S \text{ if } \langle s,t\rangle \in R_g \text{ then } \vDash_t^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \hat{K}_g\phi & \text{iff} & \vDash_s^{\mathcal{M}} \neg K_g \neg\phi \\
\vDash_s^{\mathcal{M}} [\langle\mathcal{E},e\rangle]\phi & \text{iff} & \vDash_s^{\langle\mathcal{M},s\rangle\otimes\langle\mathcal{E},e\rangle} \phi
\end{array}
$$

where $\otimes$ is the *event update operator* such that an *updated epistemic model* $\langle\mathcal{M},s\rangle \otimes \langle\mathcal{E},e\rangle$ is a 2-tuple $\langle\mathcal{M}', s[e]\rangle$, where $\mathcal{M}'$ is a 3-tuple $\langle S', R', V'\rangle$, where

$S' = \{s[e] | s \in S \text{ and } e \in E \text{ and } \vDash_s^{\mathcal{M}} P(e)\}$,
$R' = \{\langle g, R'_g\rangle | R'_g = \{\langle s[e], t[f]\rangle | s[e], t[f] \in S' \text{ and } \langle s,t\rangle \in R_g \text{ and } \langle e,f\rangle \in Q_g\}\}$,
and
$V' = \{\langle s[e], V_s\rangle | s[e] \in S'\}$.

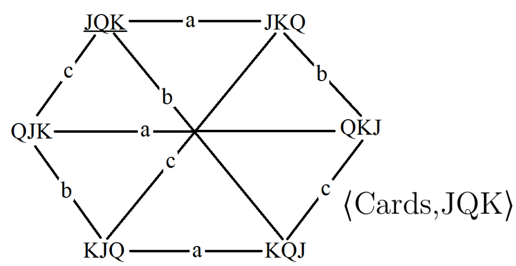
Notes: $s[e]$ refers to the state s of an epistemic model after update with an event model, where event e exists in the event model and the precondition for e is satisfied in s.
If updating $\langle\mathcal{M},s\rangle$ with $\langle\mathcal{E},e\rangle$ would cause it to be the case that $s[e] \notin S'$, then it's illegal to update $\langle\mathcal{M},s\rangle$ with $\langle\mathcal{E},e\rangle$.
An updated epistemic model is a special case of epistemic models.
An updated epistemic model $\langle\mathcal{M},s\rangle \otimes \langle\mathcal{E},e\rangle$ has factual state $s[e] \in S'$.

A further note about state notation in updated epistemic models: Square brackets in state names help further distinguish different states in updated epistemic models. This is useful notation because states in updated epistemic models can be distinguished by the occurrence of different events alone. For example, the state $s[e]$ represents the state s after the event e has occurred. But more than one event may occur at a single state. For example, we may suppose that it is possible for either event e or event f to occur at s, in which case the updated event model would have separate states $s[e]$ and $s[f]$.

Now using DEL, we can express Alice's private disclosure. For now let's assume that Carol can see that Alice showed Bill her card, and it's just that Carol can't see which card was shown to Bill.

The outcome of this is that Bill knows which deal is actually the case, Carol knows that Bill knows which deal is actually the case, and both Alice and Carol each remain uncertain which of two possible deals is actually the case. Both Alice and Bill also know that Carol doesn't know exactly which deal is the case.

Let's call the event model for this scenario "Show". Then the initial situation is the same as before:



As Alice shows Bill that she has a Jack while Carol can see that Alice is showing Bill the card, we observe model update from ⟨Cards,JQK⟩ to ⟨Cards′,JQK[aJ$_{\text{a}}$b]⟩ = ⟨Cards,JQK⟩ $\otimes$ ⟨Show,aJ$_{\text{a}}$b⟩. The event model ⟨Show,aJ$_{\text{a}}$b⟩ is:

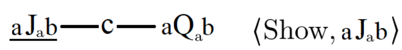


Figure 3: The ⟨Show,aJ$_{\text{a}}$b⟩ event model

The name of the state describes the event that occurred; for example, "aJ$_{\text{a}}$b" means "Alice shows Bill that her card is the Jack". Note that the set E in an event model is merely a set of events irrespective of whether every one of those events is possible at the actual state in the epistemic model. In the case of the event model Show, it's important that aQ$_{\text{a}}$b is a member of E because Carol cannot identify which card it is that Alice is showing Bill.

The updated epistemic model immediately after Alice shows everybody that she has a Jack is:

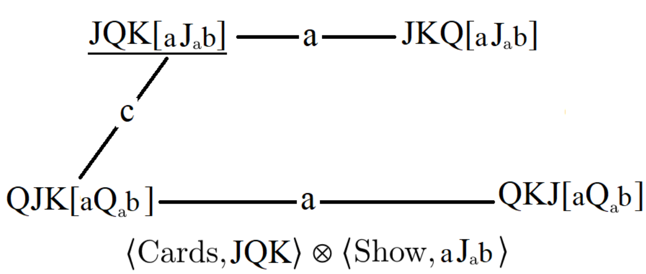


Figure 4: Cards after update with event ⟨Show,aJ$_{\text{a}}$b⟩

Where PAL was unable to express the epistemic outcome of this event of private disclosure, DEL succeeds handily. DEL accomplished this by dropping PAL's assumption that events are public. But PAL also assumes that events are truthful, that is to say, that the informational content of any event is a proposition that actually the case.

In the next chapter I demonstrate that DEL's continued reliance on this second assumption hinders it from expressing misinformation events. We will return to the Cards example to diagnose the sort of modifications that will be needed in order to drop this assumption in DEL while maintaining logical consistency. The distinction between content and event ignorance provides us clues on doing so.

3 Deception and DEL

In this chapter we'll dive into the details of the Cards example to illustrate places where DEL could be improved regarding expressing circumstances of misinformation. We'll also use Cards to dissect the notion of private events with regard to the different epistemic effects of content ignorance and event ignorance. DEL's handling of event ignorance will yield clues on the sort of extension needed in order to drop the assumption of events' informational contents' truth.

I'll begin by addressing an existing proposal on extending DEL to capture deception. My criticism argues that expression of deception in DEL is more meaningful if encoded with greater generality. My proposed solution will instead follow the guideline that a logic of misinformation should be developed before a logic of deception. With this background I'll move on and use the Cards model to analyze specific points of failure of DEL to capture misinformation. This analysis highlights the places where DEL should be modified to develop an extension of DEL capable of expressing misinformation.

3.1 Existing literature and relevance to the current project

To date, formalizing deception in epistemic logic has only been accomplished for specific event categories such as lying and bluffing.[5, 12] I argue that formalizing deception by explicitly encoding specific event categories is not only ad hoc but also unhelpful for formalizing deception as a whole. This is because the concept of deception is much broader than specific event types. Rather than develop a logic where various examples of deception are extended piecemeal from an existing logic, I argue that it would be more promising to analyze the concept of deception, as I have done in Chapter 1, to indicate the conceptual shifts required to extend and adapt an existing logic to the task of deception.

As such, in contrast to past work my objective is to develop a logic which does not make explicit different categories of events. Further, I adopt the methodical approach of formalizing conceptual aspects of deception; I begin with the more basic notion of misinformation, from which the categorization of an event as a lie or a bluff would be emergent from the structure of the resulting updated epistemic model. This emergence of properties is much in the same way that modal epistemic logic does not explicitly stipulate agents' consideration bases but rather takes them as emergent from the structure of agent-typed epistemic accessibilities between epistemic states.

Besides serving as a non-ad-hoc foundation for a logic of deception, I predict that such a logic will lead us to more philosophically interesting insights about deception. Such hypotheses are expected from the greater freedom of interactions between events from reduced restrictions on event classification. Moreover, unlike past attempts to formalize deception, *misinformation* will not need explicit encoding as a special case in my envisioned logic. This approach allows for a logic of misinformation where instances of misinformation are emergent rather than forced into the system as a sort of logical reification.

DEL also lacks support for trust between agents. While past work on a logic of lying incorporates Liu's insights regarding discrete agent differentiation with respect to actor discrimination and trust, I want to avoid explicit categorization or typing of agents.[5, 10] My approach will be instead to only explicitly stipulate relationships of trust between agents and have agent identification as emergent from the resultant epistemic models.

For example, instead of typing Alice as a liar, my approach would instead produce a resultant epistemic model in which the trust function for every other agent outputs a "do not trust advisory" when the input is Alice. This approach makes possible the emergence of complex relationships of trust where some receivers think of Alice as a liar but the rest do not. Non-ad-hoc implementation of such a relationship would be impossible on simple explicit agent typing.

Again, I believe that emergent rather than explicit categorization is more likely to produce philosophically interesting hypotheses as logical outcomes of a logic of deception. Nevertheless, to my knowledge there is no discussion of formalizing nondiscrete trust specifically in DEL. I acknowledge past work on probabilistic extensions of DEL as loose inspiration for approaching the task of encoding nondiscrete trust in DEL.[3, 8]

3.2 Shortcomings of deception with DEL

In this section I will revisit the aspects of deception: intentionality, misinformation, and trust. I'll analyze each aspect alongside DEL to determine how and where DEL fails to capture it, which in turn directly inform our development of extensions of DEL in the following chapters.

3.2.1 DEL cannot express intentionality

DEL lacks mechanisms for expressing agents' intentionality. We can see this because the outcome of event update should be affected by the intention of the actor. If DEL had a mechanism for expressing intentionality, then event update in DEL would be expected to accept as an input the intention of the actor as an input to the updated epistemic model.

3.2.2 DEL cannot express misinformation

Earlier it was seen that DEL is capable of expressing events that leave other agents in (various states of) ignorance, or in other words events that refrain from providing information. On the other hand, DEL cannot express events that provide false information, and as a result is incapable of capturing events of dishonest disclosure.

Let's consider again Cards and recall the criticism that PAL is only able to express events where *all* agents are truthfully told that something is the case. DEL overcomes that restriction, but it is still only able to express events where all or some agents are *truthfully* told that something is the case. For example, while DEL is able to express the epistemic outcomes of Alice showing only Bill her card, DEL is unable to express the event where Alice tells (lies to) Bill that she has the King.

The reason for this is the requirement that epistemic update follows the precondition function of events in the event model. A precondition function implies that there must be some prerequisites fulfilled in order for the event to take place.[6] Essentially, that DEL stipulates the precondition of an event to be fulfilled before it is a legal event forces it to also assume out of hand that the informational content of the behavior encoded as an event is true. Therefore, DEL is incapable of expressing misinformation because misinformation events do not exist in DEL.

This suggests that as a first step we need to eliminate or at least loosen the stringency of the precondition function. Insofar as we're interested in formalizing misinformation, it makes sense that there should be less restrictions on events that can occur, because in reality we are not restricted to acting in ways which are consistent with facts of reality.

But in order to permit this we must also have mechanisms to deal with issues such as agents' coming to seriously consider contradictions when misinformation events occur. This is because event update in DEL simply takes the informational content of events at face value. In DEL, agents do not analyze the information they're receiving before update. Relaxing the restriction on the precondition function means observing an event purporting p will lead agents already considering $\neg p$ into serious consideration of a contradiction.[7]

As such the first step for developing a logic of misinformation extended from DEL is to add a mechanism for misinformation handling in DEL. Specifically there must be some way for agents to recover from the consideration of a contradiction. We can say that while agents may "tentatively consider" contradictions that arise as a result of the receipt of misinformation, the receipt of that faulty information does not also ultimately lead them

[7] This suggests that encoding misinformation is deeply intertwined with encoding trust. We will eventually see that there's a fundamental relationship between misinformation and trust in the context of modal epistemic logics.

to "seriously consider" a contradiction. I believe that this treatment also aligns well with intuitions about the way that people reason when they encounter new information which may or may not be true[8], and which they realize may or may not be true.

3.2.3 DEL cannot express trust

Secondly, DEL appears to lack mechanisms for encoding trust. This is also in part due to the assumption carried forward partly from PAL that the propositional content of events cannot be false.

PAL requires that event informational content is true while DEL only weakly requires that it not be false. This was the reason why DEL but not PAL could express private disclosure. This restriction on the type of events admissible in the system conceals its lack of trust encoding because the assumption that the information content of events is not false allows agents to unquestioningly take up the informational content of received events without finding themselves afterwards with an internally inconsistent consideration base.

Recall from our discussion in Chapter 1 that in order for an agent to exist in a state of misinformation, agents must exhibit some propensity to "jump to conclusions". Agents in DEL cannot be deceived because in DEL the event truth precondition enforces the truth of the information content of events and agents are perfectly deductive. In a logic of misinformation, agents must be able to trust each other to some degree, or in other words that they will seriously consider a proposed proposition p even without sufficient deductive evidence to believe it.

3.3 Ignorance in DEL

Analyzing the effect of private disclosure in Cards demonstrates that there are two types of ignorance that can result from private disclosure. We'll distinguish these two by calling them *content ignorance* and *event ignorance*. The flipside of these epistemic outcomes are the types of private events which effect them; we'll respectively distinguish between *open* private events and *closed* ("secret") private events.

3.3.1 Content ignorance

Recall that content ignorance was presented earlier to showcase the increase in expressive power of DEL over PAL (Chapter 2).

[8] Or perhaps we should say, "which may or may not be consistent with their pre-existing consideration base".

The transition for the Cards model for content ignorance looked like this[9]:

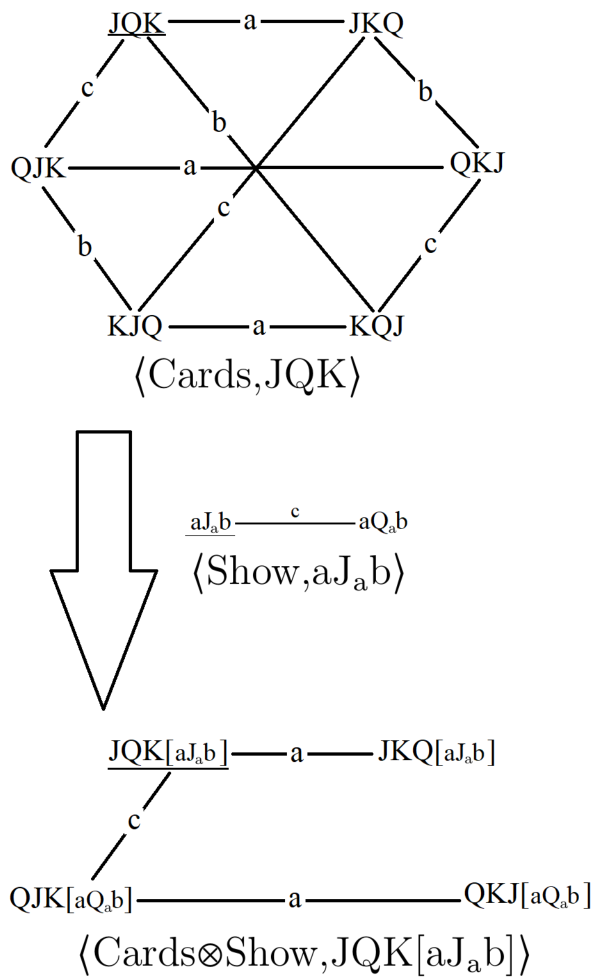


Figure 5: Update with the ⟨Show,aJ_ab⟩ event model

Notice that aJ_ab produces content ignorance because Carol is aware of the occurrence of an event but not of which card is being shown to Bill. We will call this phenomenon "content ignorance" because Carol is left unaware only of the content of the event that was communicated, not of its having occurred in the first place.

