The Surprise Deception Paradox

Benjamin Icard

The Surprise Deception Paradox

Benjamin Icard
LIP6, Sorbonne Université, CNRS, France

Abstract

This article tackles an epistemic puzzle formulated by R. Smullyan that we call the *'Surprise Deception Paradox'*. On the morning of April 1^{st} 1925, his brother announced that he would deceive him during the day, but apparently nothing happened. Since R. Smullyan waited all day to be deceived by some action, he was actually deceived, but by the lack of an action, that is to say by omission. Afterwards, Smullyan felt immediately puzzled: because he expected to be deceived, he was not deceived; but since he was not deceived the way he expected, he was actually deceived. We use dynamic belief revision logic to look more clearly into this puzzle. We argue that Smullyan's reasoning is not a self-referential paradox but shares common features with the more famous Surprise Examination Paradox. In Smullyan's riddle, we show that a misleading default mechanism makes R. Smullyan surprised by the deception he has been preyed to. We also use this solution to discuss whether such defaults, compared to other forms of truth-telling deception, may qualify as lies or not.

Keywords — Deception, Omission, Default Inference, Dynamic Belief Revision Logic, Doxastic Paradox, Surprise, Veridical Deception, Lying

1 Introduction

In *What is the Name of this Book?*, Raymond Smullyan tells a personal anecdote concerning his first introduction to logic [49]. While he was suffering from flu on April 1^{st} 1925, his older brother — Emile —, said to him: "*Well,* Raymond*, today is April Fool's Day, and I will deceive you as you have never been deceived before!*". After this announcement, Raymond waited all day long to be deceived but (apparently) nothing happened. Late at night, though he was no longer expecting to be deceived (or so he thought), he still felt highly concerned by his brother's announcement. Noticing his perplexity, Emile asks him in a Socratic way:

Emile: "*So, you expected me to deceive you, didn't you?*
Raymond: *Yes.*
Emile: *But I didn't, did I?*
Raymond: *No.*
Emile: *But you expected me to, didn't you?*
Raymond: *Yes.*
Emile: *So I deceived you, didn't I?*
Raymond: *Yes.*"

After this explanation, Raymond lays in his bed and starts reasoning. He rapidly falls into puzzling thoughts. On the one hand, supposing that he wasn't deceived, then he did not get what he expected, because he expected to be deceived after his brother's

announcement, and hence he was actually deceived. But on the other hand, supposing that he was actually deceived, then he exactly did get what he expected, and hence he was not *deceived* stricto sensu.

Emile's trick may be decomposed into three distinct stages: the *ex ante*, *ex interim* and *ex post* stages. At the *ex ante stage*, Emile announces that he will deceive Raymond later on the day. At this moment, Raymond is not yet deceived but he learns he will be. At the *ex interim stage*, deception actually occurs: Raymond believes that Emile will deceive him by doing some specific action (commission) but in fact, he deceives him by doing none (omission). At the *ex post stage*, Raymond is no longer deceived but Emile's later explanation makes him puzzled and surprised.

In this paper, we analyze the dynamics of Emile's deception and show that this latter paradox is specious. However, by comparing Emile's announcement with the teacher's announcement in the surprise examination paradox [e.g. 41, 47, 48], we argue that Emile's announcement should not succeed in principle. His deception succeeds, and Raymond feels surprised, only because Emile's announcement leads Raymond to draw the misleading default interpretation that he will be deceived only by commission. Leaving this mechanism aside, Emile's announcement should theoretically fail, as in Gerbrandy's formalizations of the three-day version of the surprise exam paradox [29, 5]. Raymond is surprised at the end of the day because he did not expect at all that he could be deceived both by commission and by omission. This is a source of strong surprise.

In Section 2, we adapt extant dynamic belief revision logic [6, 11] to analyze Smullyan's story in details. Our language $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ is a propositional doxastic syntax enriched with a belief upgrade operator taken from [6, 7]. This upgrade operator is used to provide abbreviations for two defined operators inspired by the quantification over actions in arbitrary announcement logic [26, 33]. Those defined operators aim to express Raymond's mistaken *default interpretation* of Emile's announcement as meaning that he will be deceived by some action (*deception by commission*), when in fact he will be deceived by none (*deception by omission*) (subsection 2.1). This notion of default interpretation motivates the use of doxastic plausibility models to provide $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ with semantic clauses (subsection 2.2). That being done, we rely on this apparatus to define various notions of deception involved in Smullyan's story, such as deception as a state versus as an attitude, deception on a content versus simpliciter, etc. (subsection 2.3). Finally, $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ helps to set out several facts of Smullyan's tale before analyzing them in more details (subsection 2.4).

In Section 3, we use $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ to describe the stages of the April Fool's deception. We distinguish Emile's *ex ante announcement* in which Emile informs Raymond that he will deceive him during the day (subsection 3.1), from the *ex interim deception* coming later on that day (subsection 3.2). In this case, we focus on the two distinct states of deception Raymond goes through successively (subsection 3.2.1) and then express Emile's two distinct attitudes leading to those states (subsection 3.2.2). Finally, we model Emile's *ex post explanation* that describe his whole deceptive plot (subsection 3.3).

Section 4 delves deeper into the *ex post* stage of deception by analyzing Raymond's puzzling thoughts and surprise after Emile's explanation. We first point out that Raymond's reasoning is self-referential but not paradoxical as it stands (subsection 4.1). Then, we compare Emile's announcement with the teacher's announcement in the Surprise Examination Paradox to argue that Emile's announcement of a surprising deception should fail in principle (subsection 4.2). Basically, we rely on Gerbrandy's formal-

ization of surprise [29] to provide a radical upgrade reading of EMILE's announcement in line with the three-day version of the surprise exam paradox (subsection 4.2.1). Using plausibility models as in [5], we argue that EMILE's announcement does succeed only as a result of RAYMOND being deceived by misleading default interpretation (subsection 4.2.2). Finally, we argue that RAYMOND is in fact surprised in two different ways by EMILE on April 1^{st} 1925 (subsection 4.2.3).

We conclude in Section 5 by showing that Smullyan's riddle is informative on the way one can deceive by saying the truth, as in so-called "*veridical deception*". Classically, lying is considered the standard form of veridical deception since, traditionnally, a speaker can lie by saying the truth, provided that the speaker is insincere when making her utterance. We discuss EMILE's announcement in light of this standard definition of lying and by taking into account debates related to other forms of veridical deception, such as false implicatures [e.g. 23, 1, 39].

2 A Language for Analysis

2.1 A Dynamic Doxastic Syntax $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$

We present $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ which relies on extant dynamic belief revision theory [e.g. 11, 6, 7]. The syntax of $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ contains a static modal operator for (conditional) belief, as well as a dynamic modal operator for upgrading beliefs. The set of all $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$-*formulas* is given by the following Backus-Naur Form:

$$\langle \textbf{\textit{Formulas}} \rangle \quad \varphi, \psi \quad ::= \quad \top \mid p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \boldsymbol{B}^{\psi}\varphi \mid [\Uparrow \varphi]\psi,$$

Atomic formulas are abbreviated by p. Complex formulas, such as φ or ψ, can be either tautologies (using the common abbreviation $\top$), atomic formulas, negation of a formula, conjunction of formulas, static formulas $\boldsymbol{B}^{\psi}\varphi$ (for beliefs in φ conditional on ψ), or dynamic formulas of the form $[\Uparrow \varphi]\psi$ (for radical upgrade with a formula φ leading to ψ). Additional connectives ($\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$) are defined as usual by using the primitive symbols of language $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$. As long as this does not introduce ambiguity, we now omit parentheses when writing formulas for easier reading.

In the specific context of Smullyan's story, the informal reading of operator $\boldsymbol{B}^{\psi}\varphi$ is more precisely *"*RAYMOND *believes that φ conditional on formula ψ"*,[1] while the reading of the dynamic operator $[\Uparrow \varphi]\psi$ is: *"after* RAYMOND *commits a radical upgrade with φ, ψ is the case"*. Operator $\boldsymbol{B}$ stands for *"*RAYMOND *plainly believes that φ"* and can be defined from $\boldsymbol{B}^{\top}\varphi$ corresponding to *"*RAYMOND *conditionally believes that φ on $\top$"*: $\boldsymbol{B}\varphi := \boldsymbol{B}^{\top}\varphi$. The connection between Smullyan's story and the complex formulas of $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ is made more precise in sections 3 and 4.

2.2 Doxastic Plausibility Models for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$

As indicated previously, we choose to give a semantic interpretation to $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ by using doxastic plausibility models defined by [e.g. 11, 6, 7, 12], instead of more classical Kripke models used in public announcement logics for instance [30, 4, 43]. This choice is motivated by the fact that, after hearing EMILE's announcement of deception on April 1^{st} 1925, RAYMOND performs the *default interpretation* that he will be deceived by some action, as in so-called *deception by commission*, instead of no action, as *deception by*

[1] Following [9], we can think of conditional beliefs $\boldsymbol{B}^{\psi}\varphi$ as 'contingency' plans for belief change: in case RAYMOND will find out that ψ was the case, he will believe that φ was the case.

omission. From a more qualitative perspective, this default interpretation corresponds to the *most plausible way* people would interpret EMILE's early announcement: *"Well,* RAYMOND*, today is April Fool's Day, and I will deceive you as you have never been deceived before!"*. In order to capture this notion of default interpretation based on high plausibility, we use doxastic plausibility models to interpret language $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ semantically.

A plausibility model for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ is a relational structure $\mathbb{S} = \langle \mathcal{S}, \leqslant, \| \, . \, \| \rangle$ with:

- $\mathcal{S}$ which is a finite non-empty set of "*possible states*" $\boldsymbol{s}$ (or "*worlds*").
- $\leqslant$ which is a "*plausibility relation*" for RAYMOND such that $\leqslant \subseteq \mathcal{S} \times \mathcal{S}$. This relation $\leqslant$ is a total preorder, that is to say a transitive, strongly connected and thus reflexive relation over the set $\mathcal{S}$.
- $\| \, . \, \|$: $At \to \wp(\mathcal{S})$ which is a standard "*valuation map*" where At is the set of all atomic formulas p, and $\wp(\mathcal{S})$ is the set of subsets of $\mathcal{S}$.

The conventional reading of the plausibility order is the following one: when $\boldsymbol{s} \leqslant \boldsymbol{t}$ (for all $\boldsymbol{s}$, $\boldsymbol{t} \in \mathcal{S}$), it means that RAYMOND considers the state $\boldsymbol{t}$ to be *"at least as plausible as"* the state $\boldsymbol{s}$. The truth of a formula φ at the actual state $\boldsymbol{s}$ in the doxastic plausibility model $\mathbb{S}$, denoted $\mathbb{S}, \boldsymbol{s} \models \varphi$, is defined inductively by extending the valuation map to all the complex formulas of $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$:

- $\mathbb{S}, \boldsymbol{s} \models \top$ *Always.*
- $\mathbb{S}, \boldsymbol{s} \models p$ *iff* $\boldsymbol{s} \in \| p \|$.
- $\mathbb{S}, \boldsymbol{s} \models \neg\varphi$ *iff* $\mathbb{S}, \boldsymbol{s} \not\models \varphi$.
- $\mathbb{S}, \boldsymbol{s} \models (\varphi \wedge \psi)$ *iff* $\mathbb{S}, \boldsymbol{s} \models \varphi$ and $\mathbb{S}, \boldsymbol{s} \models \psi$.
- $\mathbb{S}, \boldsymbol{s} \models \boldsymbol{B}^{\psi}\varphi$ *iff* for all $\boldsymbol{t} \in Max_{\leqslant}^{\|\psi\|}$: $\mathbb{S}, \boldsymbol{t} \models \varphi$.

