Logics of communication and knowledge
Sietsma, F.A.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
9.1 Introduction

In the first part of this thesis I considered models of truthful communication. Furthermore, in Chapter 8 I considered a model of belief and belief revision, which can alternatively be viewed as a model of preference and preference aggregation. Here, I will investigate what happens when agents hear a lie, which they may believe or not. This chapter has a somewhat more philosophical flavour than the previous chapters, which are of a more technical nature.

The first question I would like to ask is the following: What is a lie?

The church father St. Augustine, who wrote at length about lying in De Mendacio [St. Augustine, 1988], holds a subtle view on what lying is and what it is not. I will take his view as our point of departure. Here is his famous quote on what lying is not.

For not every one who says a false thing lies, if he believes or opines that to be true which he says. Now between believing and opining there is this difference, that sometimes he who believes feels that he does not know that which he believes, (although he may know himself to be ignorant of a thing, and yet have no doubt at all concerning it, if he most firmly believes it:) whereas he who opines, thinks he knows that which he does not know. Now whoever utters that which he holds in his mind either as belief or as opinion, even though it be false, he lies not. For this he owes to the faith of his utterance, that he thereby produce that which he holds in his mind, and has in that way in which he produces it. Not that he is without fault, although he lie not, if either he believes what he ought not to believe, or thinks he knows what he knows not, even though it should be true: for he accounts an unknown thing for a known.
St. Augustine, *De Mendacio* (On Lying), ca. AD 395 [St. Augustine, 1988]

And on what lying is:

Wherefore, that man lies, who has one thing in his mind and utters another in words, or by signs of whatever kind. Whence also the heart of him who lies is said to be double; that is, there is a double thought: the one, of that thing which he either knows or thinks to be true and does not produce; the other, of that thing which he produces instead thereof, knowing or thinking it to be false. Whence it comes to pass, that he may say a false thing and yet not lie, if he thinks it to be so as he says although it be not so; and, that he may say a true thing, and yet lie, if he thinks it to be false and utters it for true, although in reality it be so as he utters it. For from the sense of his own mind, not from the verity or falsity of the things themselves, is he to be judged to lie or not to lie. Therefore he who utters a false thing for a true, which however he opines to be true, may be called erring and rash: but he is not rightly said to lie; because he has not a double heart when he utters it, neither does he wish to deceive, but is deceived. But the fault of him who lies, is the desire of deceiving in the uttering of his mind; whether he do deceive, in that he is believed when uttering the false thing; or whether he do not deceive, either in that he is not believed, or in that he utters a true thing with will to deceive, which he does not think to be true: wherein being believed, he does not deceive though it was his will to deceive: except that he deceives in so far as he is thought to know or think as he utters.

St. Augustine, [St. Augustine, 1988]

I cannot do better than to follow St. Augustine in assuming that the intention to mislead is part of the definition of a liar. Thus, to me, lying that is communicating $p$ in the belief that $\neg p$ is the case, with the intent to be believed.

The deceit involved in a lie that $p$ is successful, if $p$ is believed by the addressee after the speaker’s utterance. This is my perspective. As is common in dynamic epistemic logic, I model the agents addressed by the lie, but I do not (necessarily) model the speaker as one of those agents. Dynamic epistemics model how to incorporate novel information after the decision to accept that information, just like in belief revision. I do not claim that this decision is irrelevant, far from that, but merely that this is a useful abstraction allowing me to focus on the information change only. This further simplifies the picture: I do not need to model the intention of the speaker, nor do I need to distinguish between knowledge and belief of the speaker: he is the observer of the system and his beliefs are taken to be the truth by the listeners. In other words, instead of having a precondition ‘the speaker believes that $p$ is false’ for a lie, I have as a precondition ‘$p$ is false’.
9.1. Introduction

In the previous chapters on truthful communication, the relations of the models I used were equivalence relations. In other words, the models were S5 models. In Chapter 8 I already briefly mentioned the fact that while truthful communication corresponds to S5 models, belief is often taken to correspond to KD45 models. I will now focus on these KD45 models. The logic also allows for even less specific notions than knowledge or belief. My analysis applies to all equally, and for all such epistemic notions I will use a doxastic modal operator $B_a p$, for ‘agent $a$ believes that $p$’. My analysis is not intended as a contribution to epistemology. I am aware of the philosophical difficulties with the treatment of knowledge as (justified) true belief [Gettier, 1963].