Further, if Carol had been unaware of the event's occurrence as well, then the set of states to which she has access would not change as a result of update. We can say that Carol's "worldview", the set of states which she entertains as candidates for the factual state, would

[9]⟨Cards,JQK⟩ (seen at the top) is the epistemic model before update with the event model ⟨Show,aJ_ab⟩ (seen in the middle next to the large arrow). The large arrow indicates the event update operation on ⟨Cards,JQK⟩ with ⟨Show,aJ_ab⟩ to produce the updated epistemic model ⟨Cards⊗Show,JQK[aJ_ab]⟩ (seen at the bottom).

be unchanged if she were unaware of the event's having occurred at all. Nevertheless, we see here that under content ignorance Carol's worldview does change.[10]

Notice also that despite her content ignorance Carol still carries the assumption that whatever it was that Alice whispered to Bill, its informational content is true. This is a result of DEL's precondition function, as we discussed earlier. In the Cards scenario this means that Carol, who knows that she holds the only King, also therefore knows that what Alice whispered to Bill was (exclusively) either "I have the Jack" or "I have the Queen", but certainly not "I have the King."

3.3.2 Event ignorance

Event ignorance leads to an altogether different updated epistemic model than content ignorance.

Consider the following alternative transition for the Cards model (see next page):

[10] I'll rigorously define "worldview" later.

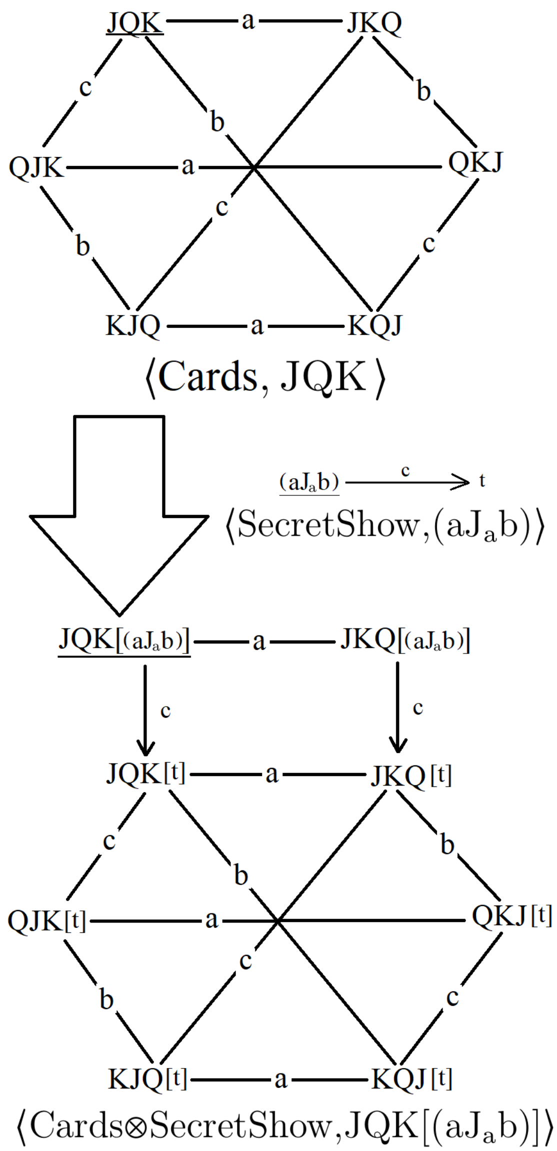


Figure 6: Update with the ⟨SecretShow,($aJ_a b$)⟩ event model

($aJ_a b$) produces event ignorance because Carol is not aware of which card is being shown to Bill, as well as not aware of a card being shown to Bill. The state used in the event model for situations of event ignorance is t, the "null event". In DEL, the precondition for the null event is trivial truth $\top$, or in other words there are no states eliminated by update with t.

Note that whenever there is a one-way accessibility relation, that implies that there does not exist any reflexive access for that agent in the first state identified in the one-way accessibility pair. In DEL this is built in to epistemic update, specifically because elements of the set of an agent's accessibility relation are preserved through update only when the corresponding precondition-respecting events also stand in relation of accessibility with respect to that agent.

We call this phenomenon “event ignorance” because Carol is unaware of the occurrence of the event of Alice whispering to Bill.

Notice that Carol’s worldview is unchanged even after update. A fortiori, Carol is not aware of Bill’s worldview having changed. Accordingly, when Carol is subjected to event ignorance it must mean that Carol loses accessibility to any state in the updated epistemic model where an event other than the null event has occurred.

Put another way, one-way accessibility implies that the agent is being led away from the starting state without their being aware of it. Carol is unaware of the event taking place, so after event update Carol must have a one-way accessibility away from the actual state (where a non-null event has taken place) to a factually identical state where the null event has “taken place”. Moreover Carol cannot have reflexive access to the actual state (or any other state where a non-null event has taken place) because that would mean Carol is aware of (at least) the possibility of a non-null event having taken place.

4 Misinformation handling in DELMI$_{\text{int}}$

So what happens if we forgo event preconditions? Then it's possible for an agent to consider a proposition and at the same time also consider its negation. But this means it's possible for an agent to seriously consider contradictions to be the case.

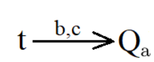
While it certainly is unacceptable for agents to seriously consider contradictions, I argue that it isn't unacceptable for agents to merely tentatively consider contradictions. This is because tentative consideration of contradictions mirrors an intuitive understanding of our thought process when we learn new things about the world inconsistent with our existing knowledge. When we come to observe that p (or so it seems) after having hitherto understood that $\neg p$, we must come away from the experience with one or the other exclusively intact. How do we understand that rejection is an appropriate response? Precisely because we encounter the threat of contradiction, and such an encounter can be provided for in a logic of misinformation by way of "tentative" consideration.

Thus a logic of misinformation must minimally have some mechanism for resolving contradiction arising as a result of update. In a basic logic of misinformation, it should be sufficient for the ensuing epistemic revision to be identical for all agents. In a more complex logic of misinformation different agents might be expected to respond to the threat of contradiction in different ways, and perhaps even under different contexts such as that based on identifying the actor of the most recent event. For our current purposes let's adhere to our objective of developing first only a basic logic of misinformation.

4.1 Misinformation in Cards with DEL

Let's see what happens in the Cards scenario if we encode misinformation by simply dropping the precondition requirement in DEL. For example, suppose that Alice tells everybody else that she has the Queen.

Then the event model looks like this[11]:

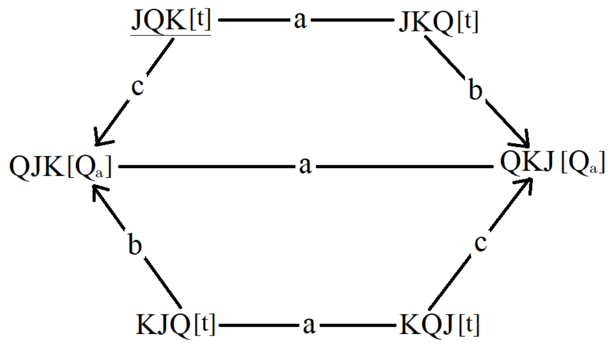


$\langle$LieQueenAll, "Q_a"$\rangle$

Figure 7: The $\langle$LieQueenAll, "Q_a"$\rangle$ event model

[11] "LieQueenAll" is shorthand for "[Alice] lies that she has the Queen, to all agents". The other event models in this investigation are similarly named.

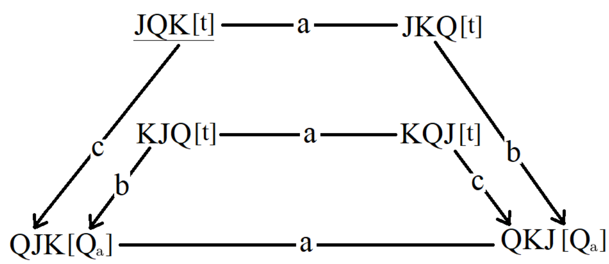
Setting aside for now the issue of what agents b and c do when they arrive at contradiction, here's what the updated model looks like:



⟨Cards⊗LieQueenAll,JQK[“Q_a”]⟩

Figure 8: After update with the ⟨LieQueenAll,“Q_a”⟩ event model, using DEL

There is a remarkable similarity of this model to that resulting from event ignorance in Cards. Rearranging the presentation a bit makes this even more evident:



⟨Cards⊗LieQueenAll,JQK[“Q_a”]⟩

Figure 9: Alternative view of Cards after update with the ⟨LieQueenAll,“Q_a”⟩ event model, using DEL

Those agents which do not become aware of their having been misinformed are directed via one-way accessibility to the subset of the states of the pre-update epistemic model, as if the original Cards model had been updated with the event at some factual state where Alice wasn't lying. Compare this with update under event ignorance, where those agents which are not aware of the occurrence of the event are directed via one-way accessibility to the full set of states of the pre-update epistemic model as if no event had occurred.

Can we capture misinformation in an extension of DEL by repurposing the strategy of how DEL expresses event ignorance? We certainly can.

While we see again here that DEL already has the means to indicate the occurrence of misinformation, but only lacks a mechanism with which to express how agents should react to the discovery of having received misinformation. To recapitulate, DEL doesn't have the means to express misinformation handling. At the states where agents are aware of their having received misinformation, agents seriously consider contradictions as a direct result of the misinformation update. DEL is able to express that an agent realizes that they have been the recipient of misinformation by having that agent arrive at a contradiction after event update.

However, DEL lacks a mechanism to eliminate such contradiction. Agents are aware by the inconsistency of their own consideration b‘ase that they have been the recipient of some misinformation (whether during the most recent event or at some point in the past), but there is no implementation of their reaction to this realization. Without a suitable reaction to the discovery of the receipt of misinformation, agents' tentative and serious considerations collapse into one.

Looking back to Figure 9, notice that all those receivers that become aware of the misinformation arrive at the self-contradictory consideration base by way of losing all accessibilities to any other state, including itself. For example, at the factual state JQK[t], Alice is aware that if it is actually the case that JKQ[t], then Carol would immediately be aware of Alice having lied. The intuition is that Carol already holds the Queen, so she knows that Alice cannot actually hold the Queen as stated. Formally, Carol would lose all accessibility out of the state JKQ[t], meaning that Carol's consideration base would contain contradictions (including the relevant contradiction that it is the case that Alice does not hold the Queen *and* Alice does hold the Queen).[12]

For example, at state JQK[t] Bill is aware of the fact that Alice has lied about having the Queen, because Bill knows that he has the Queen. Moreover, he knows that Carol is now under the mistaken consideration that Alice really does have the Queen. But, since Bill now has no accessibility from the factual state to any other state (including the factual state itself), Bill seriously considers any and all contradictions, including the relevant contradiction that Alice does not have the Queen (inferred from Bill's having the Queen) and Alice does have the Queen (inferred from Alice's claim to having the Queen). This is because for any agent g at states where there is no accessibility to any other state it is trivially true that $K_g\phi$ for any $\phi \in \mathcal{L}_{\text{DEL}}$.

While explosion of agents' consideration bases betrays the transfer of misinformation,

[12]Later we'll see that actually this is always the case. All agents that become aware of their having received misinformation do so by arriving at all possible contradictory considerations including the relevant contradiction as a result of trivial truth of the consideration operator from a lack of accessibility. This always happens in the foregoing logic because in order to become aware of misinformation in the first place there must have already been no accessibility to states contradictory to their consideration base. Later we will describe this by saying that fibbees always lose all accessibilities during event update prior to the reconsideration step because their initial worldviews necessarily precluded any states consistent with the purported informational content of the misinformation event.

still it must not be the case that agents ultimately come to seriously consider contradictions. We ensure this by implementing a misinformation handling mechanism as an extension to DEL. To this we now turn our attention.

4.2 Intuitions

There is one little complication to repurposing event ignorance for misinformation. Consider that in event ignorance under Cards, Alice's worldview is always a strict superset of Carol's worldview. This represents the fact that Alice is aware of Carol's being unaware of the occurrence of the event at the same time that Carol is unaware of Alice's awareness of the factual state (Alice showed Bill her card, and Carol didn't notice that Alice's card has been shown to Bill).

In contrast, under misinformation, we would expect that Alice's worldview is a superset of Carol's only in the case when Carol has been successfully misinformed. If Carol were privy to some incontrovertible evidence that Alice is lying, then it would actually be Carol's worldview that is a superset of Alice's! This would represent the fact that Carol is aware of Alice's utterance being false at the same time that Alice is unaware of Carol's awareness of the factual state (Alice is lying, and Carol realizes that).

So in misinformation there is not necessarily, so to speak, an epistemic advantage of Alice over Carol, since Carol may know she has been lied to without Alice knowing that that's the case. In such a case it would actually be Carol who has the epistemic advantage over Alice.

I will illustrate this difference between event ignorance and misinformation. Let us try to update Cards in DEL with this event, which is a mixture of misinformation and event ignorance: Alice lies to Bill that she has the Queen, and this is a closed private event.