Here $Max_{\leqslant}^{\|\psi\|}$ is the set of states that satisfy ψ and that, among all ψ-states, are maximal for the ordering $\leqslant$: $Max_{\leqslant}^{\|\psi\|} = \{\boldsymbol{u} \in \| \psi \| \mid \text{for all } \boldsymbol{v} \in \| \psi \| \; \boldsymbol{v} \leqslant \boldsymbol{u}\}$, where $\| \psi \|$ is the extension of $\| \, . \, \|$ to arbitrary formulas $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, such that, given $\mathbb{S}$: $\| \psi \| = \{\boldsymbol{t} \in \mathcal{S} \mid \mathbb{S}, \boldsymbol{t} \models \psi\}$. In other words, in the semantic clause corresponding to $\boldsymbol{B}^{\psi}\varphi$, the ψ-states on which RAYMOND conditions his beliefs in φ are the ψ-states of the ordering that RAYMOND considers to be the most plausible ones *and* that also satisfy φ. As we said, plain beliefs are recovered from conditional beliefs by conditioning on $\top$, and the clause can be reduced to:

- $\mathbb{S}, \boldsymbol{s} \models \boldsymbol{B}\varphi$ *iff* for all $\boldsymbol{t} \in Max_{\leqslant}^{\|\top\|}$: $\mathbb{S}, \boldsymbol{t} \models \varphi$.

Following [11] and [7], we define a belief radical upgrade with a formula φ, written $[\Uparrow \varphi]$, as a mapping of the following kind:

$$[\Uparrow \varphi]\colon \mathbb{S} \mapsto \mathbb{S}^{[\Uparrow\varphi]}$$

Here $\mathbb{S}$ is the initial plausibility model and $\mathbb{S}^{[\Uparrow\varphi]}$ is the transformed model obtained once the intended operation $[\Uparrow \varphi]$ has been performed on $\mathbb{S}$. The semantic clause for belief radical upgrade is defined by:

- $\mathbb{S}, \boldsymbol{s} \models [\Uparrow \varphi]\psi$ *iff* $\mathbb{S}^{[\Uparrow\varphi]}, \boldsymbol{s} \models \psi$.

The doxastic plausibility model $\mathbb{S}^{[\Uparrow\varphi]}$ is defined from $\mathbb{S}$ in the following way:

$$\mathbb{S}^{[\Uparrow\varphi]} = \langle \, \mathcal{S}^{[\Uparrow\varphi]}, \leqslant^{[\Uparrow\varphi]}, \| \, . \, \|^{[\Uparrow\varphi]} \, \rangle$$

where:

$$\begin{array}{lcl} \mathcal{S}^{[\Uparrow\varphi]} & := & \mathcal{S} \\ \leqslant^{[\Uparrow\varphi]} & := & (\leqslant \cap(\mathcal{S}\times \parallel \varphi \parallel))\cup(\leqslant \cap(\parallel \neg\varphi \parallel \times\mathcal{S})) \cup (\parallel \neg\varphi \parallel \times \parallel \varphi \parallel) \\ \parallel . \parallel^{[\Uparrow\varphi]} & := & \parallel . \parallel \end{array}$$

Here $\mathcal{S}^{[\Uparrow\varphi]}$ is identical to $\mathcal{S}$ from the initial model $\mathbb{S}$, while on the atomic fragment of $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, $\parallel . \parallel^{[\Uparrow\varphi]}$ is similar to $\parallel . \parallel$. The special feature of $\mathbb{S}^{[\Uparrow\varphi]}$ is the plausibility order $\leqslant^{[\Uparrow\varphi]}$. The reordering of states defined by $\leqslant^{[\Uparrow\varphi]}$ ensures that the states where φ is true are promoted in plausibility. More precisely, the first part $(\leqslant \cap(\mathcal{S}\times \parallel \varphi \parallel))$ of $\leqslant^{[\Uparrow\varphi]}$ states that the relative ordering of worlds where φ is true is the same as in the original order $\leqslant$. The second part $(\leqslant \cap(\parallel \neg\varphi \parallel \times\mathcal{S}))$ states that the relative ordering of worlds where φ is false is the same as in the original order $\leqslant$. Finally, the third part $(\parallel \neg\varphi \parallel \times \parallel \varphi \parallel)$ states that the worlds where φ is true become equally or more plausible than the worlds where φ is false. So, in the upgraded model $\mathbb{S}^{[\Uparrow\varphi]}$, although all φ-states are made *equally or more plausible than* all $\neg\varphi$-states, everything else is left the same *within* the φ and $\neg\varphi$ zones themselves.

For clarity's sake, let us put more emphasis on the dynamic attitude corresponding to operator $[\Uparrow \varphi]\psi$. The upgrade part of this operator, that is to say $[\Uparrow \varphi]$, indicates that the agent does not know the source of φ to be *absolutely reliable*, thus *completely honest* and *truthful*. But the agent believes the source to be *highly reliable*, thus *strongly honest* and *truthful*.[2] The whole complex formula $[\Uparrow \varphi]\psi$ according to which ψ is the case after the belief upgrade $[\Uparrow \varphi]$, can be seen as a *default rule of interpretation* based on the intuition that ψ is *more plausible* than $\neg\psi$ in all the cases in which φ obtains. This notion of *higher plausibility* which is the core intuition governing non-monotonic reasoning has been captured in [51] by a formal rule written $(\varphi \Rightarrow \psi)$ meaning that formulas "φ *are normally* ψ" or, to put it another way, that worlds satisfying the conjunction $(\varphi \wedge \psi)$ are *more plausible* than worlds satisfying the conjunction $(\varphi \wedge \neg\psi)$. For this reason, we assume that the default rule $(\varphi \Rightarrow \psi)$ can be reasonably approximated by the formula $[\Uparrow \varphi]\boldsymbol{B}\psi$ in language $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$.

Various other dynamic revision operators exist in the literature [e.g. 11, 7, 12], based on previous work on belief revision policies [e.g. 45, 27]. Within all this diversity, two well-known representatives are hard updates such as e.g. public announcements, usually written "!", in which sources are considered completely truthful and receive absolute trust [30, 4, 43], or softer upgrades such as e.g. "conservative upgrades", usually written "↑", in which sources are taken with much weaker confidence as only barely truthful [7, 12]. Those policies can also be read as types of default interpretations but regarding the trust Raymond should put into Emile on April 1^{st} 1925, radical upgrades are the most appropriate match: high confidence, though not perfect confidence. From a theoretical perspective, Raymond cannot have perfect confidence in Emile: the day is April Fool's Day and Emile explicitly states that he will deceive him during that day. On a more practical note, Raymond's behaviour all along the day shows that he has, and keeps having, strong confidence in Emile, — even at the end of the day when nothing has happened and even after his brother's late explanation. So, weaker confidence operators,

[2] According to van Benthem's terminology, the agent does not have a *hard* doxastic attitude towards the source of information, but a *softer* one. The agent is not sure that the source always tells the truth but she is *highly confident* that the source does.

such as conservative upgrades, would not match adequately with RAYMOND's effective behaviour.

As proved in [7] and [11], $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ is axiomatized completely by a set of axioms and inferences rules for propositional logic, belief operator $\boldsymbol{B}$ and the dynamic upgrade operator $\Uparrow$. Here we simply put emphasis on two axioms of central importance for analyzing RAYMOND's paradoxical thoughts, in addition to other more classical principles (see subsection 4.1 in that respect). For $\varphi,\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, the two reduction axioms we use are:

$$\begin{array}{lllll} (i) & \vdash & [\Uparrow \varphi]\neg\psi & \leftrightarrow & \neg[\Uparrow \varphi]\psi \\ (ii) & \vdash & [\Uparrow \varphi](\psi \wedge \phi) & \leftrightarrow & ([\Uparrow \varphi]\psi \wedge [\Uparrow \varphi]\phi) \end{array}$$

2.3 Expressing Deception with $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$

We first define the distinction between *action of commission* leading to a formula ψ versus *omission of an action* leading to ψ to draw further distinctions regarding RAYMOND's deception on April 1^{st} 1925.[3] Broadly speaking, we distinguish the fact for RAYMOND to be in states of deception on April 1^{st} 1925 from EMILE's deceptive attitudes leading to RAYMOND to fall in such states.

2.3.1 Commission versus Omission of an Action

Definition 1 (*commission such that* ψ) We let $[\Uparrow^{+}]\psi$ stands for the action of commission leading to formula ψ on April 1^{st} 1925. Given some $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, this corresponds to the following abbreviation:

$$[\Uparrow^{+}]\psi \quad := \quad [\Uparrow \varphi]\psi \text{ for some } \varphi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

The abbreviation $[\Uparrow^{+}]\psi$ provided in Definition 1 states that formula ψ obtains as the result of a radical upgrade with some formula φ. In the context of Smullyan's story, this abbreviation $[\Uparrow^{+}]\psi$ can be read as *"after* RAYMOND *commits some radical upgrade, ψ is the case"*.

Definition 2 (*omission such that* ψ) We let $[\Uparrow^{-}]\psi$ stands for the omission of an action leading to formula ψ on April 1^{st} 1925. Given some $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, this corresponds to the abbreviation:

$$[\Uparrow^{-}]\psi \quad := \quad [\Uparrow \varphi]\psi \text{ for no } \varphi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

By contrast with Definition 1, abbreviation of Definition 2 states that formula ψ now obtains as the result of the lack of any radical upgrade with some φ. In the context of Smullyan's story, the abbreviation $[\Uparrow^{-}]\psi$ can be read as *"after* RAYMOND *omits to perform any radical upgrade, ψ is the case"*.

Observation 1 (*commission vs omission such that* ψ) Notice that given a formula $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, a doxastic plausibility model $\mathbb{S} = \langle \mathcal{S}, \leqslant, \| \cdot \| \rangle$ for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ with a state $\boldsymbol{s} \in \mathcal{S}$, we have $\mathbb{S}, \boldsymbol{s} \models [\Uparrow^{-}]\psi$ *iff* $\mathbb{S}, \boldsymbol{s} \not\models [\Uparrow^{+}]\psi$. Indeed, suppose that we have $\mathbb{S}, \boldsymbol{s} \models [\Uparrow^{-}]\psi$. So, by Definition 2, we have $\mathbb{S}, \boldsymbol{s} \models [\Uparrow^{-}]\psi$ *iff* $\mathbb{S}, \boldsymbol{s} \models [\Uparrow \varphi]\psi$ for no $\varphi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ *iff* $\mathbb{S}, \boldsymbol{s} \not\models [\Uparrow \varphi]\psi$ for all $\varphi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, so that we have in particular $\mathbb{S}, \boldsymbol{s} \not\models [\Uparrow \varphi]\psi$ for some $\varphi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, i.e. $\mathbb{S}, \boldsymbol{s} \not\models [\Uparrow^{+}]\psi$ by Definition 1.

[3] Notice that the nature of omission regarding causality [e.g. 46, 18] raises deontological issues linked to the moral responsibility of agents [19, 54] in computational ethics and related logics [15, 16, 21]. These questions of crucial importance go beyond the scope of this article which is more concerned with modelling attitudes of deception based on commission versus omission and epistemic paradoxes involved in Smullyan's recollection.