It is also possible to model the speaker explicitly in a modal logic of lying (and I will do so in examples) and extend my analysis to multi-agent systems wherein the deceptive interaction between speakers and hearers is explicit in that way. However, I do not explore that systematically here.

The intention to be believed can also be modeled in a (modal) logical language, namely by employing, for each agent, a preference relation that is independent from the accessibility relation for belief. This is to account for the fact that people can believe things for which they have no preference, and vice versa. This perspective is, e.g., employed in [Sakama et al., 2010] - this contains further references to the expansive literature on beliefs and intentions.

The moral sides to the issue of lying are clarified in the ninth of the ten commandments (‘Thou shalt not bear false witness’) and the fourth of the five Buddhist precepts (‘I undertake the precept to refrain from false speech’). On the other hand, in the Analects of Confucius, Confucius is quoted as condoning a lie if its purpose is to preserve social structure:

The Governor of She said to Confucius, ‘In our village we have an example of a straight person. When the father stole a sheep, the son gave evidence against him.’ Confucius answered, ‘In our village those who are straight are quite different. Fathers cover up for their sons, and sons cover up for their fathers. In such behaviour is straightness to be found as a matter of course.’ Analects, 13.18.

Among philosophical treatises, the quoted text of St. Augustine is a classic. For more, see [Bok, 1978] and [Arendt, 1967] and the references therein.

Rather than dwell on the moral side of the issue of lying, here I will study its logic, focusing on simple cases of lying in game situations, and on a particular kind of public announcement that may be deceptive and that I call ‘manipulative update’. Thus, I abstract from the moral issues. I feel that it is important to understand why lying is tempting (why and how it pays off) before addressing the choice between condemnation and absolution.

The rest of the chapter is structured as follows. First, in Section 9.2, I develop a logic of lying in public discourse, treating a lie as an update with a communication believed to be truthful. Next, I turn to lying in games, by analyzing the game
of Liar’s Dice, first in terms of game theory (Section 9.3), next in terms of (an implementation of) my logical system (Section 9.4). Section 9.5 concludes with a reflection on the difference between my logic of lying as manipulative update and lying in Liar’s Dice.

9.2 The Logic of Lying in Public Discourse

We get lied to in the public domain, all the time, by people who have an interest in obfuscating the truth. In 1993 the tobacco company Philip Morris tried to discredit a report on Respiratory Health Effects of Passive Smoking by founding, through a hired intermediary, a fake citizen’s group called The Advancement of Sound Science or TASSC, to cast doubt on it. Exxon-Mobile used the same organisation to spread disinformation about global warming. Their main ploy: hang the label of ‘junk science’ on peer-reviewed scientific papers on smoking hazards or global warming, and promote propaganda disguised as research and ‘sound science’. It worked beautifully for a while, until the New York Times exposed the fraud [Montague, April 29, 1998]. As a result, many educated people are still in doubt about the reality of global warming, or think the issues are just too hard for them to understand.

It has frequently been noted that the surest result of brainwashing in the long run is a peculiar kind of cynicism, the absolute refusal to believe in the truth of anything, no matter how well it may be established. In other words, the result of a consistent and total substitution of lies for factual truth is not that the lie will now be accepted as truth, and truth be defamed as lie, but that the sense by which we take our bearings in the real world - and the category of truth versus falsehood is among the mental means to this end - is being destroyed.


Now this situation where complete cynicism reigns is one extreme attitude to confront lying. This is of course at the price of also no longer believing the truth. This attitude will be explored in my analysis of the game Liar’s Dice, where the rules of the game allow any utterance regardless of its truth. The only thing that counts is winning. As everyone knows this, this is some kind of fair play.