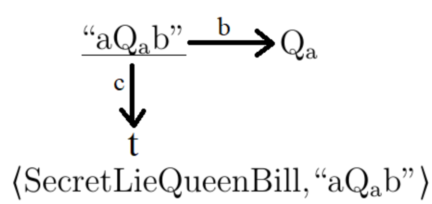
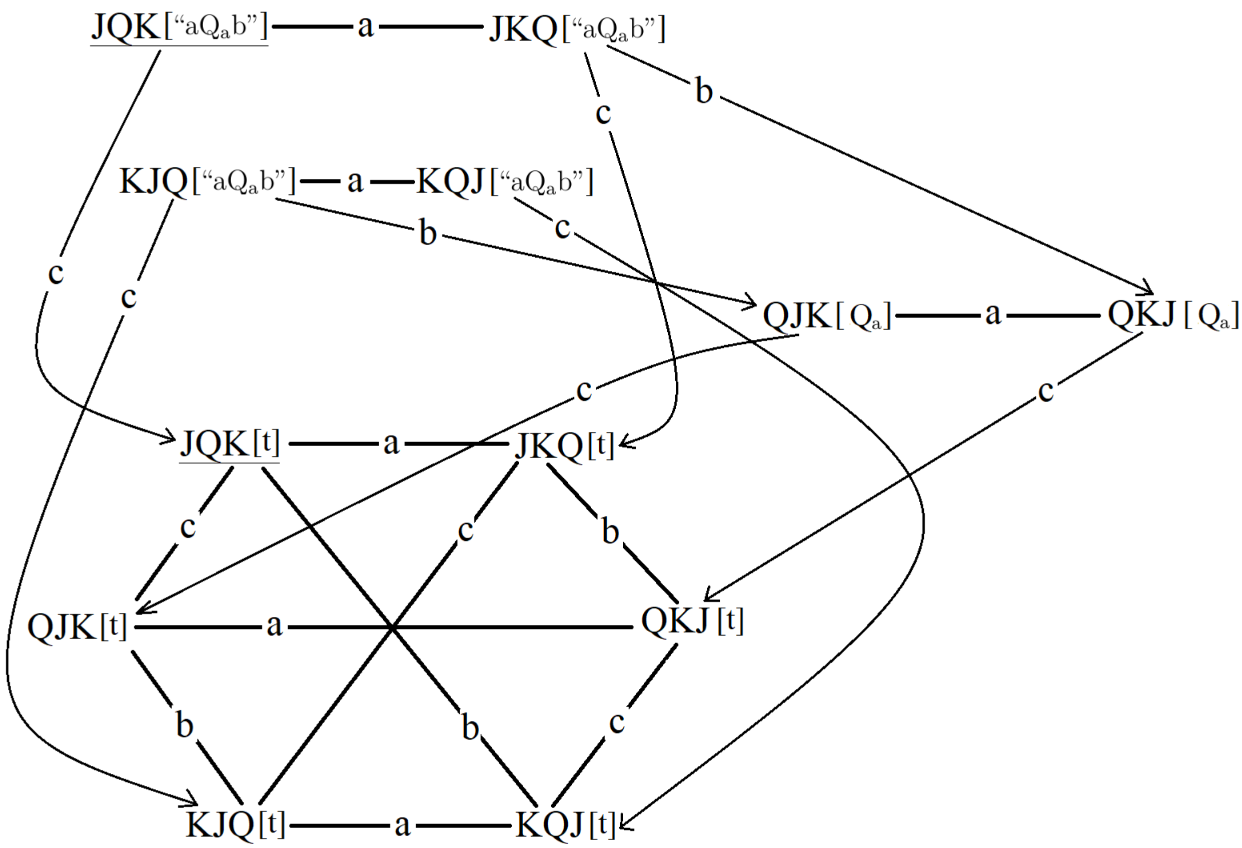


Figure 10: The ⟨SecretLieQueenBill, "$\mathrm{aQ_ab}$"⟩ event model

From the resultant updated epistemic model we can see that update for misinformation is a lot like update for event ignorance except with the additional accessibility for the case of realizing that one has been lied to:



⟨Cards⊗SecretLieQueenBill,JQK[“aQ_ab”]⟩

Figure 11: After update with the ⟨SecretLieQueenBill,“aQ_ab”⟩ event model, using DEL

There are two states (namely KJQ[“aQ_ab”] and KQJ[“aQ_ab”]) to which no agents have accessibility from the factual state. Again, remedying this sort of outcome is a purpose of misinformation handling.

The inadequacy of DEL's presentation of this event is clear in that in the states where Bill knows that Alice lied, namely JQK[“aQ_ab”] and KQJ[[“aQ_ab”], Bill has no accessibility to any state. So Bill considers all possible propositions, including all contradictions, as a result of the realization that he's received misinformation. As discussed, in the existing mechanisms for DEL Bill is able to identify misinformation but unable to properly react to it. What is needed is for Bill to at most only tentatively consider the self-contradictory consideration base before rejecting it.

Since misinformation handling boils down to what agents should do when they realize they're the recipient of misinformation, let's refer to particular types of reactions to such realization as “fibbee protocols”. For now we'll analyze only two fibbee protocols: “intolerance” and “negation”.

Intolerance will be to react to misinformation by ignoring it. The epistemic outcome of this is to keep intact one's consideration base. Negation will be to react to misinformation by incorporating the negation of the incoming information to one's consideration base. These fibbee protocols have intuitive appeal and are simple to implement in a modular design for a logic of misinformation.

Let's begin with the fibbee protocol of intolerance. We will need a new accessibility which reverts the effect of event update but only for the enlightened fibbee. This follows from our observation that epistemic models resulting from update with a misinformation event are similar to those resulting from update with an event ignorance event except that it's possible for recipients to result in having epistemic advantage over the actor.

Taking account of our development so far: We see that to capture misinformation in DEL we treat it as a special case of event ignorance where it may be the case that the audience becomes aware of having been lied to. This is specifically when update with the apparent information content of the event leads to contradiction in one's consideration base, under which there is evident need for the enactment of a fibbee protocol. With a little bit of foresight it should be reasonably straightforward to extend DEL such that its extension serves as a basic logic of misinformation with modularity for testing different fibbee protocols.

4.2.1 Veridicality

Veridicality, also known as the T axiom schema, is $K_g\phi \rightarrow \phi$ for arbitrary $g \in AGT$ and $\phi \in \mathcal{L}$.[2, 4] The inclusion of the T axiom schema in an epistemic logic means that the content of any serious consideration of agents is guaranteed to be true.

Obviously, veridicality does not hold in a logic of misinformation. This is because in a state of being misinformed it is the case that one considers something that is actually false. As such veridicality does not hold if it is possible to enter into a state of being misinformed.

4.2.2 Fibbee protocol "intolerance"

To modularly implement fibbee protocols in DEL, we could stipulate new members of the accessibility relation whenever it is the case that a contradiction is considered, such that the agent is redirected to their revised considerations on the awareness of having received misinformation. This implies a two-step event update. The first step would involve the detection of misinformation, and the second step would involve the fibbee protocol-specific reaction to misinformation having been detected. Then, different fibbee protocols would differ only in respect to the sort of revised considerations they have in place of contradiction.

This modularity is very desirable for future work. Suppose later we want to test the effect of a new fibbee protocol on our logic. All we'd have to do is change that latter part of the event update step corresponding to the specific fibbee protocol in use. The rest of the logic would remain identical.

4.3 Dynamic epistemic logic with misinformation by fibbee protocol "intolerance" (DELMI$_{\text{int}}$)

4.3.1 On modality notation

We will return to the $\Box$ and $\Diamond$ notation of S5 instead of the K_g and $\hat{K}_g$ of epistemic S5 and DEL. This is to avoid implicitly making the claim that truth is not necessary for knowledge. Without this change in notation, a logic of misinformation would imply that an agent despite being misinformed that p nevertheless "knows" that p. Since it isn't our objective to argue one way or the other regarding the definition of knowledge, we'll simply change our notation going forward. Besides appearance, the logical behavior of our modality will remain the same.[13]

4.3.2 Worldviews

Definition. An agent's *worldview* is the set of all internally consistent states that is accessible either immediately or transitively from the factual state.

Definition. The *worldview function* is $D : \bigcup\langle\mathcal{M}, s\rangle \times \text{AGT} \to \mathcal{P}(S)$, where $\bigcup\langle\mathcal{M}, s\rangle$ is the set of all epistemic models in the logic and S is the set of states comprising part of the definition of $\mathcal{M}$.

For readability we'll use subscripts and superscripts to present the parameters for worldviews. For example, $D(\langle\mathcal{M}, s\rangle, g)$ will instead have the notation $D_g^{\langle\mathcal{M},s\rangle}$.

Specifically, for some epistemic model $\langle\mathcal{M}, s\rangle$ and set of agents AGT,
$D_g^{\langle\mathcal{M},s\rangle} = \{t | \langle s, t\rangle \in R_g^+ \text{ and } g \in \text{AGT and } \neg\exists\phi \in \mathcal{L}_{\text{DELMI}_{\text{int}}} \vDash_t^{\mathcal{M}} \Box_g\phi \wedge \Box_g\neg\phi\}$, where R_g^+ is the transitive closure of the epistemic accessibility relation for agent g.

Using worldviews we can now capture the context-based truth definitions of a logic of misinformation, since agents rely on arriving at the contradictory consideration (generally,

[13]Further, we technically shouldn't call a logic of misinformation an *epistemic* logic. We won't go so far as to change that too, if only just for convenience of reference and to acknowledge that this logic of misinformation is an extension of DEL.

all contradictions) to react to misinformation and yet *ultimately* do not seriously consider the contradictory consideration.

4.3.3 Epistemic advantage and disadvantage

In a logic of misinformation, an agent g has an epistemic advantage over another agent f when g has a worldview which properly encompasses the worldview of f. For example, if it is the case that p but Alice lies to Bill and Bill now thinks that $\neg p$, then Alice has epistemic advantage over Bill because Bill is unaware of Alice's having lied to him. On the other hand, if Alice lies to Bill but Bill is unshaken in his information from stronger source that it is actually p which is the case, then Bill has epistemic advantage over Alice because Alice is unaware of Bill's being aware of Alice's having lied to him.

We will see these intuitions play out in later examples, but for now we formally define these concepts of epistemic advantage and disadvantage.

Definition. An agent g is in a state of *epistemic advantage* relative to agent f when the worldview of agent g is a strict superset of the worldview of agent f. That is, agent g has epistemic advantage relative to agent f in $\langle \mathcal{M}, s\rangle$ just when $D_g^{\langle \mathcal{M},s\rangle} \supset D_f^{\langle \mathcal{M},s\rangle}$.

And vice versa:

Definition. An agent g has *epistemic disadvantage* relative to agent f in $\langle \mathcal{M}, s\rangle$ just when $D_g^{\langle \mathcal{M},s\rangle} \subset D_f^{\langle \mathcal{M},s\rangle}$.

Definition. If neither $D_g^{\langle \mathcal{M},s\rangle} \supset D_f^{\langle \mathcal{M},s\rangle}$ nor $D_g^{\langle \mathcal{M},s\rangle} \subset D_f^{\langle \mathcal{M},s\rangle}$ then we say that the two agents g and f are in *neither epistemic advantage nor disadvantage* relative to each other.

Of course, for all of these definitions it is provided $g, f \in AGT$ where AGT is the set of agents corresponding to the model $\langle \mathcal{M}, s\rangle$.

4.3.4 Definitions

The language of ***d**ynamic **e**pistemic **l**ogic with **mi**sinformation by fibbee protocol "**int**olerance"* ($\text{DELMI}_{\text{int}}$) $\mathcal{L}_{\text{DELMI}_{\text{int}}}$ is the set of formulas captured by the following BNF:

$$\phi ::= p|\neg\phi|\phi \wedge \phi|\Box_g \phi|[\langle \mathcal{E}, e\rangle]\phi$$

where $p \in$ ATM and ATM is the set of all atomic propositions (atoms), $g \in$ AGT and AGT is the set of all epistemic agents, and $\langle \mathcal{E}, e\rangle$ is an event model (see below).

An *unpointed pre-update epistemic model* in $\text{DELMI}_{\text{int}}$ $\mathcal{M}$ is defined as a 3-tuple $\langle S, R, V \rangle$, where

S is a non-empty set of states,

R is a mapping from agents to sets of pairs of states; i.e. $R : \text{AGT} \to \mathcal{P}(S^2)$, where

R_g is a set of pairs of states called the *epistemic accessibility relation for agent g*, and

V is a mapping from states to a mapping from atoms to a member of the set $\{0, 1\}$ called the *valuation function*; i.e. $V : S \to (ATM \to \{0, 1\})$, where

V_s is a mapping from atoms to a member of the set $\{0, 1\}$ called the *valuation for state s*; i.e. $V_s : ATM \to \{0, 1\}$, where

$V_s(p)$ is a member of the set $\{0, 1\}$ called the *truth assignment for atom p at state s.*

A *pre-update epistemic model* in $\text{DELMI}_{\text{int}}$ is a 2-tuple $\langle \mathcal{M}, s_0 \rangle$ where $s_0 \in S$ is the factual state.

Once again we also first need to define what an *event model* is before we can move on to truth definitions.

An *unpointed event model* in $\text{DELMI}_{\text{int}}$ $\mathcal{E}$ is defined as a 3-tuple $\langle E, Q, P \rangle$, where

E is a non-empty set of *events*,

Q is a mapping from agents to sets of pairs of events; i.e. $Q : \text{AGT} \to \mathcal{P}(E^2)$, where

Q_g is a set of pairs of events called the *event accessibility relation for agent g*,

P is a mapping from events to formulas called the *purported informational content function*; i.e. $P : E \to \mathcal{L}_{\text{DEL}}$, where

$P(e)$ is a formula called the *purported informational content for event e.*

Notice that the precondition function P in DEL has been replaced with the "purported informational content function" in $\text{DELMI}_{\text{int}}$. This is because there is now no need for a function to track the preconditions for events, at least so far as events of saying something are concerned; in principle, any agent can say anything regardless of the truth-value of the informational content of what is said. At the same time, it is still important to track what is being purported as true by what is being said, and so this function still serves an important but distinctly different purpose in $\text{DELMI}_{\text{int}}$.

An *event model* in $\text{DELMI}_{\text{int}}$ is a 2-tuple $\langle \mathcal{E}, e_0 \rangle$, where $e_0 \in E$ is the factual event.

The truth definitions for $\text{DELMI}_{\text{int}}$ are:

$$
\begin{array}{lll}
\vDash_s^{\mathcal{M}} p & \text{iff} & V_s(p) = 1 \\
\vDash_s^{\mathcal{M}} \neg\phi & \text{iff} & \nvDash_s^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \phi \wedge \psi & \text{iff} & \vDash_s^{\mathcal{M}} \phi \text{ and } \vDash_s^{\mathcal{M}} \psi \\
\vDash_s^{\mathcal{M}} \phi \vee \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg(\neg\phi \wedge \neg\psi) \\
\vDash_s^{\mathcal{M}} \phi \rightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg\phi \vee \psi \\
\vDash_s^{\mathcal{M}} \phi \leftrightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi) \\
\vDash_s^{\mathcal{M}} \Box_g\phi & \text{iff} & \forall t \in D_g^{\langle \mathcal{M},s\rangle} \vDash_t^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \Diamond_g\phi & \text{iff} & \vDash_s^{\mathcal{M}} \neg \Box_g \neg\phi \\
\vDash_s^{\mathcal{M}} [\langle \mathcal{E}, e\rangle]\phi & \text{iff} & \vDash_s^{\langle \mathcal{M},s\rangle \otimes \langle \mathcal{E},e\rangle} \phi
\end{array}
$$

where $\otimes$ is the *event update operator* such that an *updated epistemic model* $\langle \mathcal{M}, s\rangle \otimes \langle \mathcal{E}, e\rangle$ is a 2-tuple $\langle \mathcal{M}', s[e]\rangle$, where $\mathcal{M}'$ is a 3-tuple $\langle S', R', V'\rangle$ called an *unpointed updated epistemic model.*

The updated epistemic model $\langle \mathcal{M}', s[e]\rangle = \langle \mathcal{M} \otimes \mathcal{E}, s[e]\rangle$ is determined by a two-step process:

Step 1: Assimilation

$S'_0 = \{s[e] | s \in S \text{ and } e \in E\}$,
$R'_0 = \{\langle g, R'_{0,g}\rangle | R'_{0,g} = \{\langle s[e], t[f]\rangle | s[e], t[f] \in S'_0 \text{ and } \langle s,t\rangle \in R_g \text{ and } \langle e, f\rangle \in Q_g\}\}$,
and
$V' = \{\langle s[e], V_s\rangle | s[e] \in S'_0\}$.