Consistent with conceptual intuitions, Observation 1 shows that the omission of an action resulting in ψ amounts to the *absence*, or *lack*, of an action of commission resulting in ψ.

2.3.2 Deception as a State

We start out by defining what it means for RAYMOND to be in a doxastic state of deception, first concerning a specific formula ψ and then simpliciter.

Definition 3 (*deceived on* ψ) We write $\texttt{deceived}_\psi$ the fact for RAYMOND to be deceived on some formula ψ on April 1^{st} 1925. The corresponding abbreviation is:

$$\texttt{deceived}_\psi \quad := \quad (\psi \wedge \boldsymbol{B}\neg\psi) \text{ for some } \psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

Outside of Smullyan's story, Definition 3 would be considered inadequate for the reason that it does not express the presence of the deceiver and sets aside his or her intentions to deceive the addressee. But firstly, the truly central dimensions of EMILE's deceptive plan are the dimensions of commission and omission, not EMILE's intention to deceive RAYMOND, which is naturally assumed. Secondly, EMILE's intention to deceive does not play any particular role in the doxastic paradoxes involved in Smullyan's story (see subsections 4.1 and 4.2). For those reasons, we assume that EMILE has the intention to deceive RAYMOND, but we do not formalize it explicitly in $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ or by introducing further definitions.

Definition 4 (*deceived simpliciter*) We write $\texttt{deceived}$ the fact for RAYMOND to be deceived simpliciter on April 1^{st} 1925. The corresponding abbreviation is:

$$\texttt{deceived} \quad := \quad \texttt{deceived}_\psi \text{ for some } \psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

Definition 4 simply states that being deceived on some formula $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ is sufficient for RAYMOND to be deceived *stricto sensu* by EMILE on April 1^{st} 1925.

2.3.3 Deception as an Attitude

After having defined deception as a state, we define deception as an attitude here. For capturing the subtleties of EMILE's deceptive trick, we need to be able to capture the distinction between deception by an action, i.e. *deception by commission*, versus deception by the lack of an action, i.e. *deception by omission*.

Definition 5 (*deception by commission on* ψ) We abbreviate by d^+_ψ the action of commission leading RAYMOND to be deceived on formula ψ. That is:

$$d^+_\psi \quad := \quad [\Uparrow^+]\texttt{deceived}_\psi \text{ for some } \psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

Based on Definition 3, Definition 5 states that RAYMOND is deceived by commission on some formula $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ in case there is some action such that after this action, RAYMOND is deceived on ψ, since ψ is true but RAYMOND believes ψ to be false.

Definition 6 (*deception by omission on* ψ) We use d^-_ψ to abbreviate the omission of an action leading RAYMOND to be deceived on formula ψ. That is:

$$d^-_\psi \quad := \quad [\Uparrow^-]\texttt{deceived}_\psi \text{ for some } \psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

Again, based on Definition 3, Definition 6 states that RAYMOND is deceived by omission on some formula $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ in case there is no action such that, as the result of this lack of action, RAYMOND is deceived on ψ, since ψ is true but RAYMOND believes ψ to be false.

Observation 2 (*deception by commission vs omission such that* ψ) Following Observation 2, given a formula $\psi \in \mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, a doxastic plausibility model $\mathbb{S} = \langle \mathcal{S}, \leqslant, \| \cdot \| \rangle$

for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ with a state $\boldsymbol{s}\in\mathcal{S}$, we have $\$, \boldsymbol{s} \models d^-_\psi$ *iff* $\$, \boldsymbol{s} \not\models d^+_\psi$. Again, Observation 2 is consistent with conceptual intuitions since the fact that Emile will deceive Raymond by omission on a formula ψ on April 1^{st} 1925 implies that he will not deceive him by commission on ψ, and the other way around.

Definition 7 (*deception on* ψ) We abbreviate by d_ψ the fact that Emile deceive Raymond on formula $\psi \in\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ on April 1^{st} 1925, either by commission or by omission.

$$d_\psi \quad := \quad (d^+_\psi \vee d^-_\psi) \text{ for some } \psi \in\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$$

As emphasized in Observation 2, d^+_ψ and d^-_ψ cannot be true together for a given formula ψ since d^+_ψ is semantically equivalent to $(d^+_\psi \wedge \neg d^-_\psi)$, while d^-_ψ is semantically equivalent to $(\neg d^+_\psi \wedge d^-_\psi)$. Naturally, this semantic incompatibility does not impose any syntactic restriction in Definition 7.

Definition 8 (*deception simpliciter, versions of*) Let us use abbreviations $\mathtt{d}^+$, $\mathtt{d}^-$ and $\mathtt{d}$ for attitudes of deception by commission simpliciter, deception by omission simpliciter and deception simpliciter. That is:

- Deception by commission simpliciter: $\mathtt{d}^+ := \mathtt{d}^+_\psi$ for some $\psi \in\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$
- Deception by omission simpliciter: $\mathtt{d}^- := \mathtt{d}^-_\psi$ for some $\psi \in\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$
- Deception simpliciter: $\mathtt{d} := \mathtt{d}_\psi$ for some $\psi \in\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$

In definition 8, formula $\mathtt{d}$ means that, in case Emile actually deceive Raymond, he deceives him either by commission ($\mathtt{d}^+$) or by omission ($\mathtt{d}^-$). Conceptually, formula $\mathtt{d}^+$ states that Emile will deceive Raymond by *doing some specific action* on April 1^{st} 1925. That is: Emile will make *some move* such that Raymond will be *deceived on some fact* afterwards. Such a strategy is known as *"deception by commission"* in the literature [17]. By contrast, formula $\mathtt{d}^-$ states that Emile will deceive Raymond by *not making any action* on April 1^{st} 1925, that is: by the absence of an action. Now Emile won't perform any action but, in doing so, he will deceive his brother by depriving him of some informational facts. This strategy is also called *"deception by omission"* in the literature [17].

2.4 Contextualizing $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$ to Smullyan's story

Formulas $\mathtt{d}^+$ and $\mathtt{d}^-$ should be seen as *types* of deceptive attitudes. Those formulas express the two kinds of deceptive means Emile can use to deceive his brother after his morning announcement. Let us abbreviate by formula D Emile's morning announcement in $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$, more precisely the formula capturing its content:

Definition 9 (*informal reading of* Emile*'s announcement*) We write D the content of Emile's morning announcement on April 1^{st} 1925. That is:

D : "*Well,* Raymond*, today is April Fool's Day, and I will deceive you as you have never been deceived before!*"

In Definition 9, abbreviaiton D is only a first, and still informal, way of writing Emile's morning announcement.[4] Intuitively, three distinct combinations of formulas $\mathtt{d}^+$ and $\mathtt{d}^-$ align with D being true: $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$, $(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)$ and $(\mathtt{d}^+ \wedge \mathtt{d}^-)$. The first combination $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ means that Raymond is deceived *only* by commission on April 1^{st} 1925. The second combination $(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)$ that he is deceived *only* by omission on that date. The last combination $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ means that Raymond is deceived *both*

[4] In Definition 12, we provide a formal expression of this announcement, written $\boldsymbol{D}$, which clearly insists on the surprise dimension of Emile's announcement of deception, based on [29, 5].

by commission *and* by omission on April 1^{st} 1925. Contrariwise, in case RAYMOND is *neither* deceived by commission *nor* by omission on April 1^{st} 1925, then formulas $\mathtt{d}^+$ and $\mathtt{d}^-$ are false, so we have the combination $(\neg\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$.

Let us now point out facts that are objectively true in Smullyan's story. To begin with, RAYMOND's upgrade with EMILE's announcement D is an action. After this action, RAYMOND is indeed *deceived by commission* $\mathtt{d}^+$ but, in fact, he is not deceived by commission *only*, so the combination $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ is false: $\neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. In addition to that, RAYMOND also turns out te be *deceived by omission* at the end of the day, so the combination $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ is true. As a result, the combination stating that RAYMOND is deceived *only* by omission is false: $\neg(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)$.

From a more pragmatic perspective, the announcement D triggers the conclusion that EMILE will deceive RAYMOND only by commission, which is encoded by the conjunctive formula $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. This is the conclusion that RAYMOND believe after EMILE's announcement D. This conclusion relies on the default inference one would make to interpret some announcement of deception in common circumstances. When hearing such an announcement, which can be written by $[\Uparrow D]$, one would naturally infer that deception only by commission $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ is *more plausible* than the negation of it $\neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. In other words, one would infer that the conjunction $D \wedge (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ is a *more plausible* option than the conjunction $D \wedge \neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. Using the classical writing of default rules given in [51], the default rule at stake in Smullyan's story can be written $D \Rightarrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$.

The issue is that the formula $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ is false in the context of Smullyan's tale: it is not the case that RAYMOND will be deceived only by commission after his brother's announcement. Actually, EMILE's explanation reveals that RAYMOND is deceived also by omission at the end of the day. Though being perfectly natural as a default rule, the inference $D \Rightarrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ should not be used by RAYMOND on April 1^{st} since using this rule leads him to draw the false conclusions that he will be deceived only by commission and thus, that he won't be deceived also by omission. We will show later that having those false beliefs is what causes his surprise afterwards. Before doing this, we take the opportunity to describe EMILE's misleading announcement in more details.

3 Smullyan's April Fool's Trick

3.1 Ex Ante Interpretation

Semantically, RAYMOND uses the default rule $D \Rightarrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ to interpret EMILE's announcement as meaning that he will deceive him only by commission, i.e. by believing that $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$: $\boldsymbol{B}(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. Let us now define a reasonable approximation of this default interpretation in language $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$.

Definition 10 (*default deception*) Let us write `default` the abbreviation of the formula $[\Uparrow D]\boldsymbol{B}(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ that we previously accepted as a reasonable approximation of the default rule $D \Rightarrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. That is:

$$\texttt{default} \quad := \quad [\Uparrow D]\boldsymbol{B}(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$$

3.2 Ex Interim Deception

3.2.1 Raymond's Two States of Deception

We now describe these two states of deception RAYMOND successively goes through during the day. On April 1^{st} 1925, RAYMOND is actually `deceived two times` by his

brother. To begin with, EMILE's morning announcement that D brings about RAYMOND's *first state of deception*. By applying the default rule $\texttt{default}$ to D, RAYMOND computes the wrong conclusion that he will be deceived only by commission on April 1^{st} 1925: $(\texttt{d}^+ \wedge \neg\texttt{d}^-)$. Accordingly, RAYMOND believes a formula, i.e. $(\texttt{d}^+ \wedge \neg\texttt{d}^-)$, which is false. So based on definition 3, RAYMOND is deceived on formula $\neg(\texttt{d}^+ \wedge \neg\texttt{d}^-)$ after his brother's announcement, which can be described by the conjunction (before and after eliminating double negation):[5]

$$\begin{aligned}\texttt{deceived}_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)} &:= \neg(\texttt{d}^+ \wedge \neg\texttt{d}^-) \wedge \boldsymbol{B}\neg\neg(\texttt{d}^+ \wedge \neg\texttt{d}^-)\\ &:= \neg(\texttt{d}^+ \wedge \neg\texttt{d}^-) \wedge \boldsymbol{B}(\texttt{d}^+ \wedge \neg\texttt{d}^-)\end{aligned}$$

Late on that day, a *second state of deception* adds to this first state when EMILE explains his April Fool's plot. Indeed, EMILE reveals that RAYMOND was not deceived only by commission but that he was in fact deceived both by commission and by omission: $(\texttt{d}^+ \wedge \texttt{d}^-)$. Since RAYMOND believes this possibility to be false all day, his *second state of deception* consists in being deceived on this formula $(\texttt{d}^+ \wedge \texttt{d}^-)$:

$$\texttt{deceived}_{(\texttt{d}^+\wedge\texttt{d}^-)} := (\texttt{d}^+ \wedge \texttt{d}^-) \wedge \boldsymbol{B}\neg(\texttt{d}^+ \wedge \texttt{d}^-)$$

To sum up, RAYMOND goes through *two states of deception* in Smullyan's tale. First, he is deceived on the fact that he will be deceived only by commission: $\texttt{deceived}_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)}$. Then, he is deceived on the fact that he will be deceived both by commission and by omission: $\texttt{deceived}_{(\texttt{d}^+\wedge\texttt{d}^-)}$. RAYMOND's first state of deception is made possible by applying the default rule $\texttt{default}$ to EMILE's announcement D. RAYMOND's second state of deception results from the fact that RAYMOND does not realize that he will be deceived also by omission after realizing that he has not been deceived (only) by commission. The next subsection is devoted to studying EMILE's attitudes that lead to those two states of deception.