The other extreme is the attitude where all lies are believed. This will be the logic of successful lies, where I take successful to mean that the addressees accept the lie as truth, even at the price of believing inconsistencies. Below I will give a logic of possibly deceptive public speech acts, to model the effects of lying as in politics. Proposition 9.2.10 below can be seen as a clear vindication that Arendt is right about the grave consequences of lying in politics.

---

I will use Kripke models as defined in Chapter 2 to model the beliefs of a group of agents, and the modal language presented there to reason about them. I will use $B_a \phi$ as a shorthand for $[a] \phi$. It expresses that agent $a$ believes $\phi$. I will use action models with substitutions as defined in Chapter 3, Definition 3.3.4 to model the event that the agents hear a lie. The constraint I will put on these models is that they are KD45 models, as defined in Chapter 2. The class of KD45 models is characterized by the following axioms:

\[
\neg B_a \perp \\
B_a \phi \rightarrow B_a B_a \phi \\
\neg B_a \phi \rightarrow B_a \neg B_a \phi
\]

The first axiom states that no agent believes an inconsistency. The second is called positive introspection, and it states that if an agent believes something, then he believes that he believes it. The third axiom is negative introspection: if an agent does not believe something, then he believes that he does not believe it.

If I would also want to model the intention to deceive, I would need to use doxastic preference models ($W, V, R, S$), where $S$ is a second relation for preference. Then it is reasonable to let $S$ satisfy the KD45 postulates, or the constraint of linkedness that I presented in Chapter 8. But rather than carry such preference relations along in the exposition, I will indicate at appropriate places how they can be dealt with.

As I already indicated in Chapter 8 there is a problem with the logic of KD45 structures with KD45 updates, namely that this model class is not closed under execution of such updates. A single-agent example suffices to demonstrate this: consider a KD45 agent incorrectly believing that $p: \neg p \land B_i p$. Now inform this agent of the truth of $\neg p$. Then his accessibility relation becomes empty and is no longer serial. Another way to see that KD45 is no longer satisfied is by observing that the axiom $\neg B_a \perp$ no longer holds. The agent now believes everything! This means that the logic that incorporates updates with any action model as modal operators such as proposed in [van Benthem et al., 2006] cannot be complete with respect to the class of KD45 Kripke models. Therefore, I will not include a modal operator that consists of the update with an arbitrary action model in my logic. Rather, I will introduce certain updates representing a lie that will preserve the KD45 properties.

First, take the prototypical example of lying about $p$. Picture an initial situation where agent $a$ knows that $p$, and agent $a$ knows that agents $b$ and $c$ do not know that $p$. One way to picture this initial situation is like this:
The gray shading indicates that 0 is the actual world. Because the relations are no longer assumed to be reflexive, in this chapter I will explicitly draw all reflexive relations. Note that agent $a$ believes that $p$ (agent $a$ even knows that $p$, but this difference is immaterial to my analysis), but agents $b, c$ also consider it possible that agent $a$ believes the opposite (which is the case in world 1), or that agent $a$ has no beliefs whatsoever about $p$ (the situation in worlds 2 and 3).

In typical examples of bearing witness in court, the situation is often a bit different. In cases of providing an alibi, for example, the question ‘Was the accused at home with you during the evening of June 6th?’ is posed on the understanding that the witness is in a position to know the true answer, even if nobody can check that she is telling the truth.

Let us assume that everyone knows that $a$ knows whether $p$. The picture now becomes:

Assume agent $a$ sends a group communication to $b, c$ to the effect that $\neg p$. Would the following action model be a correct representation of the lie that $\neg p$?