$\langle S'_0, R'_0, V'\rangle$ is called the *assimilation submodel of* $\langle \mathcal{M} \otimes \mathcal{E}, s[e]\rangle$.

Step 2: Reconsideration

In the case of $\text{DELMI}_{\text{int}}$ the fibbee protocol applied in this step is "intolerance":
$S' = S'_0 \cup S_g$, where
$S_g = \{s_g | s \in S \text{ and } s \in D_g^{\langle \mathcal{M},s\rangle}\}$
$R' = \{\langle g, R'_g\rangle | g \in AGT\}$, where

$$
R'_g = \begin{cases} R'_{0,g} \cup \langle t[f], t_g\rangle \cup R_g & \text{if} \vDash_{s[e]}^{\mathcal{M}\otimes\mathcal{E}} \Box_g(P(e) \wedge \neg P(e)) \\ R'_{0,g} & \text{otherwise} \end{cases},
$$

$t_g \in S_g$,

and

R_g is a component of $\mathcal{M}_g$, called the *reconsideration submodel for agent* g *in* $\langle \mathcal{M} \otimes \mathcal{E}, s[e] \rangle$, where

$\mathcal{M}_g = \mathcal{M} \cap g = \langle S_g, R_g, V \rangle$.

Notes: An updated epistemic model $\langle \mathcal{M}, s \rangle \otimes \langle \mathcal{E}, e \rangle$ has factual state $s[e] \in S'$.
It's possible for S' to contain states which is not contained in any agent's worldview, i.e. to which no agent has accessibility from the factual state. By convention such states may be omitted in presentations of epistemic models with no loss in expressiveness.

The intuition behind this definition of event update is that agents will first tentatively consider the purported informational content of an observed event before assessing whether that information conflicts with their existing consideration base. These two steps correspond respectively with the assimilation step, in which the incoming information is naively *assimilated* into the agent's existing consideration base; and the reconsideration step, in which the agent immediately afterwards *considers again* (reconsiders) the naively updated consideration base in order to root out contradictions that have arisen from naive update and tentative consideration.

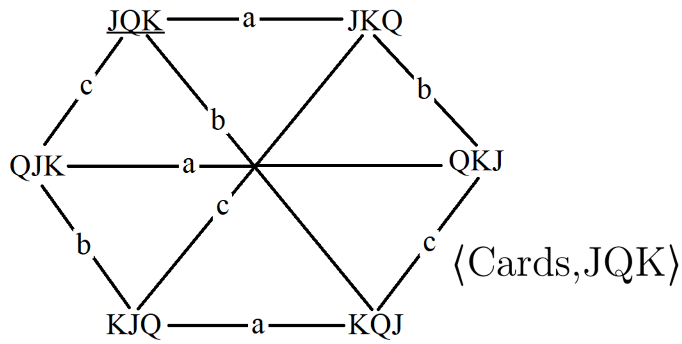
Further, the reconsideration step prunes away only those contradictions by adding accessibility to states with valuations identical to corresponding states of the agent's former worldview. This way, intolerance means for an agent to disregard the new information completely when realizing that they've received misinformation. Adding that accessibility accomplishes precisely the effect of worldview reversion, such that the consideration base is preserved.

4.4 Back to Cards

Let's return to Cards with these new tools at our disposal. Does $\text{DELMI}_{\text{int}}$ fare better than DEL at expressing misinformation events?

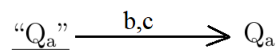
4.4.1 Misinformation

Once again this is the pre-update epistemic model:



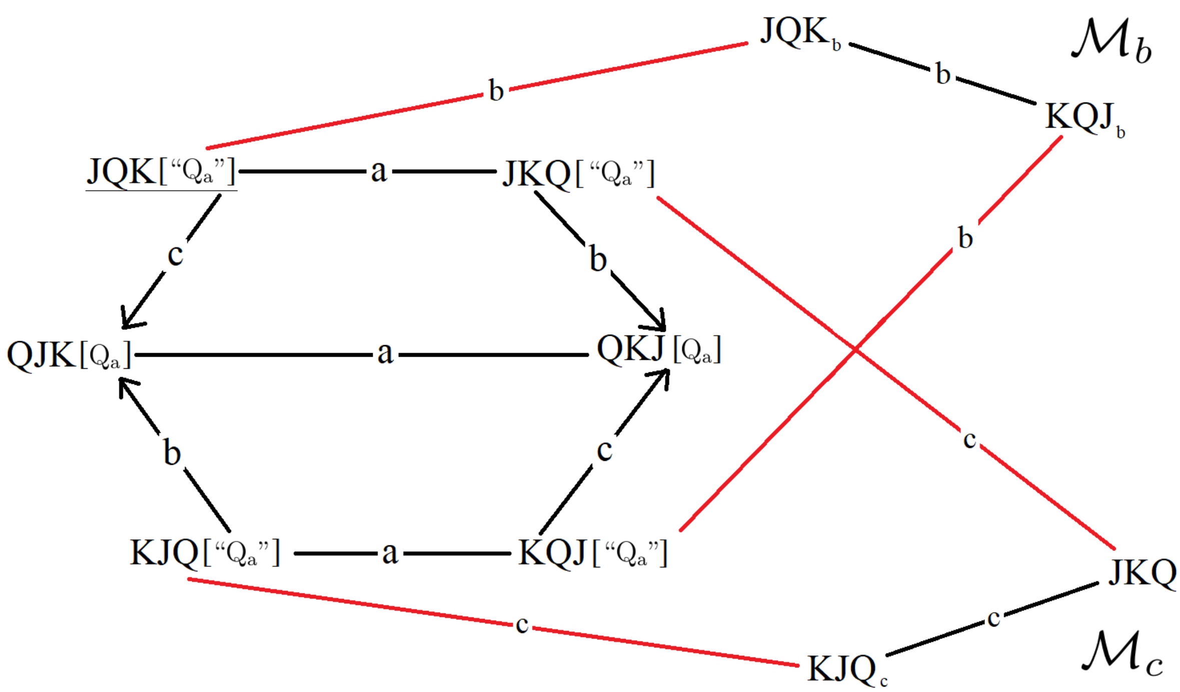
⟨Cards,JQK⟩

And here was the event model:



⟨LieQueenAll,“Q_a”⟩

The updated epistemic model now is[14]:



⟨Cards⊗LieQueenAll,JQK[“Q_a”]⟩

Figure 12: After update with the ⟨LieQueenAll,“Q_a”⟩ event model, using $\text{DELMI}_{\text{int}}$

Rearranging the positions of some of the states in the figure such that reconsideration

[14]The assimilation submodel is to the left while the two reconsideration submodels are to the top and bottom right. The first accessibilities added in during the reconsideration step are highlighted in red. These grant the fibbee accessibility from the assimilation submodel to the reconsideration submodel. That is, the red accessibilities represent Bill’s and Carol’s updating as per the fibbee protocol intolerance provisional on their discovery of having received misinformation from Alice.

submodels are at the top and the assimilation submodel is at the bottom demonstrates the relationship between misinformation and event ignorance:

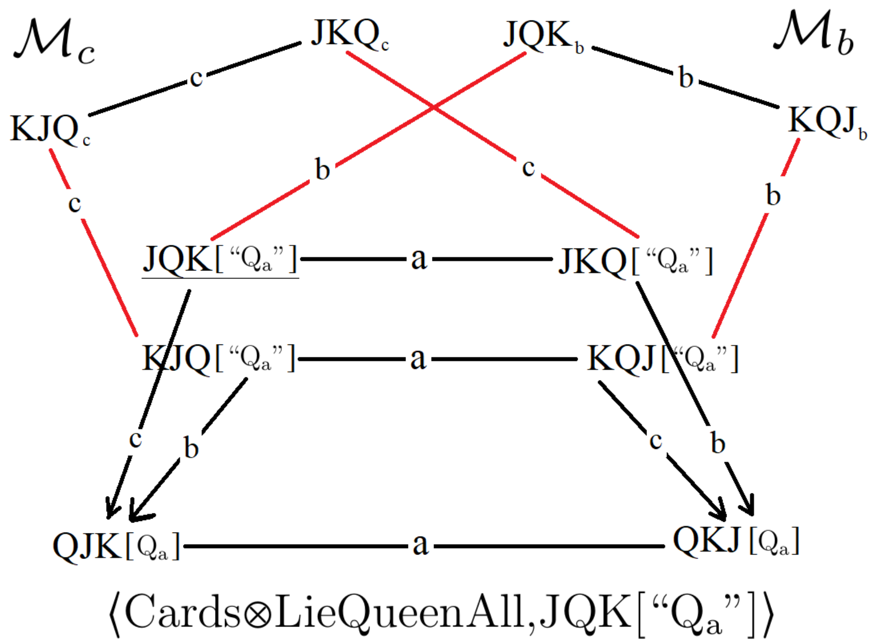
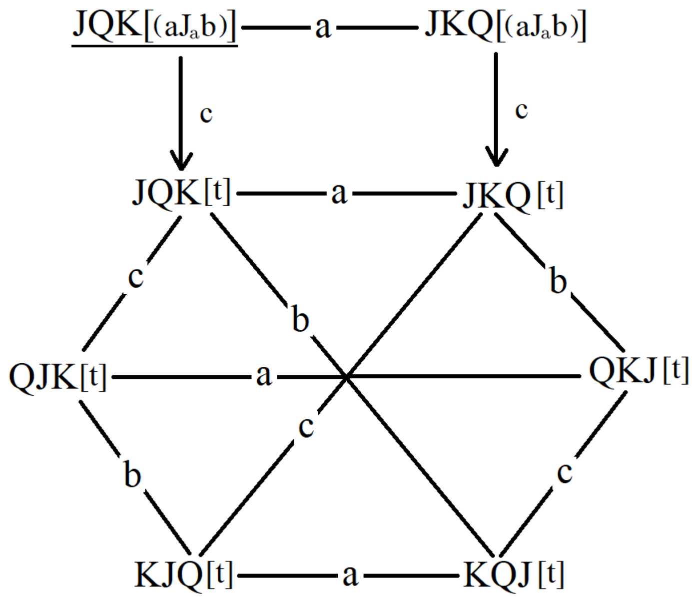


Figure 13: Alternative view of after update with the ⟨LieQueenAll,“Q_a”⟩ event model, using $\text{DELMI}_{\text{int}}$

Similar to event ignorance, in cases of successful misinformation there are agents whose worldviews are misdirected to the subset of the states underneath among which there is no state considered with facts the same as that of the factual state. This misdirection occurs entirely within assimilation submodels, which makes sense because successful misinformation does not trigger any change in the agent's consideration base at the reconsideration step of event update.

To observe this, compare the updated epistemic model with that of event ignorance:



⟨Cards⊗SecretShow,JQK[(aJ$_a$b)]⟩

However, unlike in event ignorance there is also fibbee protocol action in the new reconsideration step of event update. Their effect is expressed as the reconsideration submodels observed. While those succumbing to the misinformation unwittingly follow one-way accessibility to the bottom half of the assimilation submodel, those aware of having been the recipient of misinformation gain special access to their reconsideration submodel. This occurs during the reconsideration step of event update.

The net epistemic effect of all of this is that a fibber will have epistemic advantage over those fibbees who fail to realize that they've been lied to, while fibbees which realize that they've been lied to will have epistemic advantage over the fibber. Thereby $\mathrm{DELMI}_{\mathrm{int}}$ overcomes the added complexity that misinformation bears beyond event ignorance.

To summarize, the full picture with the event model and pre-update epistemic model is:

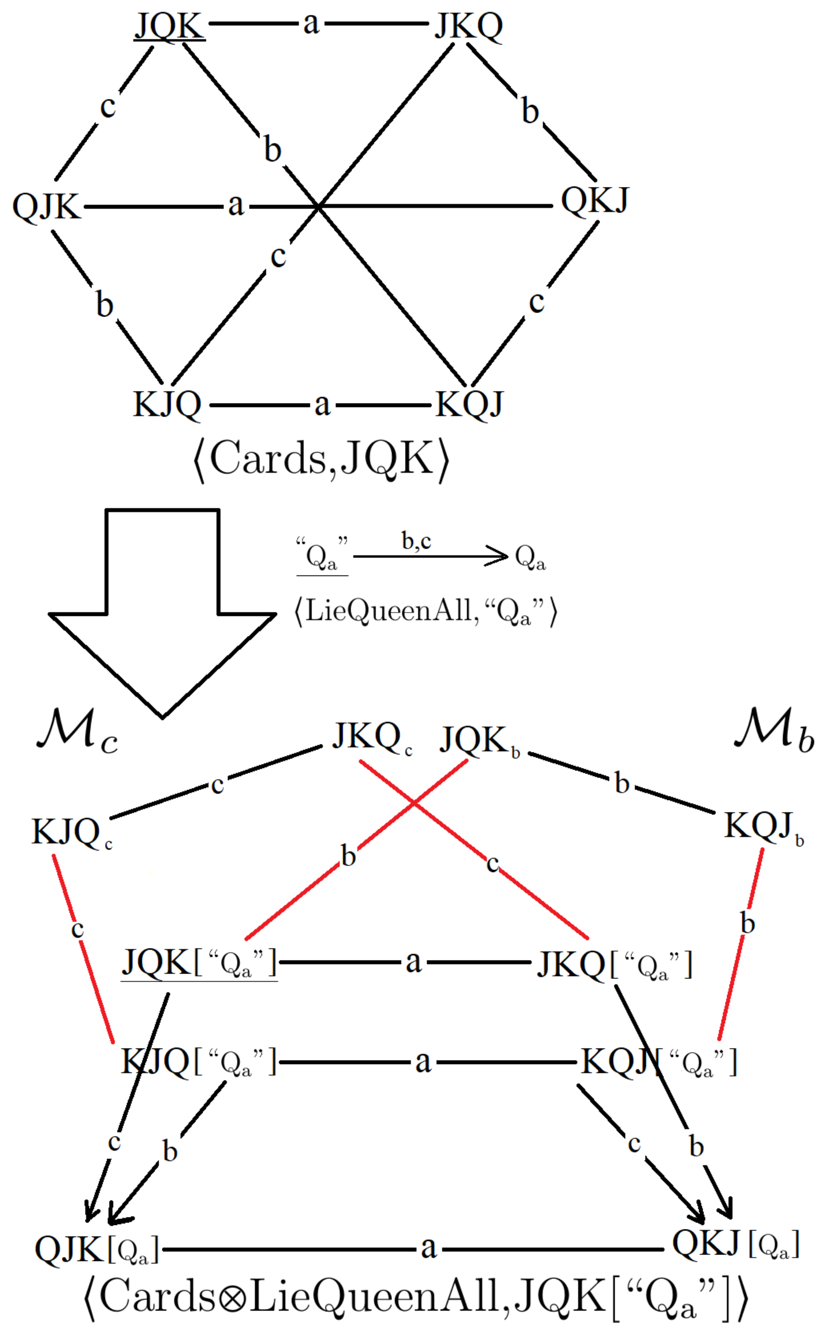


Figure 14: Update with the ⟨LieQueenAll, "Q_a"⟩ event model, using $\text{DELMI}_{\text{int}}$

Since in the Cards scenario there are three players and only three unique cards, there is initially complete information about all possible states for each agent. It's only because of this that Alice will always be aware of the possibility of Bill's or Carol's realization of having received misinformation.