3.2.2 Emile's Two Attitudes of Deception

Raymond's *first state of deception* is that he wrongly believes that he will be deceived only by commission on April 1^{st} 1925: $\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)\wedge\boldsymbol{B}\neg\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)$, or after eliminating double negation: $\neg(\texttt{d}^+ \wedge \neg\texttt{d}^-) \wedge \boldsymbol{B}(\texttt{d}^+ \wedge \neg\texttt{d}^-)$. As we have said, this state results from the $\texttt{default rule}$ he applies to EMILE's announcement. This announcement D is in fact the event that leads RAYMOND to be deceived a first time, since after this announcement D the following formula holds as an assumption:

$$[\Uparrow D](\neg(\texttt{d}^+ \wedge \neg\texttt{d}^-) \wedge \boldsymbol{B}(\texttt{d}^+ \wedge \neg\texttt{d}^-))$$

By applying definition 3 and the definition of the commision operator $\Uparrow^+$:

$$[\Uparrow^+]\texttt{deceived}_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)}$$

So by applying definition 5, we have:

$$\texttt{d}^+_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)}$$

Late in the afternoon, comes RAYMOND's second state of deception since he still does not realize that he is actually deceived both by commission and by omission: $(\texttt{d}^+ \wedge \texttt{d}^-)$. Interestingly, this second state of deception does not result from EMILE making any action, but specifically from the lack of an action, or omission. Let us sum up the event that leads to RAYMOND's *second state of deception*. Late on the afternoon, the following formula holds:

[5] From now on, we eliminate double negations immediately when it is possible.

$$[\Uparrow^-]((\mathtt{d}^+ \wedge \mathtt{d}^-) \wedge \boldsymbol{B}\neg(\mathtt{d}^+ \wedge \mathtt{d}^-))$$

By applying definition 3 and the definition of the omission operator $\Uparrow^-$:

$$[\Uparrow^-]\mathtt{deceived}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$$

So by applying definition 6:

$$\mathtt{d}^-_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$$

We are now able to summarize Emile's deceptive plot on April 1^{st} 1925 into a single formula, as given in Definition 11 hereafter.

Definition 11 (*deception plot*) We define Emile's deceptive plot on April 1^{st} 1925, written `deception`, which relies on using both deception by commission and deception by omission as follows:

$$\mathtt{deception} := (\mathtt{d}^+_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)} \wedge \mathtt{d}^-_{(\mathtt{d}^+\wedge\mathtt{d}^-)})$$

This definition is in line with the deceptive plot happening in Smullyan's story since, according to the left conjunct $\mathtt{d}^+$, Emile deceives Raymond by commission, and, according to the right conjunct $\mathtt{d}^-$, Emile also deceives Raymond by omission. Specifically, Emile deceives Raymond by commission on formula $\neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ and deceives Raymond by omission on formula $(\mathtt{d}^+ \wedge \mathtt{d}^-)$.

In the next section, we use dynamic plausibility models to show that Emile's explanation acts as a way of teaching Raymond the meaning of this formula *Deception*, that is to say his plan to deceive his brother on April 1^{st} 1925.

3.3 Ex Post Explanation

We know that late on April Fool's Day, Raymond has still not understood the trick. His mother intervenes and Emile finally unveils the inner workings of his deceptive plot. As a matter of fact, Emile's explanation can be seen as a Socratic dialogue to teach him the meaning of sentence *Deception*. To show this, we now use plausibility models to represent the dynamics of this dialogue. The epistemic landscape of all the deceptive possibilities can be given by the set:

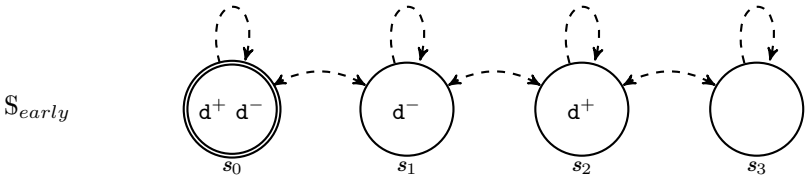
$$\{\boldsymbol{s}_0^{(\mathtt{d}^+\wedge\mathtt{d}^-)}, \boldsymbol{s}_1^{(\neg\mathtt{d}^+\wedge\mathtt{d}^-)}, \boldsymbol{s}_2^{(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}, \boldsymbol{s}_3^{(\neg\mathtt{d}^+\wedge\neg\mathtt{d}^-)}\}$$

For simplicity's sake, our modelling only represents the logical space resulting from complex formulas $\mathtt{d}^+$ and $\mathtt{d}^-$, setting aside the modelling of atomic formulas.[6] The formula holding at each state is right-upperscripted. State $\boldsymbol{s}_3$ corresponds to Emile not deceiving Raymond either by commission or by omission, $\boldsymbol{s}_2$ to deceiving him only by commission, $\boldsymbol{s}_1$ to deceiving only by omission and $\boldsymbol{s}_0$ to deceiving him by both options. The actual state is $\boldsymbol{s}_0$ since in Smullyan's story, Emile deceives Raymond by both commission and by omission. Graphically, to represent the fact that the state $\boldsymbol{t}$ is *at least as plausible as* the state $\boldsymbol{s}$ for Raymond (i.e. $\boldsymbol{s} \leqslant \boldsymbol{t}$), we draw a dashed right arrow "$\dashrightarrow$" from state $\boldsymbol{s}$ to state $\boldsymbol{t}$. When states $\boldsymbol{s}$ and $\boldsymbol{t}$ are *equally plausible* for him (that

[6] States are only labelled with complex formulas $\mathtt{d}^+$ and $\mathtt{d}^-$ since, when modelling Emile's explanation, we are only concerned with the evolution of Raymond's beliefs regarding deception (only) by commission and deception (only) by omission, both represented by complex formulas $\mathtt{d}^+$ and $\mathtt{d}^-$, respectively. For similar reasons, we only represent $\mathtt{d}^+$ and $\mathtt{d}^-$ when modelling the effect of Emile's announcement of surprise in subsubsection 4.2.2.

is $s \leqslant t$ and $t \leqslant s$), we draw a left-right dashed arrow "$\leftarrow\!\!-\!\!-\!\!-\!\!\rightarrow$" between s and t. For easier readability, we omit representing the transitivity of plausibility orders in any of the models.

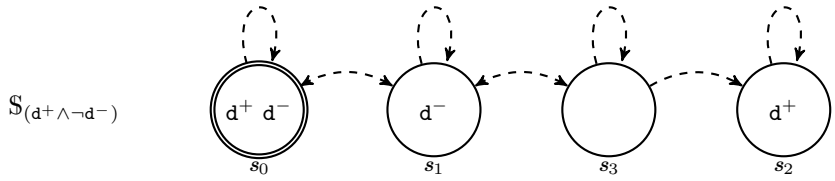
Early in the morning on April Fool's Day, before any announcement, RAYMOND has no particular belief in the possibility of his brother deceiving him on that day. More precisely, being deceived and not being deceived are equally plausible options, and so are all the states of the epistemic landscape: s_0, s_1, s_2 and s_3. The following model depicts this initial configuration, before any announcement:



According to RAYMOND's recollection, EMILE's explanation starts as follows:

EMILE: "*So, you expected me to deceive you, didn't you?*
RAYMOND: *Yes.*"

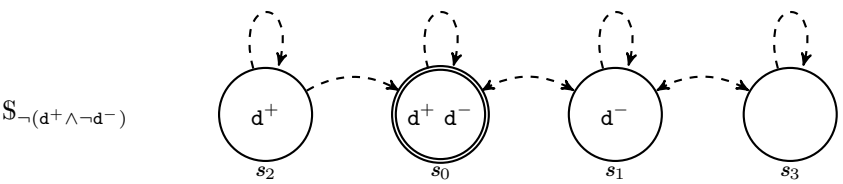
Here EMILE's question can be understood as "*So, you expected me to deceive you by some action, didn't you?*". Since this question is in the affirmative form, it can be formally expressed by RAYMOND upgrading his beliefs with formula $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ corresponding to the belief he reached through the default interpretation of EMILE's announcement. The matching operation $[\Uparrow (\mathtt{d}^+ \wedge \neg \mathtt{d}^-)]$: $\mathbb{S}_{early} \to \mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}$ returns the following model:



We can see that $\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}, s_0 \models \boldsymbol{B}(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$. Here RAYMOND acknowledges that he came to believe that $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ after his brother's announcement that D. Immediately, though, EMILE informs him that this default interpretation was wrong:

EMILE: "*But I didn't, did I?*
RAYMOND: *No.*"

By those lines, EMILE invites his brother to retract his belief in formula $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$. The corresponding dynamic operation is $[\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)]$: $\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)} \to \mathbb{S}_{\neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}$. Graphically, the model becomes:

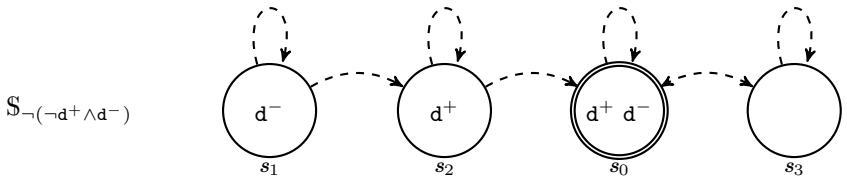


In this model, RAYMOND now believes that formula $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ is false, that is: $\mathbb{S}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}, \boldsymbol{s}_0 \models \boldsymbol{B}\neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. By doing so, he acknowledges that he was wrong to believe that $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ after his brother's announcement and thus, that he has been deceived on $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$. At this step of EMILE's explanation, RAYMOND is actually uncertain whether he has been deceived at all since $\boldsymbol{s}_0$, $\boldsymbol{s}_1$ and $\boldsymbol{s}_3$ are the most plausible states of the ordering, yet equally plausible.