It is easy to see that this cannot be right. The result of this update is a model that has no actual worlds, i.e., an inconsistent model, since the actual world has $p$ true, and the precondition of the actual action is $\neg p$. 

\[ a \rightarrow \neg p, b, c \rightarrow \neg p \]
Rather, the misleading communication should be modeled as a KD45 action model, as follows:

\[
\begin{array}{c}
\text{0: } \top \\
\text{1: } \neg p
\end{array}
\]

The misleading agent \(a\) knows that no truthful communication is being made, but the two agents \(b,c\) mistakenly believe that \(\neg p\) is truthfully being asserted. The fact that the originator of the lie does believe that \(p\) is true can be taken on board as well, of course:

\[
\begin{array}{c}
\text{0: } B_a p \\
\text{1: } \neg p
\end{array}
\]

This update can equally be seen as agent \(a\) lying about \(p\), or as an observer, not modeled in the system, lying about agent \(a\) believing that \(p\). It cannot be called an explicit of a lie by agent \(a\), because it cannot be distinguished from the (in fact more proper) perspective of an observer ‘knowing’ (believing, and with justification, as he is omniscient) that \(B_a p\).

In the context of doxastic preference models, the precondition for the actual action could be extended even further, with the intent to mislead: in \(a\)’s most preferred worlds, his victims believe that \(\neg p\). I will omit the formal details in the interest of readability.

Updating the initial model with this action model gives:

\[
\begin{array}{c}
(0,0): p \\
(1,1): \neg p
\end{array}
\]

This is a model where \(a\) believes that \(p\), where \(b,c\) mistakenly believe that \(\neg p\), and where \(b,c\) also believe that \(a\) believes that \(\neg p\). Note that the model is KD45: beliefs are still consistent (\([a] \phi \rightarrow (a) \phi\) holds in the model), but the model is not truthful anymore (there are \(\phi\) and \(a\) for which \([a] \phi \rightarrow \phi\) does not hold, i.e., there are false beliefs).

This way to model lying suggests a natural generalization of the well-studied concept of a public announcement. In the logic of public announcements [Plaza,
1989, Gerbrandy, 1999], a public announcement !φ is always taken to be a true statement. A more realistic version of public announcements leaves open the possibility of deceit, as follows. A possibly deceptive public announcement φ is a kind of ‘if then else’ action. In case φ is true, the announcement is a public update with φ, in case φ is false, the public is deceived into taking φ as true. The manipulative update with p by an outside observer (the announcer/speaker, who is not modeled as an agent in the structure), in a setting where the public consists of a, b, c, looks like this:

There are two actual events, one for the situation where p is true - in this case, the public is duly informed - and one for the situation where p is false - in this case the public is misled to believe that p. This action model can be simplified, as follows:

Call this the two-pointed manipulative update for p. I will refer to this action model as UP. I will refer to the variation on this action model where only event 0 is actual as U0P. This action model denotes the lie with p. I will refer to the variant with only event 1 actual as U1P. This action model denotes the public announcement with p.

Let me introduce operations for these actions. The manipulative update with φ is denoted ʃφ, and its two variants are denoted ¡φ (for the lie that φ) and !φ (for the public announcement that φ).

I will include these updates as modal operators in my language. Define the logic of individual belief and manipulative update LBM as follows:

\[ \phi ::= p \mid \neg \phi \mid \phi_1 \land \phi_2 \mid B_i \phi \mid [\mathbb{I} \phi_1] \phi_2 \mid [\mathbb{I} \phi_1] \phi_2 \mid [\mathbb{I} \phi_1] \phi_2 \]

Interpretation as sketched above:
9.2. The Logic of Lying in Public Discourse

- \([\mathcal{I}\phi]\psi\) is true in a model \(M\) at a world \(w\) if \(\psi\) is true in both \((w, 0)\) and \((w, 1)\) of the updated model \(M \otimes U\).

- \([i\phi]\psi\) is true in a model \(M\) at a world \(w\) if \(\psi\) is true in \((w, 0)\) of the updated model \(M \otimes U^0\).

- \([!\phi]\psi\) is true in a model \(M\) at a world \(w\) if \(\psi\) is true in \((w, 1)\) of the updated model \(M \otimes U^1\).