In more complex scenarios that fibbees will fall either into the "misdirected" group below in the assimilation submodel or the "privy"[15] group above in reconsideration submodels. I predict that this would be such that privy fibbees not only consider possible the factual state but also something that the actor themself does not consider: that those fibbees are aware that what they've received from the actor is misinformation.

There are two other things to note. The first is that the effect of lying is, at least so far as the liar is concerned, assumed to be equivalent to the null event. This is effectively the same as the treatment of event update for actors in DEL. The second is that some states in the updated epistemic model are inaccessible from the actual state, for all agents.

4.4.2 Misinformation and content ignorance

What happens if misinformation is mixed with content ignorance?

If the occurrence of the event is known to Carol, Carol will assume that Alice is telling the truth, since that part of DEL is unchanged. Interestingly, even if Bill does not realize that Alice is lying, he will take for granted that Carol is following the two remaining possible states because of the assumption of unconditional trust.

We find that the result is (see next page):

[15] "Privy" in the sense that these agents are in the know regarding the misinformation going around.

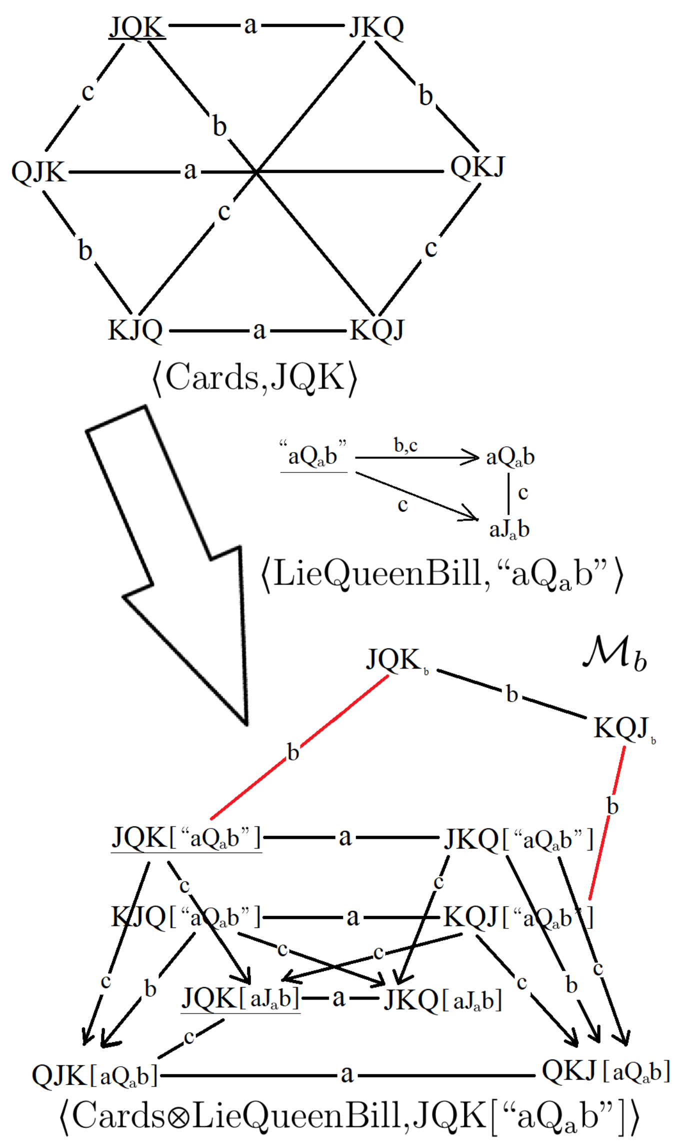


Figure 15: Update with the ⟨LieQueenBill,“aQ_ab”⟩ event model, using $\text{DELMI}_{\text{int}}$

The bottom portion of the updated model corresponds precisely with Carol's succumbing to misinformation. Since Carol is assuming the truth of the event, there is no further consideration of the states where Alice is lying about the card she holds. Also, in the case where Bill is unaware of the falsehood of the purported informational content of Alice's event, Bill's update results in his worldview being a strict subset of Carol's. This is to be expected as in comparison with the outcome in the case of content ignorance.

This shows that $\text{DELMI}_{\text{int}}$'s misinformation handling mechanism has no problem with expressing in Cards the epistemic outcomes of misinformation and content ignorance simultaneously.

4.4.3 Misinformation and event ignorance

What happens if misinformation is mixed with event ignorance?

If the occurrence of the lie is not known to Carol, then not only will Carol be misdirected to a submodel where her worldview is as if no event occurred, Bill will also be aware of this—regardless of whether Bill is himself misdirected or not.

The result is:

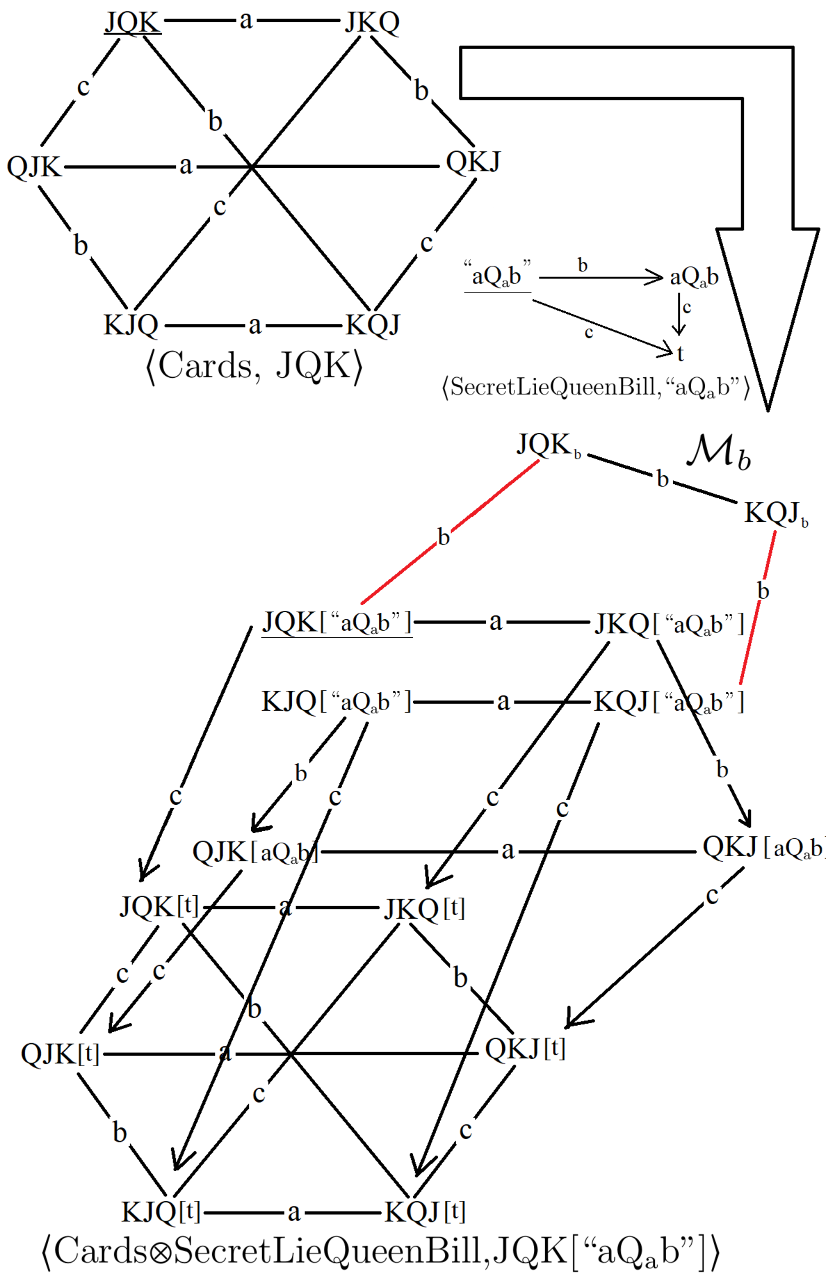


Figure 16: Update with the ⟨SecretLieQueenBill, "aQ_ab"⟩ event model, using $\text{DELMI}_{\text{int}}$

The bottom portion of the updated model corresponds precisely with Carol's being unaware of the occurrence of the event, and that Bill, even if himself misdirected, is aware of Carol's being misdirected. This demonstrates that $\mathrm{DELMI}_{\mathrm{int}}$ is able to express the intuition that it is possible for one to have an epistemic advantage over another agent while being in relative epistemic disadvantage to yet another agent. In this case Bill would be in epistemic disadvantage relative to Alice while in epistemic advantage relative to Carol.

It is shown that the misinformation handling mechanism of $\mathrm{DELMI}_{\mathrm{int}}$ has no problem with expressing in Cards the epistemic outcomes of misinformation and event ignorance simultaneously.

4.4.4 Misinformation and simultaneous event ignorance

To further demonstrate $\mathrm{DELMI}_{\mathrm{int}}$'s ability to express multiple levels of epistemic advantage, consider updating the pre-update epistemic model $\langle \mathcal{M}, \mathrm{JQK} \rangle$ with the event model $\langle \mathcal{E}, \text{“aXg”}2 \rangle$, where “aXg”2 is shorthand for the event “a$\mathrm{Q_a}$b” $\wedge$ “a$\mathrm{K_a}$c”, that Alice secretly lies to Bill and Carol that she holds the Queen and the King respectively.

The result is (see next page):

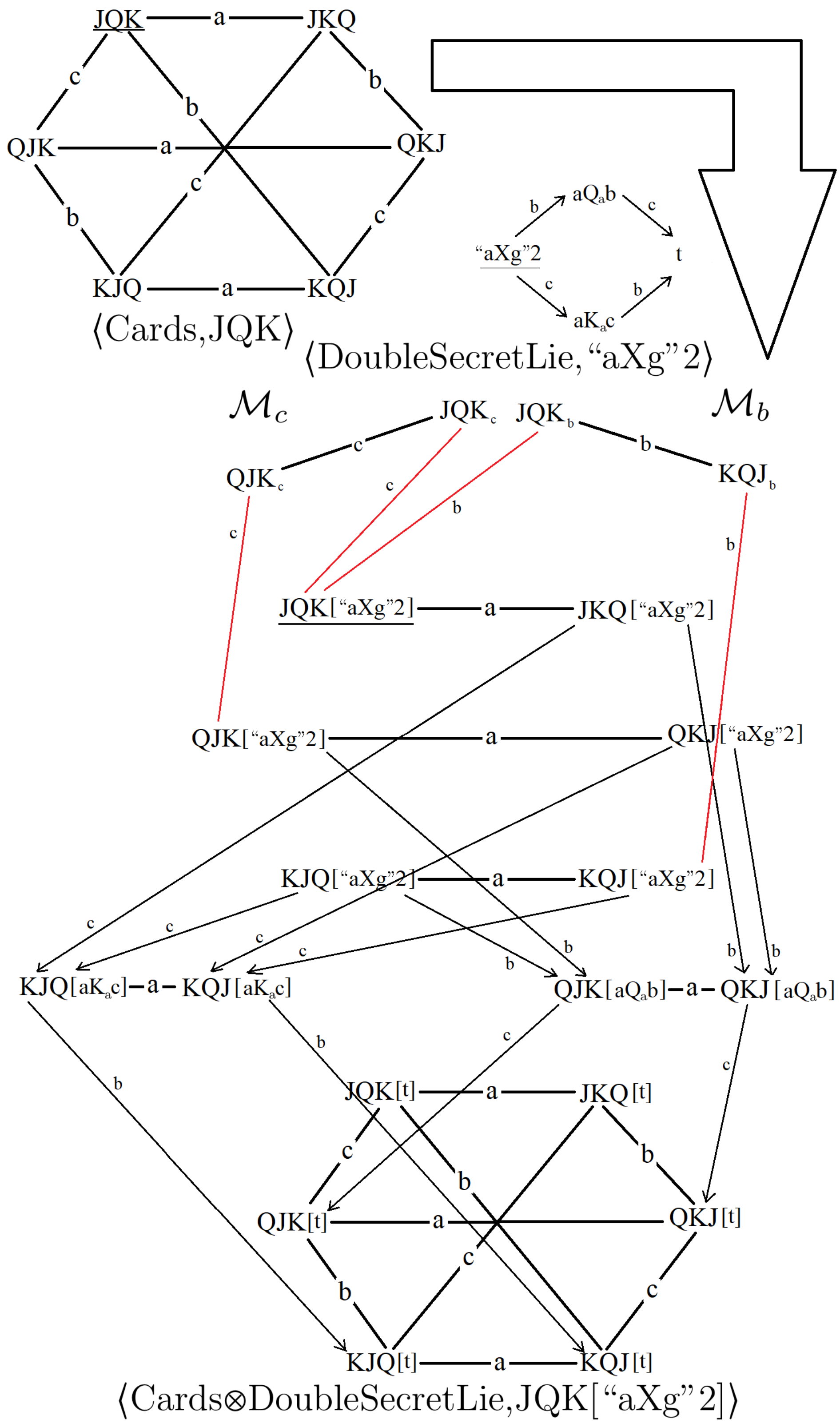


Figure 17: Update with the ⟨DoubleSecretLie,“aXg”2⟩ event model, using $\text{DELMI}_{\text{int}}$

$\mathcal{M}_c$ is different from those of the preceding models because the reconsideration step will expand the existing accessibility relations for fibbees only with respect to conditions under which they would realize that they've become misinformed. In the case of Carol this means she will become a privy fibbee only if the actual state involves her holding the King.

No agent has access between the states $\mathrm{JQK_c}$ and $\mathrm{JQK_b}$. While this is so due to the construction during the reconsideration step, transitivity allows that both Bill and Carol, if both aware of becoming misinformed, will consider the possibility of it being the state where both of them simultaneously reject the purported informational content of Alice's event (namely, JQK).

4.5 Downsides of fibbee protocol "intolerance"

The main drawback of the intolerance protocol is that once misinformation has managed to work its way into one's epistemic framework, it cannot be excised even by later truth-telling, since later received truths will be ignored out of hand as misinformation.

This is a major drawback, which is to be expected of a protocol as simpleminded as intolerance. Perhaps the greatest worry is that intolerance does not comport well with our experience of reality. It is a common experience throughout one's life (and especially in youth) to shed past considerations that have since been discovered to have been mistaken. Those that adhere strictly to past considerations, perhaps even constructing elaborate structures of conspiracy to design a worldview which still agrees with those past considerations, are said to be stubborn or closed-minded, and epistemically are in no better position than our hypothetical agents adhering to the intolerance protocol.