But EMILE's demonstration goes even further. He explains that not only has RAYMOND been wrong, and thus deceived, on formula $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ but that he has also been deceived on formula $(\mathtt{d}^+ \wedge \mathtt{d}^-)$. The reason is that RAYMOND kept dismissing the true possibility that he would be deceived both by commission $d^+$and by omission d^-after realizing, as he just did, that he would not be deceived by commission *only*: $\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)$. EMILE starts explaining why in the following lines:

EMILE: "*But you expected me to, didn't you?*
RAYMOND: *Yes.*"

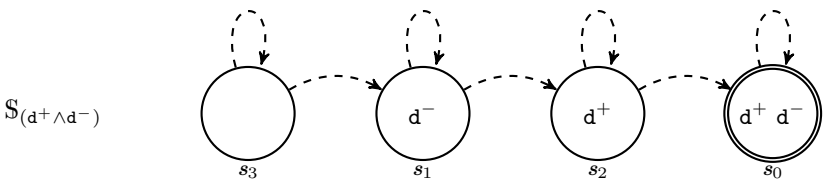
Here EMILE's repeats the same question he asked before (*"So, you expected me to deceive you, didn't you?"*) using the contrastive word *"But"* to generate a different interpretation that we can translate by reformulating EMILE's question: *"But you kept dismissing that you could be deceived also by omission, didn't you?"*. This way, EMILE invites RAYMOND to realize that he kept rejecting this option even though he observed that he was not deceived only by commission in model $\mathbb{S}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}$. Rejecting deception by omission as a possibility corresponds to the upgrade $[\Uparrow \neg(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)]$: $\mathbb{S}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)} \to \mathbb{S}_{\neg(\neg\mathtt{d}^+\wedge\mathtt{d}^-)}$ which returns the following model:



In model $\mathbb{S}_{\neg(\neg\mathtt{d}^+\wedge\mathtt{d}^-)}$, RAYMOND is now uncertain whether he is deceived both by commission and by omission, as encoded by $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ true in state $\boldsymbol{s}_0$, or not deceived at all, as encoded by $(\neg\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$ true in state $\boldsymbol{s}_3$. But still, option $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ is now a possibility. EMILE asks RAYMOND a last question to make him realize that option option $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ is in fact the true one:

EMILE: "*So I deceived you, didn't I?*
RAYMOND: "*Yes.*"

Between options $(\mathtt{d}^+\wedge\mathtt{d}^-)$ and $(\neg\mathtt{d}^+\wedge\neg\mathtt{d}^-)$, the only remaining option for RAYMOND to be deceived *per se* is of course to update his beliefs with formula $(\mathtt{d}^+ \wedge \mathtt{d}^-)$. The corresponding operation is $[\Uparrow (\mathtt{d}^+ \wedge \mathtt{d}^-)]$: $\mathbb{S}_{\neg(\neg\mathtt{d}^+\wedge\mathtt{d}^-)} \to \mathbb{S}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$ returning the final model:



We have described the distinct steps by which Raymond learns *why* and *how* he was fooled by his brother on April 1^{st} 1925. By going backward from last model $\mathbb{S}_{\mathtt{deceived}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}}$ to initial model $\mathbb{S}_{early}$, we can reconstitute the dynamics of Emile's explanation, as described in Table 1.

Dialogical Explanation	**Learning Step**	**Some Epistemic Effect**
E: "*So you expected...?* R: *Yes.*"	(a) $[\Uparrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)]$	$\boldsymbol{B}(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$
E: "*But I didn't...?* R: *No.*"	(b) $[\Uparrow \neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)]$	$\mathtt{deceived}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}$
E: "*But you expected...?* R: *Yes.*"	(c) $[\Uparrow \neg(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)]$	$\boldsymbol{B}\neg(\mathtt{d}^+ \wedge \mathtt{d}^-)$
E: "*So I deceived you...?* R: *Yes.*"	(d) $[\Uparrow (\mathtt{d}^+ \wedge \mathtt{d}^-)]$	$\mathtt{deceived}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$

Table 1: Emile's Explanation to Raymond.

Formulas (a) to (d) of Table 1 summarize the distinct steps through which Raymond learns Emile's deceptive strategy. At step (a), Raymond learns that after hearing announcement D, he made an upgrade with the default interpretation that he would be deceived only by commission: $[\Uparrow (\mathtt{d}^+ \wedge \neg\mathtt{d}^-)]$. But at step (b), Emile reveals that this default interpretation was a mistake: $[\Uparrow \neg(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)]$. By concatenating steps (a) and (b), Raymond can actually conclude that he has been deceived on formula $(\mathtt{d}^+ \wedge \neg\mathtt{d}^-)$, which is captured by $\mathtt{deceived}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}$. But at step (c), Emile informs Raymond that though realizing that he had not been deceived (only) by commission, he kept dismissing that he could be deceived only by omission: $[\Uparrow \neg(\neg\mathtt{d}^+ \wedge \mathtt{d}^-)]$. At step (d), Emile finally informs him that deception both by commission and by omission was the right conclusion to draw: $[\Uparrow (\mathtt{d}^+ \wedge \mathtt{d}^-)]$. By concatenating steps (c) and (d), Raymond can conclude that he has also been deceived on formula $(\mathtt{d}^+ \wedge \mathtt{d}^-)$, which is now captured by $\mathtt{deceived}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$.

Emile's explanation can be conceived as a way of directing Raymond's attention to the two states of deception he has been prey to: firstly, $\mathtt{deceived}_{\neg(\mathtt{d}^+\wedge\neg\mathtt{d}^-)}$; secondly, $\mathtt{deceived}_{(\mathtt{d}^+\wedge\mathtt{d}^-)}$. Through the learning process, Raymond understands *why* and *how* he has been deceived by his malicious brother. However, as we point out hereafter, this learning also throws Raymond into puzzling thoughts.

4 Some Puzzling Thoughts on Deception

In this section, we show that Raymond's late reasoning is puzzling but not of a paradoxical nature. That being shown, we argue that Emile's announcement may raise some other paradoxical issues that share echoes to the so-called "Surprise Exam Paradox" in

the epistemological and epistemic logic literature. Finally, we put more emphasis on the kind of surprise RAYMOND experiences when EMILE finally unveals his trick.

4.1 Raymond's Self-Referential Thoughts

After his brother's explanation, RAYMOND lies in bed and starts reasoning about the deception he has been subjected to. He eventually falls into a paradox that can be summed up as follows:

> *On the one hand, supposing that he,* RAYMOND*, wasn't deceived, then he didn't get what he expected (because he expected to be deceived after his brother's announcement), and hence he was actually deceived. But on the other hand, supposing that he was deceived, then he exactly did get what he expected, and hence he was not deceived stricto sensu.*

We argue that this paradox is specious by expressing RAYMOND's two lines of reasoning in $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$. More precisely, only his first line of his reasoning (*"On the one hand..."*) seems to bring about a contradiction. His second line of reasoning (*"On the other hand..."*) is not contradictory as it stands.

Let us start by examining this second line before considering the first. RAYMOND's second line of reasoning can be reformulated as follows: if RAYMOND supposes that he is deceived by his brother, he has to suppose that formula $\texttt{deceived}$ is true after having processed EMILE's announcement D: $[\Uparrow D]\texttt{deceived}$. But then, RAYMOND *exactly did get what he expected* since he expected to be deceived after his brother's announcement: $[\Uparrow D]\boldsymbol{B}(\texttt{deceived})$. From those two premises, the derivation hereafter reflects RAYMOND's conclusion that he is not deceived after all: $[\Uparrow D]\neg\texttt{deceived}$. By applying the reduction axiom for negation, i.e. axiom (*i*), we obtain a straightforward contradiction: $\neg[\Uparrow D]\texttt{deceived}$. Let us examine this derivation and look for some mistaken step:

1.	$[\Uparrow D]\texttt{deceived}$	[Hypothesis]
2.	$[\Uparrow D]\boldsymbol{B}(\texttt{deceived})$	[Hypothesis]
3.	$[\Uparrow D]\texttt{deceived} \wedge [\Uparrow D]\boldsymbol{B}(\texttt{deceived})$	1-2, Conjunction
4.	$[\Uparrow D](\texttt{deceived} \wedge \boldsymbol{B}(\texttt{deceived}))$	3, Axiom (*ii*) for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$
5.	$[\Uparrow D]\neg\texttt{deceived}_{\texttt{deceived}}$	4, Failure of Definition 3
6.	$[\Uparrow D]\neg\texttt{deceived}$	5, Failure of Definition 4
7.	$\neg[\Uparrow D]\texttt{deceived}$	6, Axiom (*i*) for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$. **Contra 1**

At step 6, it is logically incorrect to derive that $[\Uparrow D]\neg\texttt{deceived}$ by affirming that, after $[\Uparrow D]$, if RAYMOND is not deceived on formula $\texttt{deceived}$ itself, i.e. $\neg\texttt{deceived}_{\texttt{deceived}}$, then he is not deceived simpliciter, i.e. $\neg\texttt{deceived}$. This inference is too strong since RAYMOND may perfectly be deceived simpliciter, i.e. $\texttt{deceived}$, because he is deceived on formulas distinction from formula $\texttt{deceived}$ itself. This is exactly what happens in Smullyan's story in which RAYMOND is deceived on formulas $\neg(\texttt{d}^+ \wedge \neg\texttt{d}^-)$ and $(\texttt{d}^+ \wedge \texttt{d}^-)$, and thus deceived simpliciter. Therefore, the conclusion $\neg[\Uparrow D]\texttt{deceived}$ that RAYMOND obtains at step 7 is not warranted: he cannot logically conclude that after upgrading his beliefs with his brother's announcement that D, he is deceived (step 1) *only if* he is not deceived (step 7). So RAYMOND's second line of reasoning does not lead to a genuine contradiction.

Concerning RAYMOND's first line of reasoning, this line *does* lead to a genuine contradiction but only because one of RAYMOND's assumptions, viz. that he is not deceived after his brother's announcement, is false. RAYMOND's first line of reasoning can be reformulated as follows: on the one hand, if RAYMOND supposes that he is not deceived by

Emile, he has to assume that formula $\neg\texttt{deceived}$ is true after his brother's announcement D: $[\Uparrow D]\neg\texttt{deceived}$. But then, *he did not get what he expected* since he clearly expected to be deceived after his brother's announcement: $[\Uparrow D]\boldsymbol{B}(\texttt{deceived})$. From those two premises, we derive a contradiction as follows:

1.	$[\Uparrow D]\neg\texttt{deceived}$	[Hypothesis]
2.	$[\Uparrow D]\boldsymbol{B}(\texttt{deceived})$	[Hypothesis]
3.	$[\Uparrow D]\neg\texttt{deceived} \wedge [\Uparrow D]\boldsymbol{B}(\texttt{deceived})$	1-2, Conjunction
4.	$[\Uparrow D](\neg\texttt{deceived} \wedge \boldsymbol{B}(\texttt{deceived}))$	3, Axiom (*ii*) for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$
5.	$[\Uparrow D]\texttt{deceived}_{\neg\texttt{deceived}}$	4, Applying Definition 3
6.	$[\Uparrow D]\texttt{deceived}$	5, Applying Definition 4
7.	$\neg[\Uparrow D]\texttt{deceived}$	1, Axiom (*i*) for $\mathcal{L}_{(\boldsymbol{B},\Uparrow)}$. **Contra 6**

Though Raymond's reasoning is perfectly sound here, premise 1, that is to say $[\Uparrow D]\neg\texttt{deceived}$, is false. As argued in subsection 3.2.1, it is false that Raymond is not deceived after his brother's announcement D. In particular, we know that Raymond is deceived by the default interpretation he makes of Emile's announcement D as meaning deception only by commission: $[\Uparrow D]\texttt{deceived}_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)}$. Applying definition 4 on formula $\texttt{deceived}_{\neg(\texttt{d}^+\wedge\neg\texttt{d}^-)}$, we have $[\Uparrow D]$*deceived*, contrary to the assumption at step 1. In other words, the contradiction in Raymond's reasoning can be bypassed since it relies on the false assumption that $[\Uparrow D]\neg\texttt{deceived}$.