Now it turns out that the logic of individual belief and manipulative update has a simple axiomatisation in terms of reduction axioms, just like the logic of individual knowledge and public announcement. These reduction axioms are as follows. I start out with the reduction axioms for the \([\mathcal{I}\phi]\) modality:

\[
[\mathcal{I}\phi]\psi \leftrightarrow [i\phi]\psi \land [!\phi]\psi
\]

This defines the effect of \([\mathcal{I}\phi]\) in terms of those of \([!\phi]\) and \([i\phi]\). Next, there are the usual reduction axioms for public announcement:

\[
[!\phi]p \leftrightarrow \phi \rightarrow p
\]

\[
[!\phi]\neg \psi \leftrightarrow \phi \rightarrow \neg [!\phi]\psi
\]

\[
[!\phi](\psi_1 \land \psi_2) \leftrightarrow [!\phi]\psi_1 \land [!\phi]\psi_2
\]

\[
[!\phi]B_i\psi \leftrightarrow \phi \rightarrow B_i[!\phi]\psi
\]

Finally, the reduction axioms for lying:

\[
[i\phi]p \leftrightarrow \neg \phi \rightarrow p
\]

\[
[i\phi]\neg \psi \leftrightarrow \neg \phi \rightarrow \neg [i\phi]\psi
\]

\[
[i\phi](\psi_1 \land \psi_2) \leftrightarrow [i\phi]\psi_1 \land [i\phi]\psi_2
\]

\[
[i\phi]B_i\psi \leftrightarrow \neg \phi \rightarrow B_i[i\phi]\psi
\]

The final axiom of this list is the most interesting: it expresses that believing \(\psi\) after a lie that \(\phi\) amounts to the belief that a public announcement of \(\phi\) implies \(\psi\), conditioned by \(\neg \phi\).

Since all these axioms have the form of equivalences, completeness of the calculus of manipulation and individual belief follows from a reduction argument, as in the case of public announcements with individual knowledge. I refer to [van Benthem et al., 2006] for a general perspective on proving communication logics complete by means of reduction axioms.

9.2.1. THEOREM. The calculus of manipulation and individual belief is complete for the class of the (multi-)modal KD45 models.
Another way to see that the logic is complete is by means of the observation that this is the special case of the Logic of Communication and Change (LCC, [van Benthem et al., 2006]) where updates are restricted to manipulations, announcements and lies, and where doxastic programs are restricted to individual accessibilities.

Interestingly, my logic of manipulation is closely related to the variation on public announcement that is used in [Gerbrandy, 2007, Kooi, 2007] (and going back to [Gerbrandy, 1999]) to analyze the ‘surprise exam puzzle’, where public announcement of \( \phi \) is defined as an operation that restricts the doxastic alternatives of the agents to the worlds where \( \phi \) is true, i.e., all relations to \( \neg \phi \) worlds are destroyed. Using \( \dagger \phi \) for this alternative announcement, the corresponding reduction axiom is \( [\dagger \phi] B_i \psi \leftrightarrow B_i (\phi \rightarrow [\dagger \phi] \psi) \).

A forerunner of this logic is the analysis of suspicions and lies in [Baltag, 2002], which is further elaborated in [Baltag and Smets, 2008] and [van Ditmarsch, 2008]; the latter (actually a follow-up of the first version of the paper, [van Ditmarsch et al., 2012], on which this chapter was based) addresses more agency aspects in lying, such as the assumption that the addressee does not yet (firmly) believe the opposite of the lie - you don’t want to be caught out as a liar!

At first sight, this alternative semantics for announcement takes me outside of the framework sketched above. However, if \( \dagger \phi \) is an alternative announcement, then I have:

9.2.2. Proposition. \( M, w \models [\dagger \phi] \psi \text{ iff } M, w \models [\ddagger \phi] \psi \).

Alternative announcement turns out to be the same as manipulative updating, and this analysis can be viewed as a decomposition of alternative announcement into public lying and (regular) public announcement.

Regular public announcements can be expressed in terms of manipulative updating:

9.2.3. Proposition. \( \vdash [\! \phi] \psi \leftrightarrow (\phi \rightarrow [\ddagger \phi] \psi) \).

The proof is by induction on \( \psi \) and is left to the reader.

The logic of public announcement and the logic of manipulation have the same expressive power: this follows from the fact that they both reduce to multimodal KD45. But note that the logic of manipulative updating has greater ‘action expressivity’ than the logic of public announcement: the logic of \( [\! \phi] \) has no means to express an operation mapping S5 models to KD45 models, and \( [\ddagger \phi] \) is such an operation.