The silver lining is that on the other hand where there is already a strong epistemic foundation, intolerance is very well protected against later misinformation since it ignores the effect of later information entirely if it stands at odds with the existing consideration base. But in such a situation it's one's epistemic sources which are worthy of praise, not intolerance for its veridical robustness.

5 Misinformation handling in $\text{DELMI}_{\text{neg}}$

Now let's try a different fibbee protocol. Suppose that instead of ignoring the liar it could be taken that the purported informational content of the received event is the opposite of what it purports to be. Recall that we refer to this as the fibbee protocol "negation". The intuition is that this enables learning from lies as well as standard (truth-telling) events.

First I make the appropriate modifications to $\text{DELMI}_{\text{int}}$ in the following sections to produce $\text{DELMI}_{\text{neg}}$. As previously discussed the design of DELMIs is intended to be modular with respect to fibbee protocols.

Then, I show that $\text{DELMI}_{\text{neg}}$ is equivalent to $\text{DELMI}_{\text{int}}$. I argue that this equivalence holds in any modal epistemic logic. Finally I'll use this result to further analyze the difference between the intolerance and negation fibbee protocols.

5.1 Intuitions

For negation, we again need to add new members to the accessibility relation. However in this case these members should grant agents accessibility specifically only to states where the negation of the purported informational content of the misinformation event hold in the pre-update epistemic model.

5.2 Dynamic epistemic logic with misinformation by fibbee protocol "negation" ($\text{DELMI}_{\text{neg}}$)

Recall that fibbee protocols are implemented in the reconsideration step of the epistemic model update process. Therefore, $\text{DELMI}_{\text{neg}}$ is defined in the same way as $\text{DELMI}_{\text{int}}$ with the only differences being in reconsideration.

5.2.1 Definitions

The language of dynamic epistemic logic with misinformation on fibbee protocol "negate" ($\text{DELMI}_{\text{neg}}$) $\mathcal{L}_{\text{DELMI}_{\text{neg}}}$ is the same as with $\text{DELMI}_{\text{int}}$.

An *unpointed pre-update epistemic model* in $\text{DELMI}_{\text{neg}}$ $\mathcal{M}$ as well as the truth definitions for $\text{DELMI}_{\text{neg}}$ are defined the same as with $\text{DELMI}_{\text{int}}$.

$\mathcal{M}'$ is determined by a two-step process, where

Step 1: Assimilation
$S' = \{s[e] | s \in S \text{ and } e \in E\}$,
$R'_0 = \{\langle g, R'_{0,g}\rangle | R'_{0,g} = \{\langle s[e], t[f]\rangle | s[e], t[f] \in S'_0 \text{ and } \langle s, t\rangle \in R_g \text{ and } \langle e, f\rangle \in Q_g\}\}$,
and
$V' = \{\langle s[e], V_s\rangle | s[e] \in S'_0\}$,

and $\langle \mathcal{E}, e\rangle$ is an *event model*, where $\mathcal{E}$ is an *unpointed event model* defined as a 3-tuple $\langle E, Q, P\rangle$, where

E is a non-empty set of *events*,
Q is a mapping from agents to sets of pairs of events; i.e. $Q : \text{AGT} \to \mathcal{P}(E^2)$, where
Q_g is a set of pairs of events called the *event accessibility relation for agent g*,
P is a mapping from events to formulas called the *purported informational content function*; i.e. $P : E \to \mathcal{L}_{\text{DEL}}$, where
$P(e)$ is a formula called the *purported informational content for event e.*

Again this is the same as with $\text{DELMI}_{\text{int}}$ (section 4.4.3). But, the reconsideration step will be different, as the negation protocol will be applied instead of intolerance as the fibbee protocol.

Step 2: Reconsideration
In the case of $\text{DELMI}_{\text{neg}}$ the fibbee protocol applied in this step is "negation":
$S' = S'_0 \cup S_g$, where

$$S_g = \begin{cases} \{s_g | s \in S \text{ and } \vDash^{\mathcal{M}}_s \neg P(e)\} & \text{if } \vDash^{\mathcal{M}\otimes\mathcal{E}}_{s[e]} \Box_g(P(e) \wedge \neg P(e)) \\ \{\} & \text{otherwise} \end{cases}$$

$R' = \{\langle g, R'_g\rangle | g \in AGT\}, where$

$$R'_g = \begin{cases} R'_{0,g} \cup \{\langle t[f], t_g\rangle | t[f] \in S'_0 \text{ and } t_g \in S_g\} \cup R_g & \text{if} \vDash^{\mathcal{M}\otimes\mathcal{E}}_{s[e]} \Box_g(P(e) \wedge \neg P(e)) \\ R'_{0,g} & \text{otherwise} \end{cases},$$

$t_g \in S_g$,
and
R_g is a component of $\mathcal{M}_g$, called the *reconsideration submodel for agent* g *in* $\langle \mathcal{M} \otimes \mathcal{E}, s[e]\rangle$, where
$\mathcal{M}_g = \mathcal{M} \cap g = \langle S_g, R_g, V\rangle$

Notes: An updated epistemic model $\langle \mathcal{M}, s\rangle \otimes \langle \mathcal{E}, e\rangle$ has factual state $s[e] \in S'$.
It's possible for S' to contain states which are not contained in any agent's worldview, i.e. to which no agent has accessibility from the factual state. By convention such states may be omitted in presentations of epistemic models with no loss in expressiveness.

5.3 Back to Cards

Recall that the misinformation event in Cards is Alice telling both Bill and Carol that she has the Queen.

The outcome for misinformation now is:

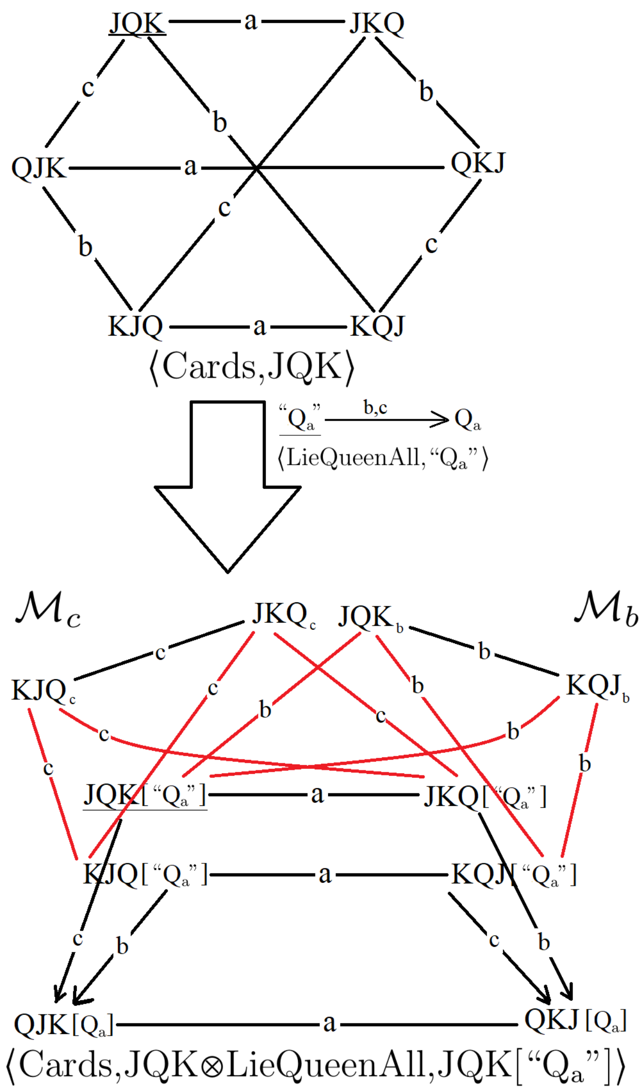


Figure 18: After update with the $\langle$LieQueenAll,"Q_a"$\rangle$ event model, using $\text{DELMI}_{\text{neg}}$

We see that the only difference with the analogous epistemic models using $\mathrm{DELMI}_{\mathrm{int}}$ is that there are more accessibilities between the assimilation and reconsideration submodels. These extra accessibilities also have no effect on the worldviews of agents because accessibilities remain transitive in $\mathrm{DELMI}_{\mathrm{neg}}$.

There are also additional states added in during reconsideration but to which the agent in concern has no access; these states can be safely ignored. These additional states do not appear in $\mathrm{DELMI}_{\mathrm{int}}$ because the definition of S' in $\mathrm{DELMI}_{\mathrm{int}}$ is agent-dependent rather than dependent on the truth-value of the purported informational content of the event.[16] For example, in $\mathrm{DELMI}_{\mathrm{neg}}$ the reconsideration submodel $\mathcal{M}_c$ contains also the states $\mathrm{JQK}_{\mathrm{c}}$ and $\mathrm{KQJ}_{\mathrm{c}}$, but these states are omitted because $\langle c, R'_c \rangle$ is the only member of R'.

Again we can see that rather than only downward epistemic misdirection from the factual state, under the misinformation event there is also privy fibbees travelling upwards such that they're specially aware of the untruth of the purported informational content of the most recent event.

Similarly the outcomes for open private misinformation, secret private misinformation, and two-way secret private misinformation in $\mathrm{DELMI}_{\mathrm{neg}}$ follow (see following pages):

[16]These states are different than the others to which no agent has access, since they're included at different stages of the update process. In principle, the update process could be modified to trim off these excess states, but still they would have no effect on resultant worldviews.

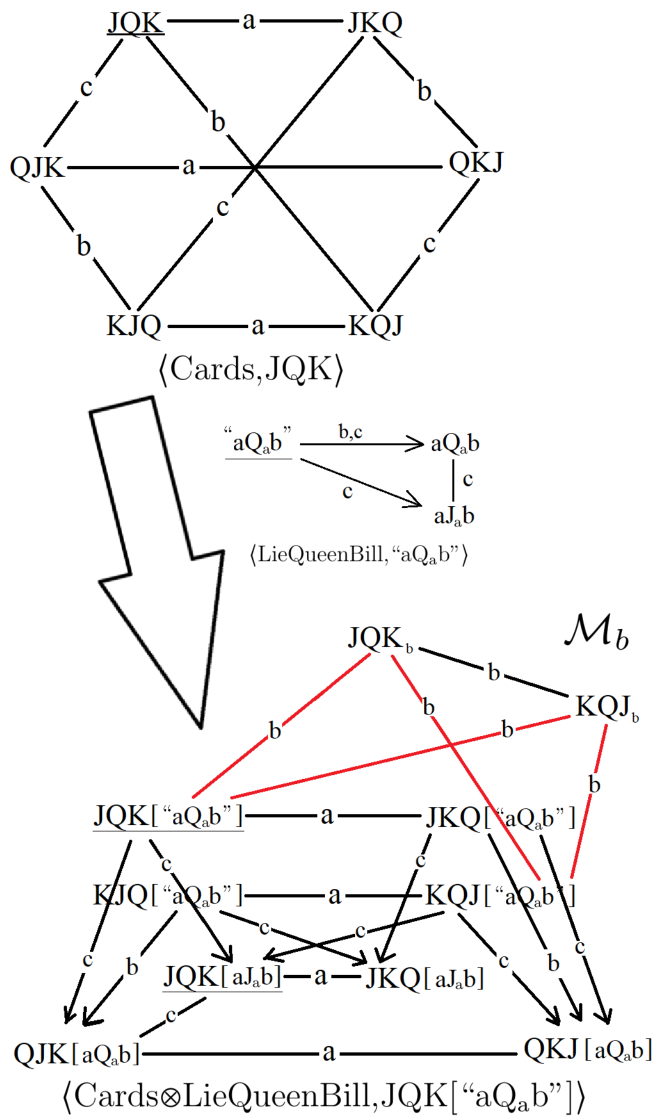


Figure 19: Update with the ⟨LieQueenBill,"aQ_ab"⟩ event model, using $\text{DELMI}_{\text{neg}}$

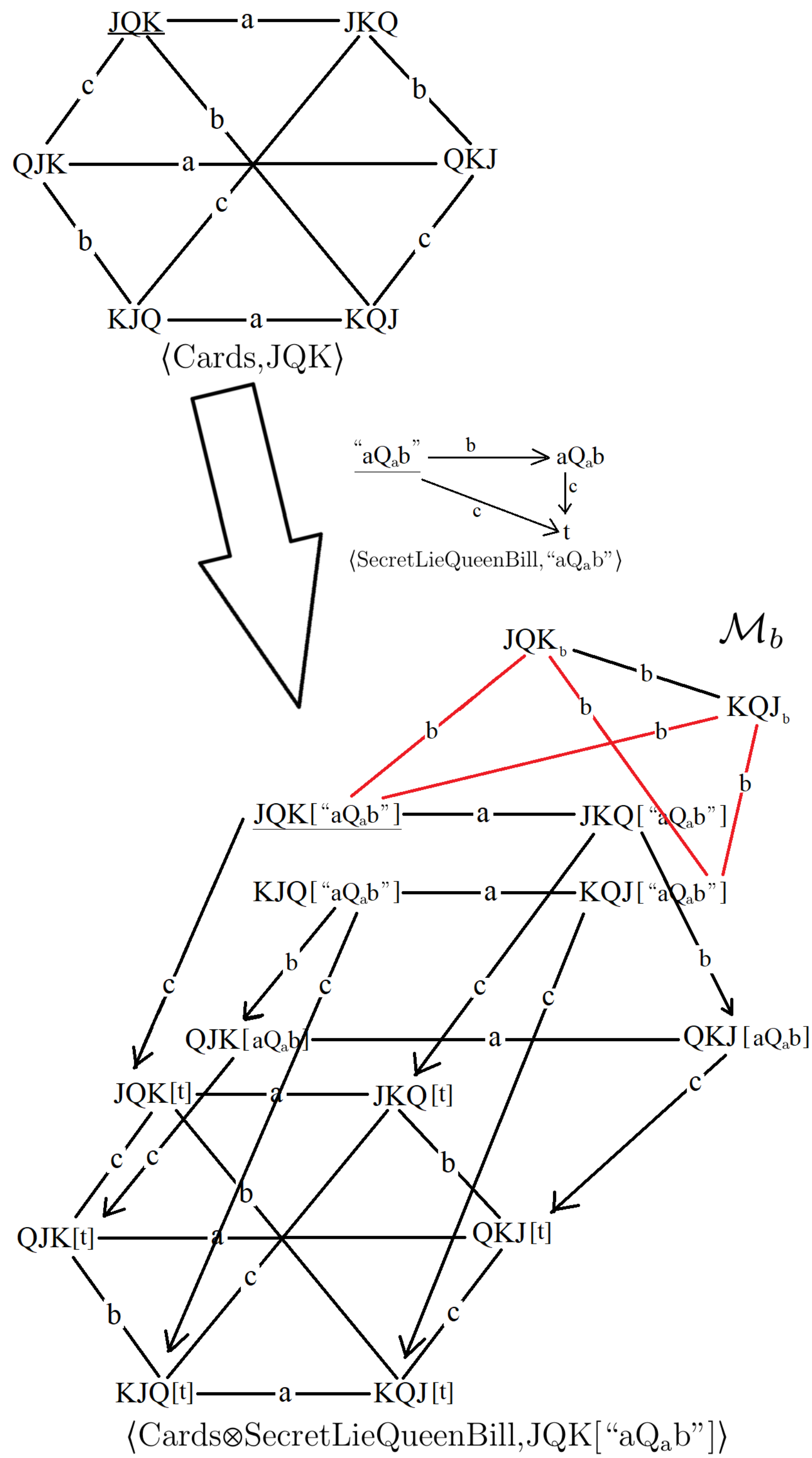


Figure 20: Update with the ⟨SecretLieQueenBill, "$\text{aQ}_\text{a}\text{b}$"⟩ event model, using $\text{DELMI}_{\text{neg}}$