To sum up, Raymond's evening thoughts are not of a paradoxical nature. In his first line of reasoning, the assumption that he is not deceived after his brother's announcement can be rejected and in Raymond's second line of reasoning, the step leading to a contradiction can be blocked. However, we will see now that paradox may strike back if we rephrase Emile's announcement in a slightly different way putting more emphasis on another dimension of Emile's announcement: the surprise Emile announces to Raymond when saying that he will deceive him "*as [he has] never been deceived before!*". Such a surprise is noticeable in Raymond's late reasoning since he is surprised by the *type of deception* Emile used: $(\texttt{d}^+ \wedge \texttt{d}^-)$. Let us now investigate on this dimension of surprise behind Emile's announcement D.

4.2 Emile's Surprise Deception

4.2.1 Expressing Surprise in Smullyan's Story

We start out by reminding Emile's early announcement: "*Well, Raymond, today is April Fool's Day, and I will deceive you as you have never been deceived before!*". We have used a specific formula, i.e. D (Definition 9), to abbreviate this announcement and, then, defined more complex formulas, i.e. $\texttt{d}^+$ and $\texttt{d}^-$ (Definitions 5 to 8), to express deception by commission and deception by omission, respectively. Using abbreviation D was sufficient at this step to express Emile's morning announcement of deception. That said, it left aside the dimension of surprise involved in Emile's announcement, as conveyed by the formulation "*... as you have never been deceived before!*". To capture this surprise aspect of Emile's announcement, we now propose to use another formula, written $\boldsymbol{D}$, encoding the specific fact that Emile's deception will be indeed surprising for Raymond.

We may ask which explicit formal definition can be given to $\boldsymbol{D}$. Extant analyzes of the so-called "Surprise Examination Paradox" [e.g. 41, 47, 48] provide some path in that respect. In the surprise exam puzzle, a teacher announces to her students that she will give a (single) surprise exam the next week. This announcement can be compared

with EMILE's announcement $\boldsymbol{D}$ along similar lines. They both announce a surprising event to come and we will see that announcing $\boldsymbol{D}$ raises similar issues as the teacher's announcement does in the surprise exam case. To perform this comparison, we focus on the *three-day* version of the surprise exam paradox since, in Smullyan's story, there are also three distinct ways in which RAYMOND can be deceived: deception only by commission ($\mathtt{d}^+ \wedge \neg \mathtt{d}^-$), deception only by omission ($\neg \mathtt{d}^+ \wedge \mathtt{d}^-$) and deception by both commission and omission ($\mathtt{d}^+ \wedge \mathtt{d}^-$). Based on those possibilities, we provide a formal expression of EMILE's announcement based on the formalizations given by [28, 29] and [20] in the surprise exam case.[7][8]

In the *three-day* version of the puzzle, the teacher makes the following announcement to the students: *"I will give you a single surprise exam either on Monday or on Tuesday or on Wednesday next week"*. Between the teacher and the students, it is common knowledge that an exam comes as a surprise if the students do not know the evening before the exam that it will happen the next day. But after this announcement, the students, who are assumed to be perfectly rational, start reasoning and finally conclude that such an exam is impossible in principle. Logically speaking, the exam would fail to be a surprise on any day of the week. The classical argument for this failure is based on so-called *backward induction* and runs as follows: if the test takes place on Wednesday, the students would know on Tuesday evening that it is on Wednesday because Wednesday is the last available working day of the week. Then, it would not be a surprise. But nor will it be a surprise if the test is on Tuesday because the students would know on Monday evening that the test is on Tuesday (for they know that the test is not on Wednesday due to the previous reasoning). The same argument applies such that the exam cannot be on Monday either. At the end of this backward reasoning, the students would conclude that the test cannot be a surprise on any day of the week. Unfortunately, the teacher unexpectedly gives the exam on Tuesday (for instance) and this is a complete surprise.

Given the various deceptive options EMILE has, a more explicit, yet informal, formulation of EMILE's announcement $\boldsymbol{D}$ might be: *"I will surprise you by deceiving you only by commission, or only by omission, or both by commission and by omission"*. We know that Raymond misinterprets this explicit announcement in the following sense: *"I will surprise you by deceiving you only by commission"*. But from a strictly semantic point of view, all the three deceptive options, namely *deception only by commission*, *deception only by omission* and *deception both by commission and by omission*, are possible surprising options for RAYMOND and should be taken into account to capture the meaning of $\boldsymbol{D}$.

This informal formulation of EMILE's announcement $\boldsymbol{D}$ can result in many formal expressions depending on the way the concept of *surprise* is formalized. Following a basic intuition on which we elaborate later (see subsection 4.2.3), a state of surprise is generally triggered by a *mismatch of beliefs* between expectations and occurring events. Indeed, the agent is surprised by some events if those events come out to be true after

[7] Those formalizations for the surprise exam rely on using a knowledge modal operator, say $\boldsymbol{K}$, for which we could provide the following interpretation in our setting: $\mathbb{S}, \boldsymbol{s} \models \boldsymbol{K}\varphi$ *iff* for all $\boldsymbol{t} \in \mathcal{S}$, if $\boldsymbol{s} \sim \boldsymbol{t}$ then $\mathbb{S}, \boldsymbol{t} \models \varphi$, with $\sim$ a comparability relation over $\mathcal{S}$ such that $\sim := \leqslant \cup \geqslant$ (where $\geqslant$ is the converse of $\leqslant$). Notice that the way it is defined, $\sim$ is a *symmetric* preorder over $\mathcal{S}$, and thus an *equivalence relation*, since for two states $\boldsymbol{s}$ and $\boldsymbol{t}$, the conditions for relations $\boldsymbol{s} \sim \boldsymbol{t}$ and $\boldsymbol{s} \sim \boldsymbol{t}$ to obtain are identical. Accordingly, our knowledge operator $\boldsymbol{K}$ is S5, as in [20], while the operator used in [28, 29] is K45. But since our analysis relies on doxastic operator $\boldsymbol{B}$ in $\mathcal{L}_{(\boldsymbol{B}, \Uparrow)}$, this difference has no impact on our analysis.

[8] See also [10] and [13] for discussions of epistemic puzzles related to the Surprise Exam Paradox in dynamic logic settings, such as Fitch's paradox, Moore's paradox and Williamson's Margin for Error paradox.

having remained unacknowledged by agents until they come out to be true. As a result, the agent experiences states of *surprise*, or even of *astonishment*, when she learns the existence of those facts and events. To see more clearly into this, let us first provide the formal abbreviation of the announcement $\boldsymbol{D}$ we favor in Definition 12.

Definition 12 (*surprise-deception reading of* Emile*'s announcement*) We write $\boldsymbol{D}$ the reading of Emile's morning announcement insisting on the surprise dimension of the announcement. That is:

$$\begin{aligned}\boldsymbol{D} \quad := \quad & ((\mathtt{d}^+ \wedge \neg \mathtt{d}^-) \wedge \neg \boldsymbol{B}(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)) \\ & \vee ((\neg \mathtt{d}^+ \wedge \mathtt{d}^-) \wedge [\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)] \neg \boldsymbol{B}(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)) \\ & \vee ((\mathtt{d}^+ \wedge \mathtt{d}^-) \wedge [\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)][\Uparrow \neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)] \neg \boldsymbol{B}(\mathtt{d}^+ \wedge \mathtt{d}^-))\end{aligned}$$

The abbreviation provided in Definition 12 is inspired by Gerbrandy's formalization of the teacher's announcement in the Surprise Exam case, using belief radical upgrades $\Uparrow$ as proposed in [5, 8] instead of announcement operators "!".[9] Firstly, formula $\boldsymbol{D}$ states that Emile's deception will be a surprise since, according to $\boldsymbol{D}$, if Raymond is deceived only by commission, formula $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ is true but Raymond denies it to be true: $\neg \boldsymbol{B}(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$. Secondly, in case Raymond is deceived only by omission, now formula $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$ is true. But again, it will be a surprise because, after having recognized that he would not be deceived only by commission: $[\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)]$, Raymond denies $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$ to be true: $\neg \boldsymbol{B}(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$. Thirdly, formula $\boldsymbol{D}$ states that if Raymond is deceived both by commission and by omission, i.e. $(\mathtt{d}^+ \wedge \mathtt{d}^-)$, it will be a surprise too because, after having recognized that $\neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ and then that $\neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$: $[\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)][\Uparrow \neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)]$, Raymond (still) denies $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ to be true: $\neg \boldsymbol{B}(\mathtt{d}^+ \wedge \mathtt{d}^-)$. In other words, following $\boldsymbol{D}$, no matter whether Raymond is deceived only by commission or only by omission, or by both, in all cases it will come as a surprise.

That being said, however, none of the deception types encoded in formula $\boldsymbol{D}$ will surprise Raymond. Again backward induction helps explain why. Suppose that Raymond has not been deceived only by commission, i.e. $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$, or only by omission, i.e. $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$. So the only remaining option for Raymond to be deceived is to be deceived both by commission and by omission: $(\mathtt{d}^+ \wedge \mathtt{d}^-)$. But then, Raymond cannot be surprised by $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ since this is the only remaining option for him to be deceived after having rejected options $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ and $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$. Suppose now that Raymond is deceived only by omission: $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$. This implies that deception only by commission, i.e. option $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$, has not taken place, so the only remaining possibility for Raymond to be deceived is $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$, and here again he cannot be surprised by this option. Having rejected options $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ and $(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)$ as possible sources of surprise, suppose finally that Raymond is deceived only by commission: $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$. Since $(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)$ is the only remaining option for Raymond to be surprised, he won't be surprised by it. Then, if Raymond's beliefs remain consistent, he cannot be surprised at all after learning Emile's morning announcement $\boldsymbol{D}$.

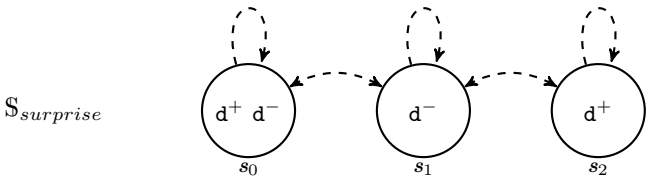
[9] A (three-day) Gerbrandy's formalization of the teacher's announcement of surprise with announcement operators would be: $(mon \wedge \neg \boldsymbol{K} mon) \vee (tue \wedge [!\neg\ mon] \neg \boldsymbol{K} tue) \vee (wed \wedge [!\neg\ mon][!\neg\ tue] \neg \boldsymbol{K} wed) \vee \boldsymbol{K} \bot$, such that "*mon*", "*tue*" and "*wed*" stand for *monday*, *tuesday* and *wednesday*, $\boldsymbol{K}$ is a K45 knowledge operator and "!" is an announcement operator. This formalization, written **S** by Gerbrandy, means that the exam will come as a surprise for the reason that if the exam is on wednesday (*wed*), the agent does not know this ($\neg \boldsymbol{K} wed$) after having learned that it is not on tuesday ($[!\neg\ tue]$); if the exam is on tuesday (*tue*), the agent does not know this ($\neg \boldsymbol{K} tue$) after having learned that it is not on monday ($[!\neg\ mon]$); if the exam is on monday (*mon*), the agent does not know it either on monday ($\neg \boldsymbol{K} mon$); if the agent has contradictory information about the date of the exam ($\bot$), the latter will also come as a surprise (according to Gerbrandy). Using the reduction axioms of dynamic epistemic logic, Gerbrandy rewrite **S** as follows: $(mon \wedge \neg \boldsymbol{K} mon) \vee (tue \wedge \neg \boldsymbol{K}(\neg mon \rightarrow tue)) \vee (wed \wedge \neg \boldsymbol{K}((\neg mon \wedge \neg tue) \rightarrow wed)) \vee \boldsymbol{K} \bot$.