As an example of reasoning with the calculus, I use the axioms to show that a manipulative update followed by a belief is equivalent to a belief followed by the corresponding public announcement:

9.2.4. Proposition. \( \vdash [\ddagger \phi] B_i \psi \leftrightarrow B_i [\! \phi] \psi \).
9.2. The Logic of Lying in Public Discourse

**Proof.**

\[
\begin{align*}
[†φ]B_iψ & \leftrightarrow ([†φ]B_iψ \land ![φ]B_iψ) \\
& \leftrightarrow ((¬φ \rightarrow B_i[![φ]ψ]) \land (φ \rightarrow B_i[![φ]ψ])) \\
& \leftrightarrow B_i[![φ]ψ].
\end{align*}
\]

□

An important difference between manipulative update and public announcement shows up when I work out the preconditions of inconsistency after an update. For public announcements I get:

**9.2.5. Proposition.** \(\vdash ![φ]\bot \leftrightarrow ¬φ.\)

**Proof.**

\[
![φ]\bot \leftrightarrow ![φ](p \land ¬p) \\
\leftrightarrow ([![φ]p \land ![φ]¬p) \\
\leftrightarrow ([![φ]p \land (φ \rightarrow ¬[![φ]p])) \\
\leftrightarrow ((φ \rightarrow p) \land (φ \rightarrow ¬p)) \\
\leftrightarrow ¬φ
\]

□

This shows that a public announcement with \(φ\) leads to an inconsistent state iff the negation of \(φ\) is true. Similarly, it is easy to work out that a public lie that \(φ\) leads to an inconsistency iff \(φ\) is true, i.e., I can derive

**9.2.6. Proposition.** \(\vdash [‡φ]\bot \leftrightarrow φ.\)

Using these propositions I can work out the preconditions for inconsistency after a manipulative update:

**9.2.7. Proposition.** \(\vdash [‡φ]\bot \leftrightarrow \bot.\)

**Proof.**

\[
[‡φ] \leftrightarrow (![φ]\bot \land [iφ]\bot) \\
\leftrightarrow (¬φ \land φ) \\
\leftrightarrow \bot
\]

□

This means that a manipulative update in a consistent state will never lead to inconsistency (although, of course, it may lead to an agent having an inconsistent set of beliefs, which is different).

The following proposition about public announcements can be proved by induction on \(φ\). It shows that if one updates with an inconsistency, the resulting model is inconsistent:
9.2.8. Proposition. \( \vdash [!\bot] \phi \leftrightarrow \top \).

In the case of manipulatively updating with an inconsistency, the result is not an inconsistent model, but a model where all accessibilities have vanished. In the particular case of the belief of agent \( a \), this gives:

9.2.9. Proposition. \( \vdash [‡\bot] B_a \phi \leftrightarrow \top \).

Proof. 
\[
\begin{align*}
[‡\bot] B_a \phi & \iff ([!\bot] B_a \phi \land [\bot] B_a \phi) \\
& \iff (\top \land B_a [!\bot] \phi) \\
& \iff B_a [!\bot] \phi \\
& \iff B_a \top \\
& \iff \top.
\end{align*}
\]

After a manipulative update with an inconsistency, the public will no longer be able to distinguish what is false from what is true.

Finally, the following proposition spells out under what conditions our ‘sense by which we take our bearings in the real world’ is destroyed. This happens exactly when we are manipulated into accepting as truth what flatly contradicts our firm belief:

9.2.10. Proposition. \( \vdash [‡\phi] B_i \bot \leftrightarrow B_i \neg \phi \).

Proof. 
\[
\begin{align*}
[‡\phi] B_i \bot & \iff ([!\phi] B_i \bot \land [\phi] B_i \bot) \\
& \iff ((\phi \rightarrow B_i [!\phi] \bot) \land (\neg \phi \rightarrow B_i [\phi] \bot)) \\
& \iff ((\phi \rightarrow B_i \neg \phi) \land (\neg \phi \rightarrow B_i \neg \phi)) \\
& \iff B_i \neg \phi.
\end{align*}
\]

I can generalize my logic to a full logic of manipulative updating, i.e., according to the full relational action description in the Logic of Communication and Change. For details, see Section 9.6.