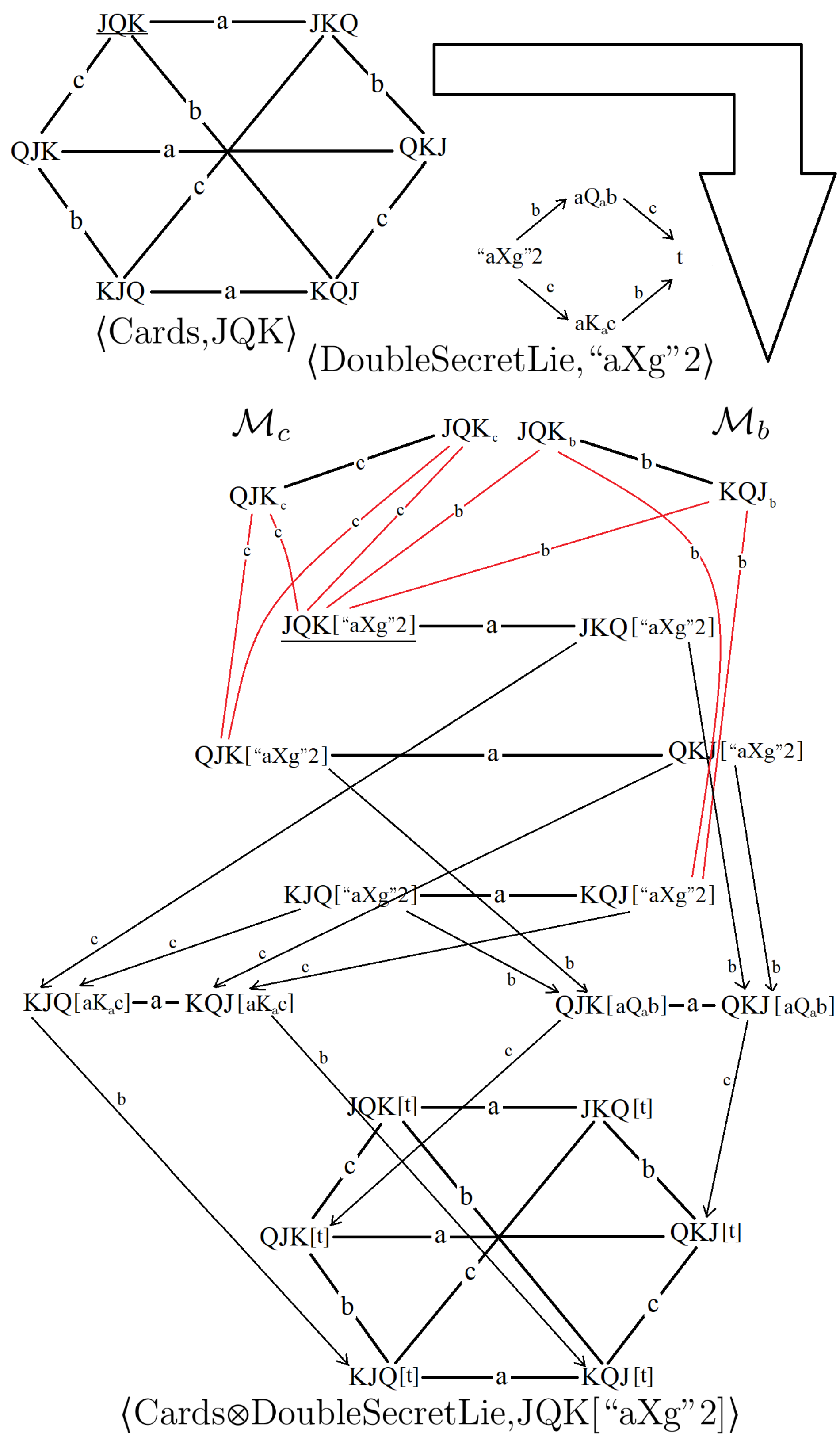


Figure 21: Update with the ⟨DoubleSecretLie,"aXg"2⟩ event model, using $\mathrm{DELMI}_{\mathrm{neg}}$

"aXg"2 stands for "Alice lies differently to two different agents", where in this case the lies are claiming to Bill to have the Queen and claiming to Carol to have the King.

We suspect at this point that $\text{DELMI}_{\text{neg}}$ is equivalent to $\text{DELMI}_{\text{int}}$. I now demonstrate that this is indeed the case. Afterwards we will discuss whether this is the case for all logics, and if not, which types of logics for which we can expect the intolerance and negation fibbee protocols to produce equivalent logics of misinformation.

5.4 Equivalence of $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$

5.4.1 The general idea

Generally speaking, the reason for the equivalence of $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ inheres in two-valued logic, specifically because in $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ formulas are either true or false but not both.

If an agent arrives at an internally inconsistent consideration base as a result of update, then it must have been the case that they were already subscribed to the negation of the purported informational content of the event beforehand. We can make this inference in logic where formulas evaluate to either true or false but not both, such as DEL; this was previously demonstrated.

Hence, whether the agent ignores the event or updates with the negation of the purported informational content of the event, the agent will arrive at the same consideration base after updating. Therefore, ignoring the event is the same as rejecting it to accept its negation.

5.4.2 Proof

Definition. The *epistemic output* of an epistemic model $\langle \mathcal{M}, s \rangle$ is the set of all formulas true in the model with a consideration operator as its main operator. Articulating this formally, let us say that:
$out : \{\langle \mathcal{M}, s \rangle | \langle \mathcal{M}, s \rangle$ is a model of the logic in use$\} \to \mathcal{P}(\mathcal{L})$ is such that
$out(\langle \mathcal{M}, s \rangle) = \{\Box_g \phi | \exists g \in AGT \vDash_s^{\mathcal{M}} \Box_g \phi, \phi \in \mathcal{L}\}$.

Definition. Two dynamic epistemic logics of misinformation are *equivalent* if and only if for every pair of epistemic models from each logic which encode the same epistemic situation their epistemic output are identical.

Theorem. $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ are equivalent.

Proof.

Consider two pre-update epistemic models $\langle \mathcal{M}, s \rangle$ and $\langle \mathcal{M}2, s2 \rangle$ for each of $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ respectively. Also suppose that they encode the same epistemic situation; that is, that the pre-update epistemic models and the event are identical. Then the set of provable formulas from each model are identical, because pre-update epistemic models encode epistemic situations in the same way in any DELMI. A fortiori, the epistemic outputs of each model are also identical.

Events are either misinformation events or they are not misinformation events.

Suppose $\langle \mathcal{M}, s \rangle$ and $\langle \mathcal{M}2, s2 \rangle$ are updated with the same misinformation event $\langle \mathcal{E}, e \rangle$.

Then for every agent $g \in AGT$, either a contradiction emerges as seriously considered during the assimilation step of update or no contradiction emerges as seriously considered during the assimilation step of update.

If a contradiction emerges as seriously considered during the assimilation step of update, then in each of $\langle \mathcal{M}, s \rangle$ and $\langle \mathcal{M}2, s2 \rangle$ it was already the case that $\Box_g \neg P(f)$ for some $f \in E$.

Then, at the assimilation step, the fibbee protocol intolerance will add to the worldview of g only states from $\langle \mathcal{M}, s \rangle$ such that it is the case that $\Box_g \neg P(f)$. This is because intolerance will cause the purported informational content of the misinformation event to be ignored, and so $\Box_g \neg P(f)$ will remain a member of the epistemic output just as it was previously for the pre-update epistemic model. On the other hand, the fibbee protocol negation will add to the worldview of g only states at which it is not the case that $P(f)$. Since that was already true of g's worldview in the pre-update epistemic model, this is equivalent to redundantly adding to the worldview of g only states from $\langle \mathcal{M}, s \rangle$ such that it is the case that $\Box_g \neg P(f)$.

If no contradiction emerges as seriously considered during the assimilation step of update, then it was not already the case that $\Box_g \neg P(f)$ for some $f \in E$ in $\langle \mathcal{M}, s \rangle$ and $\langle \mathcal{M}2, s2 \rangle$ and so the event cannot be identified as a misinformation event. Since fibbee protocols only affect what happens during the reconsideration step of event update and the reconsideration step of event update does not change the epistemic output when there is no detection of misinformation after the assimilation step, and $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ differ only in their fibbee protocols, the epistemic output of $\langle \mathcal{M}, s \rangle$ and $\langle \mathcal{M}2, s2 \rangle$ updated with $\langle \mathcal{E}, e \rangle$, respectively $\langle \mathcal{M} \otimes \mathcal{E}, s[e] \rangle$ and $\langle \mathcal{M}2 \otimes \mathcal{E}, s2[e] \rangle$, must be identical. (All epistemic models, event models, and non-reconsideration steps of event update are defined identically between $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$.)

If $\langle \mathcal{E}, e \rangle$ is not a misinformation event, then for every agent $g \in AGT$, no contradiction emerges as seriously considered during the assimilation step of update. Then for the same reason as the case where $\langle \mathcal{E}, e \rangle$ is a misinformation event not detected as such by the agent after assimilation, the epistemic outputs of $\langle \mathcal{M} \otimes \mathcal{E}, s[e] \rangle$ and $\langle \mathcal{M}2 \otimes \mathcal{E}, s2[e] \rangle$ must be identical. That is, once again the resultant epistemic models will be identical also because all epistemic models, event models, and non-reconsideration steps of event update are identical for both $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$.

So, whether intolerance or negation is the fibbee protocol, the same epistemic output will be produced for any event.

So, for every pair of pre-update epistemic models from each of $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ which encode the same epistemic situation, event update with the fibbee protocols intolerance and negation lead to updated epistemic models with the same epistemic outputs.

Then, by the definition of equivalence, $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ are equivalent.

□

5.5 Downsides of "negate"

We've already shown the equivalence of negate with intolerance, so in one fell swoop they have the same downsides.

However, outside the context of two-valued logic, the intuitions that they're fundamentally different may hold. For example, in many-valued logics, considering the negation of a proposition is not equivalent to ignoring the additional consideration of its affirmation when one does not already consider its affirmation.

In most cases this will result in narrowing the worldview, and whether this is epistemically desirable on the side of the receiver will depend on the context. For example, if the liar is informed and making a deliberate lie, then knowledge could be gained by considering the negation of the purported informational content of the event.

On the other hand, if the liar has the intent to deceive but is misinformed about the factual state, then the recipient could do themselves an epistemic misfavor by considering the negation of the potentially true purported informational content of the event. Furthermore, just like with intolerance, negation as a fibbee protocol doesn't permit recovery of eliminated subsets of worldviews.

6 Mutable trust in $\text{DELMT}_{\text{sig}}$

Earlier I argued that intentionality, misinformation, and trust are each necessary for deception. In the preceding chapters I've developed an extension of DEL able to express simple instances of misinformation.

In this chapter I will sketch out an extension of DEL with mechanisms for expressing mutable trust. Then I will remark on some insights thereby gained regarding the relationship between trust and misinformation. By juxtaposing this logic of trust with the logics of misinformation already developed, I argue that systems of mutable trust inherently presuppose an existing system of misinformation, and vice versa. In particular, preliminary corrections for intuitive inadequacies for the proposed system of mutable trust contribute further support for concluding that there is a mutual dependence between misinformation and trust.

6.1 Intuitions

There are several necessary criteria for a logic of trust. Firstly, since the trust is mutable, there will need to be some sort of update function that changes the degree of trust as a function of its past value and some relevant factor(s). Accordingly, there must secondly be some sort of trust function which maps ordered pairs of agents to some measure of the degree of trust that one agent has for the other. This trust function will need to use ordered pairs specifically, since there is no reason to expect one agent's degree of trust for another to be reciprocated to the same degree.

We simplify the treatment of trust-sensitive event update by decomposing trust into multiple components. One part is the aforementioned trust function. Another is a function which transforms the output of the trust function, representing idiosyncratic bias of the agent observing the event. Third and last is a function which assesses that output with a threshold to determine whether to accept the purported informational content of the received event. I choose the sigmoid function for this purpose only because its output is a real number in a range between 0 and 1 inclusive. This is a little more interesting than a discrete function like the step function, and better articulates the modularity of the approach taken for logics of trust here. A more complicated function could easily be substituted for the sigmoid function.

We will also present credibility as properties of agents. This avoids the oddity of reifying credibility as separate from the agents actually assessing each other's credibility. What results is a system where at every event update not only the epistemic states of each agent is updated but also the credibility of each receiving agent with respect to the acting agent.

Finally, we should also explicitly enable iterated event update of epistemic models. Mutable trust is most interesting regarding its mutation not merely over one event but through multiple events over time. Iteration permits observation of the evolution of relationships of trust between agents over the passage of time.

6.2 Dynamic epistemic logic with mutable trust using a sigmoid trust evaluation function ($\mathbf{DELMT_{sig}}$)

6.2.1 Definitions

The language of ***d**ynamic **e**pistemic **l**ogic with **m**utable **t**rust using a **sig**moid trust evaluation function* ($\mathrm{DELMT_{sig}}$) $\mathcal{L}_{\mathrm{DELMT_{sig}}}$ is the set of formulas captured by the following BNF:

$$\phi ::= p|\neg\phi|\phi \wedge \phi| \Box_g \phi|[\langle\mathcal{E}, e\rangle]\phi$$

where $p \in$ ATM and ATM is the set of all atomic propositions (atoms), $g \in$ AGT, AGT is the set of all epistemic agents, and $\langle\mathcal{E}, e\rangle$ is an event model (see 6.2.2).

g is a 3-tuple $\langle rsp_g, A_g, thr_g\rangle$, where

$rsp_g \in \mathbb{R}$ is the *incremental change in suspicion of agent* g,

$A_g : \mathbb{R} \to [0, 1]$ is the *trust evaluation function for agent* g,

where $\forall g \in AGT\ A_g(x) = \dfrac{1}{1 + e^{-x}}$

and

$thr_g \in [0, 1]$ is the *threshold of trust for agent* g.