4.2.2 Emile's Announcement of Surprise

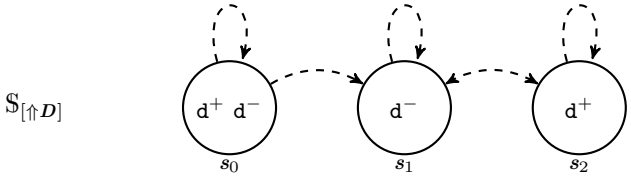
As pointed by Baltag & Smets in [5, 8], the most accepted reading of the surprise announcement is to understand the announcement as a self-referential statement, as in Gerbrandy's third expression of the teacher's announcement. In the context of Smullyan's story, this amounts to the following reading of EMILE's announcement: "*I will deceive you as you have never been deceived before, even after I am telling you all this*". This self-referential statement can be interpreted as an infinite iteration of non-self-referential upgrades as follows:

$$[\Uparrow \boldsymbol{D}]; [\Uparrow \boldsymbol{D}]; [\Uparrow \boldsymbol{D}]; ...$$

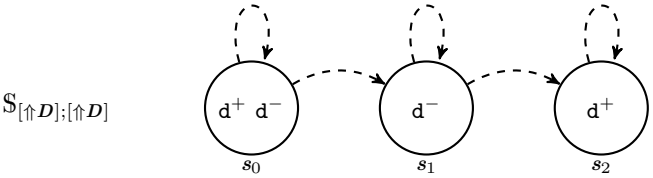
Informally, this sequence first says that EMILE's deception will be a surprise even after EMILE announcing that it will be a surprise, then it says that even after this announcement, EMILE's deception will be a surprise even after announcing it a second time, etc. To model this sequence, let us set aside the state in which RAYMOND is not deceived ($\boldsymbol{s}_3$) and only consider the states where RAYMOND is actually deceived: $\boldsymbol{s}_0$, $\boldsymbol{s}_1$, $\boldsymbol{s}_2$. Initially, before EMILE's announcement, all those states are considered to be equally plausible by RAYMOND, as shown in the following model:[10]



Let us consider the model $\mathbb{S}_{[\Uparrow \boldsymbol{D}]}$ obtained after a first iteration of $[\Uparrow \boldsymbol{D}]$ on $\mathbb{S}_{surprise}$. In $\mathbb{S}_{[\Uparrow \boldsymbol{D}]}$, all $\boldsymbol{D}$-states become more plausible than all $\neg \boldsymbol{D}$-states, strictly reducing to state $\boldsymbol{s}_0$ here.



Let us now consider the model $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$ obtained after the second iteration of $[\Uparrow \boldsymbol{D}]$, now applied on model $\mathbb{S}_{[\Uparrow \boldsymbol{D}]}$. In $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$, again, all $\boldsymbol{D}$-states from $\mathbb{S}_{[\Uparrow \boldsymbol{D}]}$ are made more plausible than all $\neg \boldsymbol{D}$-states.



Now, a fixed-point has been reached with respect to the announcement of surprise $\boldsymbol{D}$: any new update of $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$ with further operations $[\Uparrow \boldsymbol{D}]$ will leave model $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$

[10] As in subsection 3.3 when modelling EMILE's explanation, we do not show the transitivity of plausibility orders in any of the models to enhance readability.

unchanged. The reason is that in model $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$, update $[\Uparrow \boldsymbol{D}]$ is still infinitely executable but the result is empty: $\boldsymbol{D}$ is false, Raymond believes it to be false and surprise cannot happen.

Based on this conclusion that $\boldsymbol{D}$ cannot be true, we may wonder why Raymond does not simply dismiss all the iterated upgrades $[\Uparrow \boldsymbol{D}]$ he has made and thus go back to his original plausibility order expressed in model $\mathbb{S}_{surprise}$. However, as argued by [5] in the surprise examination case, Raymond would then cancel all the reasons that led him to conclude that Emile's announcement $\boldsymbol{D}$ was false. If he reverts to the initial model $\mathbb{S}_{surprise}$, then he no longer knows, more precisely believes, that Emile's announcement is false, and then surprise is possible. That Raymond concludes Emile's announcement to be false does not provide Raymond any warrant to dismiss all the iterated radical upgrades he made since, precisely, this conclusion is warranted by all those upgrades. In other words, Raymond should rationnally stick to final model $\mathbb{S}_{[\Uparrow \boldsymbol{D}];[\Uparrow \boldsymbol{D}]}$ where surprise fails in principle.

So to end this section, let us now explain why, contrary to this conclusion, Raymond is actually surprised by Emile. Subsection 3.3 where Emile's explanation is modelled explains why and how Raymond is in fact surprised in a way consistent with Gerbrandy's reading of surprise $\boldsymbol{D}$. As one can observe in model $\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}$, since $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ is true in the actual state s_0:

$$\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}, s_0 \models (\mathtt{d}^+ \wedge \mathtt{d}^-)$$

But we can also verify that the following dynamic formula is true:

$$\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}, s_0 \models [\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)][\Uparrow \neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)]\boldsymbol{B}\neg(\mathtt{d}^+ \wedge \mathtt{d}^-)$$

So, we have by conjunction:

$$\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}, s_0 \models (\mathtt{d}^+ \wedge \mathtt{d}^-) \wedge [\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)][\Uparrow \neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)]\boldsymbol{B}\neg(\mathtt{d}^+ \wedge \mathtt{d}^-)$$

Crucially, this conjunction admits formula $\boldsymbol{D}$ as a semantic consequence, by the consistency axiom for $\boldsymbol{B}$:

$$(\mathtt{d}^+ \wedge \mathtt{d}^-) \wedge [\Uparrow \neg(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)][\Uparrow \neg(\neg \mathtt{d}^+ \wedge \mathtt{d}^-)]\boldsymbol{B}\neg(\mathtt{d}^+ \wedge \mathtt{d}^-) \models \boldsymbol{D}$$

Then, applying the semantic deduction theorem, we have:

$$\mathbb{S}_{(\mathtt{d}^+ \wedge \neg \mathtt{d}^-)}, s_0 \models \boldsymbol{D}$$

In other words, following Gerbrandy's notion of surprise we encoded by $\boldsymbol{D}$ in our context, the source of Raymond's surprise is Emile revealing that he was deceived both by commission and by omission. In other words, Raymond is surprised exactly by learning that he was deceived by option $(\mathtt{d}^+ \wedge \mathtt{d}^-)$ he had not anticipated at all. In the next subsection, however, we argue that this surprise à-la-Gerbrandy is not the only kind of surprise Raymond is preyed to in Smullyan's tale.

4.2.3 Raymond's Two States of Surprise

In line of one of the Gerbrandy's reading of surprise [29], Raymond is surprised when Emile unveils his April Fool's deceptive trick at the end of the day. Here we rely on conceptual analysis to show that Raymond also encounters another form of surprise which is weaker than the one resulting from Emile's explanation.

According to [e.g. 40, 36], a cognitive agent reaches a state of surprise when she is led to recognize an *inconsistency*, that is to say a *discrepancy* or a *mismatch*, between what she believes about the world and the actual state of the world. Following a basic intuition, an event which is unexpected is considered "surprising" and the more unexpected it is, the more surprising the event turns out to be [e.g. 42, 40]. In other words, an event that was expected but does not happen *is* surprising. But an event which actually happens but was totally unexpected is *even more* surprising.

In that sense, Lorini & Castelfranchi distinguish two main kinds of surprise. The first kind is called "*mismatch-based surprise*" and results from a "*conflict between a perceived fact and a scrutinized representation*". In that case, the agent is surprised because she has some anticipatory representation of a fact or event, but she cannot make the incoming data fit with this anticipation. Accordingly, the intensity of the induced surprise depends of the probability, more crucially on the *implausibility*, which is assigned by the agent to the conflicting data she receives.

But Lorini & Castelfranchi contrast this first kind of surprise with a stronger kind named "*astonishment*" (or "*surprise in recognition*"). Contrary to the first kind, this type of surprise is of a *second-order nature* since now it is rooted on the *recognition* of the implausibility of a perceived fact compared to expectations: "[*we*] *perceive a certain fact and recognize the implausibility of this*". Lorini & Castelfranchi also precise that this astonishment can depend upon two distinct mental processes. Firstly, we can be astonished by a fact/event φ because we assigned a high probability to $\neg\varphi$ and after perceiving φ we realize that we would not have expected that event φ [see 36, 3]. Second, we can be astonished by φ in case perceiving φ makes us infer the falsity (and incongruity) of our initial disbelief that φ or belief that $\neg\varphi$.[11]

In Smullyan's tale, both kinds of deeper surprise are involved because both forms of unexpectation are involved. In the previous subsection based on Gerbrandy's reading of surprise, Emile's explanation leads Raymond to be *astonished* (or *surprised in recognition*) when Emile tells him that he has been deceived both by commission and by omission. Since Raymond kept discarding this configuration as very implausible all along the day, Raymond is astonished when he recognized that Emile used this option to trick him. But in fact, aside from this Gerbrandy's reading, Raymond is also subjected to mismatch-based surprise on April 1^{st} 1925. As we know, Raymond expects to be deceived only by commission after Emile's announcement and until his late explanation, but this does not happen. As a result of this absence, Raymond also encounters mismatch-based surprise before being astonished later.

Consistent with Lorini & Castelfranchi's distinction is the fact that Raymond's *astonishment* is stronger in terms of surprise compared to his state of *mismatch-based surprise*. Astonishement follows from his strong disbelief that he could be deceived both by commission and by omission. His belief that he would be deceived only by commission is less entrenched and only leads him to moderate surprise.

[11] Lorini & Castelfranchi contrast those "*deeper and slower forms of surprise which are due to symbolic representations of expected events*"[see 36, 1] to a *first-hand surprise* which corresponds to a perceptual mismatch between a stimulus (namely between the agents can *see* or *hear* in their immediate environment) and what the agents expect from a sensory-motor perspective.

5 Conclusion

We used dynamic belief revision to analyze a puzzle in which Raymond Smullyan's brother Emile uses a deceptive stratagem to fool him on April 1^{st} 1925. By a morning announcement, Emile misleads Raymond's expectations to be deceived (only) by commission. But in doing so, he also sets the stage for deceiving him by omission too. This stratagem leads Raymond to paradoxical thoughts and strong surprise. But contrary to what it seems, Raymond's thoughts are not paradoxical: only the success of Emile's announcement is. Based on a formal comparison with the 'Surprise Exam Paradox', we finally observed that Emile's deception by commission and omission is the main source of Raymond's surprise.

Raymond's reasoning on surprise is flawless but it results in the false conclusion that Emile will not surprise him. Of course, this should not invite Raymond to dismiss all the iterated radical upgrades he has made on April 1^{st} 1925 [see 5], but one lesson for *future* Raymond is to be more cautious with Emile's new announcements. An interesting investigation in that respect would be to consider so-called "improvement operators" [34, 14] in the future. As observed for Moorean sentences [see e.g. 7], Raymond's mistaken conclusion of no surprise is an illustration that the "*success postulate*" of the AGM framework [2] can be problematic. Still, formulas that do not give rise to any Moorean phenomena are successful in those settings. Improvement operators are less permissive in that respect since some versions authorize the plausibility of the information to increase iteratively without being accepted immediately, — no matter the logical form of the incoming formula. This specificity aligns with the realistic scenario where Raymond should now treat all Emile's announcements with cautious skepticism.