In this section I have investigated the effect of lying in public discourse. In such a setting the agents assume that they are told the truth and in the event of a lie, the agents hearing the lie do not believe that the announcement is actually a lie. This causes them to believe a false thing. In Section 9.4 I will analyze lying in a different setting, where the agents are playing a game of Liar’s Dice and following a game strategy. But first, I will give a game-theoretical analysis of the game to see how lying affects a game’s outcome.
9.3 Liar’s Dice — Game-Theoretical Analysis

In his later years as a saint, St. Augustine held the opinion that lying, even in jest, is wrong, but as the young and playful sinner that he was before his turn to seriousness he may well have enjoyed an occasional game of dice. I will examine a simplified version of two-person Liar’s Dice, and show by means of a game-theoretical analysis that it is precisely the possibility of lying - using private information in order to mislead an opponent - that makes the game interesting.

In my simplified version of Liar’s Dice, the die is replaced by a coin. A typical move of the game is tossing a coin and inspecting the result while keeping it hidden from the other player. Here is a description of what goes on, and what the options of the two players are.

- Players $a$ and $b$ both stake one euro: Player $a$ bets on heads, Player $b$ bets on tails.
- Player $a$ tosses a coin under a cup and observes the outcome (heads or tails), while keeping it concealed from player $b$.
- Player $a$ announces either $\text{Head}$ or $\text{Tail}$.
- If $a$ announces $\text{Tail}$, then she simply loses her one euro to player $b$ and game ends (for $a$ bets on heads, so she announces defeat).
- If $a$ announces $\text{Head}$, she adds one euro to the stake and the game continues.
- In response to $\text{Head}$, $b$ either passes (gives up) or challenges “I don’t believe that, you liar”) and adds 1 euro to the stake.
- If $b$ passes, $a$ wins the stake, and the game ends.
- If $b$ challenges, and the toss was heads, $a$ wins the stake, otherwise $b$ wins the stake. The game ends.

Player $a$ has two information states: Heads and Tails, while player $b$ has a single information state, for player $b$ cannot distinguish the two possible outcomes of the toss. I will give a game-theoretic analysis of how player $a$ can exploit her ‘information advantage’ to the utmost, and of how player $b$ can react to minimize her losses, on the assumption that the procedure is repeated a large number of times. The following picture gives the extensive game form. The first move is made by Chance; this move gives the outcome of the coin toss. Then player $a$ reacts, letting her move depend on the toss outcome. Finally, player $b$ decides whether to pass or challenge. This decision does not depend on the coin toss; player $b$ cannot distinguish the state where $a$ announced $\text{Head}$ after seeing heads
Chapter 9. The Logic of Lying

Figure 9.1: Extensive game form for Liar’s Dice game.

from the state where she is bluffing. In the picture of the extensive game form (Figure 9.1) this is expressed by a dotted line.

The leaves of the game tree indicate the payoffs. If the game sequence is Heads, NTail, the payoffs are $-1$ euro for player $a$ and 1 euro for player $b$. The same for the sequence Tails, NTail. Player $a$ gets 1 euro and player $b$ gets $-1$ euro for the sequences Heads, NHead, Pass, and Tail, NHead, Pass (these are the sequences where 2 gives up). The sequence Heads, NHead, Challenge is a win for player $a$, with payoff 2 euros, and $-2$ euros for player $b$. The sequence Tails, NHead, Challenge, finally, is a win for player $b$, with payoff 2 euros, and $-2$ euros for player $a$.