The shape of the trust evaluation function is well-known as the "sigmoid function":

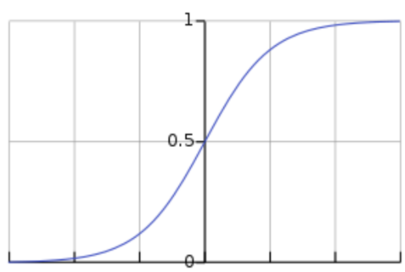


Figure 22: A sigmoid trust evaluation function

An *unpointed epistemic model* in $\text{DELMT}_{\text{sig}}$ $\mathcal{M}$ is defined as a 4-tuple $\langle S, R, V, T \rangle$, where

S is a non-empty set of states,
R is a mapping from agents to sets of pairs of states; i.e. $R : \text{AGT} \to \mathcal{P}(S^2)$, where
  R_g is a set of pairs of states called the *epistemic accessibility relation for agent g*,
V is a mapping from states to a mapping from atoms to a member of the set $\{0, 1\}$ called the *valuation function*; i.e. $V : S \to (\text{ATM} \to \{0, 1\})$, where
  V_s is a mapping from atoms to a member of the set $\{0, 1\}$ called the *valuation for state s*; i.e. $V_s : ATM \to \{0, 1\}$, where
    $V_s(p)$ is a member of the set $\{0, 1\}$ called the *truth assignment for atom p at state s.*
and
T is a mapping from agents to mappings from agents to the real numbers, where
  $T_g : \text{AGT} \to \mathbb{R}$ is the *credibility function for agent* g.

An *epistemic model* in $\text{DELMT}_{\text{sig}}$ is a 2-tuple $\langle \mathcal{M}, s \rangle$ where $s \in S$ is the actual state.

Once again we also first need to define what an *event model* is before we can move on to truth definitions.

An *unpointed event model* in $\text{DELMT}_{\text{sig}}$ $\mathcal{E}$ is defined as a 4-tuple $\langle E, Q, P, C \rangle$, where

E is a non-empty set of *events*
Q is a mapping from agents to sets of pairs of events; i.e. $Q : \text{AGT} \to \mathcal{P}(E^2)$, where
  Q_g is a set of pairs of events called the *event accessibility relation for agent g*,
P is a mapping from events to formulas called the *purported informational content function*; i.e. $P : E \to \mathcal{L}_{\text{DEL}}$, where
  $P(e)$ is a formula called the *purported informational content for event e*
C is a function $C : E \to \text{AGT}$, where
  $C(e)$ is an agent called the *actor for event e.*

An *event model* in $\text{DELMT}_{\text{sig}}$ is a 2-tuple $\langle \mathcal{E}, e_0 \rangle$ where $e_0 \in E$ is the factual event.

Now the truth definitions for $\text{DELMT}_{\text{sig}}$ follow:

$$
\begin{array}{lll}
\vDash_s^{\mathcal{M}} p & \text{iff} & V_s(p) = 1 \\
\vDash_s^{\mathcal{M}} \neg\phi & \text{iff} & \nvDash_s^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \phi \wedge \psi & \text{iff} & \vDash_s^{\mathcal{M}} \phi \text{ and } \vDash_s^{\mathcal{M}} \psi \\
\vDash_s^{\mathcal{M}} \phi \vee \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg(\neg\phi \wedge \neg\psi) \\
\vDash_s^{\mathcal{M}} \phi \rightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} \neg\phi \vee \psi \\
\vDash_s^{\mathcal{M}} \phi \leftrightarrow \psi & \text{iff} & \vDash_s^{\mathcal{M}} (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi) \\
\vDash_s^{\mathcal{M}} \Box_g\phi & \text{iff} & \forall t \in S \text{ if } \langle s,t\rangle \in R_g \text{ then } \vDash_t^{\mathcal{M}} \phi \\
\vDash_s^{\mathcal{M}} \Diamond_g\phi & \text{iff} & \vDash_s^{\mathcal{M}} \neg \Box_g \neg\phi \\
\vDash_s^{\mathcal{M}} [\langle\mathcal{E},e\rangle]\, \Box_g\, \phi & \text{iff} & \vDash_{s[e]}^{\langle\mathcal{M},s\rangle\otimes\langle\mathcal{E},e\rangle} \phi \text{ and } A_g(T_g(C(e)) \geq thr_g \\
& & \text{or } \vDash_s^{\mathcal{M}} \phi \text{ and } A_g(T_g(C(e)) < thr_g
\end{array}
$$

where $\otimes$ is the *event update operator* such that an *updated epistemic model* $\langle\mathcal{M},s\rangle \otimes \langle\mathcal{E},e\rangle$ is a 2-tuple $\langle\mathcal{M}',s[e]\rangle$, where

$\mathcal{M}'$ is a 3-tuple $\langle S',R',V',T'\rangle$ called an *unpointed updated epistemic model*, where

$S' = \{s[e] | s \in S \text{ and } e \in E'\}$,
$R' = \{\langle g,R'_g\rangle | R_g = \{\langle s[e],t[f]\rangle | s[e],t[f] \in S' \text{ and } \langle s,t\rangle \in R'_g \text{ and } \langle e,f\rangle \in Q'_g\}\}$,
$V' = \{\langle s[e],V_s\rangle | s[e] \in S'\}$,
and
$T' : AGT \rightarrow (AGT \rightarrow \mathbb{R})$, where
$T'_g = \{\langle C(e),a\rangle | a \in \mathbb{R}\}$, where

$$
a = \begin{cases} T_g(C(e)) + rsp_g & \text{if } \vDash_{s[e]}^{\mathcal{M}'} \Box_g P(e) \\ T_g(C(e)) - rsp_g & \text{if } \vDash_{s[e]}^{\mathcal{M}'} \Box_g \neg P(e) \\ T_g(C(e)) & \text{otherwise} \end{cases}
$$

Notes: An updated epistemic model $\langle\mathcal{M},s\rangle \otimes \langle\mathcal{E},e\rangle$ has factual state $s[e] \in S'$.
An updated epistemic model is a special case of epistemic model.

6.3 The interdependence of DELMIs and DELMTs

In this section I argue that the development of any modal epistemic logic accommodating mutable trust must also accommodate misinformation handling. The proceeding discussion reveals that DEL and therefore $\text{DELMI}_{\text{int}}$ and $\text{DELMI}_{\text{neg}}$ implicitly accommodated trust, albeit static rather than mutable.

This discussion delivers on my promise that a more generalized encoding of misinformation using DEL leads us to philosophically interesting insights regarding the nature of deception. I now draw general conclusions about deeper relationships between misinformation and trust in the context of deception.

6.3.1 DELMTs assume fibbee protocols

Let's consider what happens in $\text{DELMT}_{\text{sig}}$ if a trusted actor produces an event that leads a recipient agent to seriously consider a contradiction. The receiver will understand that one of her epistemic sources, whether presently or at some time in the past, was a carrier of misinformation.[17] Then the receiver will consult their previous worldview to fill out the rest of the update, that previous worldview being as per the initial epistemic model.

This is so far accommodated by $\text{DELMT}_{\text{sig}}$, but $\text{DELMT}_{\text{sig}}$ as defined only fills out the resulting updated epistemic model as basic DEL does. While the credibility function is updated as a result of what happened (the receiver deducts from her reading of the actor's trustworthiness accordingly), $\text{DELMT}_{\text{sig}}$'s recommendation of reconsideration is altogether stronger than $\text{DELMI}_{\text{int}}$'s, because complete reversion of the worldview means "higher-order social insight" is also forgone as a result of the misinformation handling.[1] In other words, in $\text{DELMI}_{\text{int}}$ a privy agent will understand that other agents may not be aware that an act of misinformation has occurred, and this will be reflected in their consideration base regarding their consideration of other agents' considerations. On the other hand, this does not happen in $\text{DELMT}_{\text{sig}}$ because privy agents revert wholesale to the epistemic model before event update.

Even so, misinformation handling is necessary in $\text{DELMT}_{\text{sig}}$. This is because having a mechanism for mutable trust in a logic which language does not trivialize it just is to implement handling for receiving information from potentially misinformative events. If $\text{DELMT}_{\text{sig}}$ is to satisfy basic intuitions about mutable trust, a fibbee protocol must be implemented, explicitly or implicitly.

Is this is peculiar to $\text{DELMT}_{\text{sig}}$, or do all modal epistemic logics of mutable trust require fibbee protocols? I now argue that the answer to this question is in the affirmative.

6.3.2 DELMIs assume trust protocols

$\text{DELMT}_{\text{sig}}$ implicitly has a fibbee protocol. As discussed previously, this is because the nontrivial implementation of mutable trust implies the existence of such a situation where a

[17] There's also the possibility that both epistemic sources are mistaken, but this would be undetectable in $\text{DELMT}_{\text{sig}}$ because it is two-valued.

receiving agent, after update with the purported informational content from an event from a trusted acting agent, has contradictory considerations in her consideration set. Thus any logic of trust must at least implicitly advertise a particular fibbee protocol.

We compare this with DEL or $\text{DELMI}_{\text{int}}$, both in which agents make event updates to their worldviews independently of the event's actor. In DEL, agents always accept the precondition of events as true. In $\text{DELMI}_{\text{int}}$, agents always provisionally accept the purported informational content of events as true, then ignore (or reject) them if in doing so they arrive at an internally inconsistent consideration set. As such $\text{DELMI}_{\text{int}}$ implicitly has a mechanism for trust even if that "mechanism" is simply naive unconditional trust.

$\text{DELMT}_{\text{sig}}$ presents agents as making event updates dependent on their relationship to the actor. Recall that this relationship is captured in the trust evaluation function. While in $\text{DELMI}_{\text{int}}$ there are only two possible outcomes of event update—accept or ignore—in $\text{DELMT}_{\text{sig}}$ there are instead four:

1. A trusted agent acts consistently with the receiver's consideration base.
2. A trusted agent acts inconsistently with the receiver's consideration base.
3. A non-trusted agent acts consistently with the receiver's consideration base.
4. A non-trusted agent acts inconsistently with the receiver's consideration base.

We recognize that $\text{DELMI}_{\text{int}}$ only expresses outcomes (1) and (2), and this is because it inherits from DEL the presumption that actors are to be trusted. Notice that we could make $\text{DELMT}_{\text{sig}}$ behave identically as $\text{DELMI}_{\text{int}}$ by simply setting all starting T values to 1 and all agents' incremental changes in suspicion to 0. These settings will make it respectively such that all agents completely trust each other by default and that inter-agent trust levels remain static regardless of the epistemic outcomes of events over time.

Looking at this from the other direction, $\text{DELMI}_{\text{int}}$'s "trust protocol" of unconditional trust is the special case of $\text{DELMT}_{\text{sig}}$'s mutable trust where all agents are trusted and the output of the trust evaluation function never changes from 1. Then, $\text{DELMI}_{\text{int}}$'s reconsideration step in event update corresponds to consideration base reversion in $\text{DELMT}_{\text{sig}}$ when the trust evaluation function returns an output below that of the relevant threshold of trust.

Altogether this demonstrates the converse that any logic accommodating misinformation contain at least some sort of rudimentary trust protocol.

Altogether we could say that implementation of misinformation handling and mutable trust in modal epistemic logic are therefore interdependent rather than independent.

7 Conclusion

I have assessed characteristics of misinformation and trust, and then formalized those characteristics in novel extensions of dynamic epistemic logic (DEL). I positioned my investigation in the context of an analysis of deception, through which I argued that intentionality, misinformation, and trust are three necessary components of deception. Formalizing misinformation and trust in DEL was accomplished through implementation of additional mechanisms for misinformation handling and mutable trust.

To formalize misinformation I developed $\mathrm{DELMI}_{\mathrm{int}}$, dynamic epistemic logic of misinformation with fibbee protocol “intolerance”. I also developed the similar $\mathrm{DELMI}_{\mathrm{neg}}$, dynamic epistemic logic of misinformation with fibbee protocol “negation”, and proved its equivalence with $\mathrm{DELMI}_{\mathrm{int}}$ in the two-valued context. The modularity of $\mathrm{DELMI}_{\mathrm{int}}$’s fibbee protocol mechanism holds promise for accommodating other fibbee protocols in future investigations.

To formalize trust I developed $\mathrm{DELMT}_{\mathrm{sig}}$, dynamic iterated epistemic logic of trust with sigmoid trust evaluation. The modularity of the trust evaluation function similarly holds promise for future work.

Then I evaluated $\mathrm{DELMT}_{\mathrm{sig}}$ and $\mathrm{DELMI}_{\mathrm{int}}$ concerning the degree that they capture intuitions about interactions between misinformation and trust. I concluded that the formalization of misinformation and trust in modal epistemic logic is mutual to a degree.

References

[1] Bolander, Thomas. (2014). "Seeing is believing: Formalising false-belief tasks in dynamic epistemic logic". *Proceedings of the First European Conference on Social Intelligence*. Barcelona, Spain. 87-107.

[2] Blackburn, Patrick; de Rijke, Maarten; Venema, Yde. (2001). *Modal Logic*. New York, U.S.A.: Cambridge University Press.

[3] Demey, Lorenz; Kooi, Barteld. (2014). "Logic and Probabilistic Update". In A. Baltag & S. Smets (eds.), *Johan van Benthem on Logic and Information Dynamics*, Outstanding Contributions to Logic 5. Switzerland: Springer International Publishing. 381-404.

[4] Garson, James W. (2013). *Modal Logic for Philosophers* (2nd ed.). New York, U.S.A.: Cambridge University Press.

[5] van Ditmarsch, Hans. (2014). "Dynamics of lying". *Synthese* **191**(5). 745-777.

[6] van Ditmarsch, Hans; van der Hoek, Wiebe; Kooi, Barteld. (2008). *Dynamic Epistemic Logic*. Dordrecht, Netherlands: Springer.

[7] Hintikka, Jaakko. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. New York, U.S.A.: Cornell University Press.

[8] Kooi, Barteld P. (2003). "Probabilistic Dynamic Epistemic Logic". *Journal of Logic, Language and Information* **12**. 381-408.

[9] Kripke, Saul. (1959). "A completeness theorem in modal logic". *The Journal of Symbolic Logic* **24**(1). 1-14.

[10] Liu, Fenrong. (2006). "Diversity of agents". In T. Agotnes, & N. Alechina (eds.), *Proceedings of the workshop on resource-bounded agents*. Malaga, Spain: European Summer School on Logic, Language and Information. 88-98.

[11] Plaza, Jan A. (1989). "Logics of public communications". In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, & Z.W. Ras (eds.), *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*. Torino, Italy. 201-216.

[12] Sakama, Chiaki. (2015). "A Formal Account of Deception". *Proceedings from the Association for the Advancement of Artificial Intelligence 2015 Fall Deceptive and Counter-Deceptive Machines Symposium*. Arlington County, Virginia, U.S.A. 34-41.

[13] von Wright, Georg H. (1953). *An Essay in Modal Logic*. Amsterdam: North-Holland Publishing Company.