The theoretical failure of Emile's announcement also raises a more epistemological discussion on whether Emile's announcement qualifies as a lie or not. Traditionally, lying consists in making "*a believed-false statement to another person with the intention that the other person believe that statement to be true*" [37]. More precisely, a speaker X lies to a hearer Y on a proposition p if and only if (i) X believes that p is false (*untruthfulness clause*), (ii) X intends Y to believe that p is true (*intention-to-deceive clause*), and (iii) X tells Y that p (*addressed statement clause*).[12] Following this so-called '*subjective account*' then, Emile's announcement cannot be considered a lie since condition (i) is not met: Emile fails to be *untruthful* since he believes his morning utterance to be true. So, Emile is not lying in the classical sense.

That said, Emile's announcement also a default interpretation leading Raymond to be deceived. Conceptually, such a misleading rule pertains to a broader category we may call *veridical deception*. In veridical deception, a non-cooperative speaker makes an utterance that is (objectively) true but whose pragmatic interpretation is false and intended to deceive some addressee. Veridical deception also includes *false implicatures*, *pretending to deceive* and *presupposition faking*. In *false implicatures*, speakers aim to make addressees believe a false information φ by telling them some true information ψ which conversationally implicates φ [e.g. 1, 38, 39]. With *pretending to deceive* [e.g. 53, 25], speakers make addressees disbelieve a true piece of information φ by arousing

[12] In more details, condition (i) requires that the speaker be insincere (or dishonest) with respect to p: she believes that p is false. Condition (ii) requires the speaker to expect her addressee to believe that p is true. Finally, condition (iii) requires the speaker to assert that p to some specific addressee for her utterance to count as a lie. Note that in this subjective definition, it is not required that content p be *objectively* false: according to condition (i), it is sufficient that the speaker *subjectively* believe that p is false. In that sense, one can perfectly lie while saying the truth.

their suspicion towards the source of φ. Finally, *presupposition faking* [e.g. 31, 53], consists in making addressees inadequately believe a true formula φ that the speakers fail to semantically or pragmatically account for.

Classically, lying is considered a strict semantic phenomenon, contrary to veridical deception which involves pragmatics. What matters is that the liar believe the semantic content of her utterance to be false (as in the *untruthfulness clause*), — no matter whether she believes the pragmatic interpretations of this utterance is false or believed to be false. By contrast, in veridical deception, and in EMILE's misleading default in particular, the speaker believes the semantic content of her utterance to be true, although its pragmatic meaning is false, or believed to be false. Accordingly, veridical deception does not involve any untruthfulness clause.

The relationship between lying and veridical deception has been discussed concerning false implicatures in particular [e.g. 23, 1, 39, 22, 52, 55]. For most authors, such as [24] and [50], false implicatures are *not lies* since the semantic meaning of the speaker's utterance is not defective, though its pragmatic meaning is. But for some others [e.g. 1, 39, 22, 52], untruthful implicatures are lies, consistent with the Gricean view that truthful statements which conversationally implicate believed-to-be-false statements *are* lies. Since the distinction between defaults and implicatures is debated [e.g. 35, 44, 3, 32], the question of whether misleading defaults are lies also remains open. Conducting empirical studies would be useful to help adjudicate on this issue. We leave this investigation for future work.

Acknowledgments

We are grateful to Alexandru Baltag, Timothée Bernard, Quentin Blomet, Denis Bonnay, Philippe Capet, Hans van Ditmarsch, Paul Égré, Raul Fervari, Jakob Süskind, Peter van Emde Boas, and Sonja Smets for stimulating exchanges and comments on intermediary versions of the manuscript. This work was supported by the programs HYBRINFOX (ANR-21-ASIA-0003), FRONTCOG (ANR-17-EURE-0017), and THEMIS (DOS-0222794/00 and DOS-0222795/00).

References

[1] Jonathan E Adler. Lying, deceiving, or falsely implicating. *The Journal of Philosophy*, 94(9):435–452, 1997.

[2] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The journal of symbolic logic*, 50(2):510–530, 1985.

[3] Kent Bach. Semantic slack: What is said and more. *Foundations of speech act theory: Philosophical and linguistic perspectives*, pages 267–291, 1994.

[4] Alexandru Baltag, Lawrence S. Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. *Proceeding TARK '98 Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, 1998.

[5] Alexandru Baltag and S Smets. Surprise?! an answer to the hangman, or how to avoid unexpected exams! In *Logic and Interactive Rationality Seminar (LIRA), slides*, 2009.

[6] Alexandru Baltag and Sonja Smets. Dynamic belief revision over multi-agent plausibility models. In *Proceedings of LOFT*, volume 6, pages 11–24, 2006.

[7] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. *Logic and the foundations of game and decision theory (LOFT 7)*, 3:9–58, 2008.

[8] Alexandru Baltag and Sonja Smets. Multi-agent belief dynamics. *course given at NASSLLI*, 2010.

[9] Alexandru Baltag and Sonja Smets. Keep changing your beliefs, aiming for the truth. *Erkenntnis*, 75(2):255, 2011.

[10] Johan van Benthem. What one may come to know. *Analysis*, 64(282):95–105, 2004.

[11] Johan van Benthem. Dynamic logic for belief revision. *Journal of applied non-classical logics*, 17(2):129–155, 2007.

[12] Johan van Benthem and Sonja Smets. Dynamic logics of belief change. In *Handbook of Logics for Knowledge and Belief*, pages 299–368. College Publications, 2015.

[13] Denis Bonnay and Paul Égré. Knowing one's limits: An analysis in centered dynamic epistemic logic. In Patrick Girard, Olivier Roy, and Mathieu Marion, editors, *Dynamic Formal Epistemology*, pages 103–126. Springer, 2011.

[14] Richard Booth, Eduardo L Fermé, Sébastien Konieczny, and Ramón Pino Pérez. Credibility-limited improvement operators. In *ECAI*, volume 263, pages 123–128, 2014.

[15] Gauvain Bourgne, Camilo Sarmiento, and Jean-Gabriel Ganascia. Ace modular framework for computational ethics: dealing with multiple actions, concurrency and omission. In *International Workshop on Computational Machine Ethics*, 2021.

[16] Ilaria Canavotto. *Where responsibility takes you.* PhD thesis, Springer, 2020.

[17] Roderick M. Chisholm and Thomas D. Feehan. The intent to deceive. *Journal of Philosophy*, 74(3):143–159, 1977.

[18] Randolph Clarke. What is an omission? *Philosophical Issues*, 22:127–143, 2012.

[19] Randolph K Clarke. *Omissions: Agency, metaphysics, and responsibility.* Oxford University Press, 2014.

[20] Hans P. van Ditmarsch and Barteld Kooi. The secret of my success. *Synthese*, 153(2):339–339, 2006.

[21] Hein Duijf. *The logic of responsibility voids.* Springer, 2022.

[22] Marta Dynel. A web of deceit: A neo-gricean view on types of verbal deception. *International Review of Pragmatics*, 3(2):139–167, 2011.

[23] Don Fallis. A conceptual analysis of disinformation. *iConference 2009 Papers*, 2009.

[24] Don Fallis. What is lying? *The Journal of Philosophy*, 106(1):29–56, 2009.

[25] Don Fallis. The varieties of disinformation. In *The Philosophy of Information Quality*, pages 135–161. Springer, 2014.

[26] Tim French and Hans Van Ditmarsch. Undecidability for arbitrary public announcement logic. In *Advances in modal logic*, pages 23–42. College publications, 2008.

[27] Peter Gärdenfors and Hans Rott. Belief revision. *Computational Complexity*, 63(6):35–132, 1995.

[28] Jelle Gerbrandy. *Bisimulations on planet Kripke*. ILLC Dissertation Series, 1999.

[29] Jelle Gerbrandy. The surprise examination in dynamic epistemic logic. *Synthese*, 155(1):21–33, 2007.

[30] Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. *Journal of logic, language and information*, 6(2):147–169, 1997.

[31] Peter Harder and Christian Erik J. Kock. *The theory of presupposition failure*. Akademisk forlag, 1976.

[32] Laurence R Horn. The border wars: A neo-gricean perspective. *Where semantics meets pragmatics*, 16:21–48, 2006.

[33] Aaron Hunter and François Schwarzentruber. Arbitrary announcements in propositional belief revision. In *DARe@ IJCAI*, 2015.

[34] Sébastien Konieczny and Ramón Pino Pérez. Improvement operators. *KR*, 8:177–187, 2008.

[35] Stephen C Levinson. *Presumptive meanings: The theory of generalized conversational implicature*. MIT press, 2000.

[36] Emiliano Lorini and Cristiano Castelfranchi. The unexpected aspects of surprise. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(06):817–833, 2006.

[37] James E Mahon. The definition of lying and deception. *The Stanford Encyclopedia of Philosophy*, 2015.

[38] Jörg Meibauer. Lying and falsely implicating. *Journal of Pragmatics*, 37(9):1373–1399, 2005.

[39] Jörg Meibauer. *Lying at the semantics-pragmatics interface*. Walter de Gruyter GmbH & Co KG, 2014.

[40] Wulf-Uwe Meyer, Rainer Reisenzein, and Achim Schützwohl. Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21(3):251–274, 1997.

[41] Donald J O'Connor. Pragmatic paradoxes. *Mind*, 57(227):358–359, 1948.

[42] Andrew Ortony and Derek Partridge. Surprisingness and expectation failure: what's the difference? In *IJCAI*, pages 106–108, 1987.

[43] Jan Plaza. Logics of public communications. *Synthese*, 158(2):165, 2007.

[44] François Récanati. *Literal meaning*. Cambridge University Press, 2004.

[45] Hans Rott. Conditionals and theory change: Revisions, expansions, and additions. *Synthese*, 81:91–113, 1989.

[46] Gilbert Ryle. Negative'actions'. *Hermathena*, pages 81–93, 1973.

[47] Michael Scriven. Paradoxical announcements. *Mind*, 60(239):403–407, 1951.

[48] Robert Shaw. The paradox of the unexpected examination. *Mind*, 67(267):382–384, 1958.

[49] Raymond M Smullyan. *What is the Name of this Book? The Riddle of Dracula and Other Logical Puzzles: Mysteries, Paradoxes, Gödel's Discovery*. Prentice-Hall, 1978.

[50] Roy Sorensen. Lying with conditionals. *The Philosophical Quarterly*, 62(249):820–832, 2012.

[51] Frank Veltman. Defaults in update semantics. *Journal of philosophical logic*, 25(3):221–261, 1996.

[52] Emanuel Viebahn. Non-literal lies. *Erkenntnis*, 82(6):1367–1380, 2017.

[53] Jocelyne M. Vincent and Cristiano Castelfranchi. On the art of deception: how to lie while saying the truth. *Possibilities and limitations of pragmatics*, pages 749–777, 1981.

[54] Elazar Weinryb. Omissions and responsibility. *The Philosophical Quarterly (1950-)*, 30(118):1–18, 1980.

[55] Alex Wiegmann and Pascale Willemsen. How the truth can make a great lie: An empirical investigation of lying by falsely implicating. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, pages 3516–3621, 2017.