Player $a$ has four strategies: (NHead, NHead) (NHead in case of heads and in case of tails), (NHead, NTail) (NHead in case of heads, NTail in case of tails), (NTail, NHead), and (NTail,NTail). Player $b$ has two strategies: Pass and Challenge. To find the strategic game form, one has to take the average of the expected payoffs for the two cases of heads and tails. E.g., if player $a$ plays (NHead, NTail) and player $b$ responds with Challenge, then in the long run in $\frac{1}{2}$ of the cases the outcome will be heads, and player $a$ wins 2 euros, and in $\frac{1}{2}$ of the cases the outcome will be tails, and player $a$ loses 1 euro. Thus, the expected payoff is $\frac{1}{2} \times 2 - \frac{1}{2} \times 1 = \frac{1}{2}$ euro for player $a$, and because the game is zero sum, $-\frac{1}{2}$ euro for player $b$. The strategic game form is given by:

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHead, NHead</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
<tr>
<td>NHead, NTail</td>
<td>0,0</td>
<td>$\frac{1}{2}$, $-\frac{1}{2}$</td>
</tr>
<tr>
<td>NTail, NHead</td>
<td>0,0</td>
<td>$-\frac{3}{2}$, $\frac{3}{2}$</td>
</tr>
<tr>
<td>NTail, NTail</td>
<td>-1,1</td>
<td>-1,1</td>
</tr>
</tbody>
</table>
It is easy to see that there is no pure strategy Nash equilibrium. A Nash equilibrium is a combination of strategies, one for each player, with the property that neither of the players can improve their payoff by unilaterally deviating from her strategy (see, e.g., [Osborne and Rubinstein, 1992]). Clearly, none of the eight strategy pairs has this property.

Now let’s consider the strategy $(\mathbb{N} \text{Tail}, \mathbb{N} \text{Tail})$ for $a$. This is the strategy of the doomed loser: even when the toss is heads the player still announces $\mathbb{N} \text{Tail}$. This is obviously not the best thing that $a$ can do. Always announcing $\mathbb{N} \text{Head}$ gives a much better payoff in the long run. In other words, the strategy $(\mathbb{N} \text{Tail}, \mathbb{N} \text{Tail})$ is strictly dominated by $(\mathbb{N} \text{Head}, \mathbb{N} \text{Head})$. Similar for the strategy of the unconditional liar: $(\mathbb{N} \text{Tail}, \mathbb{N} \text{Head})$. It is also strictly dominated by the strategy $(\mathbb{N} \text{Head}, \mathbb{N} \text{Head})$. Thus, I am left with:

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{N} \text{Head}, \mathbb{N} \text{Head}$</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
<tr>
<td>$\mathbb{N} \text{Head}, \mathbb{N} \text{Tail}$</td>
<td>0,0</td>
<td>$\frac{1}{2}, -\frac{1}{2}$</td>
</tr>
</tbody>
</table>

Suppose $a$ plays $(\mathbb{N} \text{Head}, \mathbb{N} \text{Head})$ with probability $p$ and $(\mathbb{N} \text{Head}, \mathbb{N} \text{Tail})$ with probability $1 - p$. Then her expected value is $p$ for her first strategy, and $\frac{1}{2}(1 - p)$ for her second strategy. Any choice of $p$ where the expected payoff for $p$ is different from that for $1 - p$ can be exploited by the other player. Therefore, player $a$ should play her first strategy with probability $p = \frac{1}{3}(1 - p)$, i.e., $p = \frac{1}{3}$, and her second strategy with probability $1 - p = \frac{2}{3}$. For player $b$, I can reason similarly. Suppose $b$ plays Pass with probability $q$ and Challenge with probability $1 - q$. Again, the expected values for $q$ and $1 - q$ should be the same, for otherwise this mixed strategy can be exploited by the other player. The expected value is $-q$ for her first strategy and $-\frac{1}{2}(1 - q)$ for her second strategy. Thus, she should play her first strategy with probability $q = \frac{1}{3}(1 - q)$, i.e., $q = \frac{1}{3}$. Neither player can improve on her payoff by unilateral deviation from these strategies, so the mixed strategy where $a$ plays $(\mathbb{N} \text{Head}, \mathbb{N} \text{Head})$ in $\frac{1}{3}$ of the cases and $b$ plays Pass in $\frac{1}{3}$ of the cases is a Nash equilibrium. In other words, the best thing that player $a$ can do is always announcing the truth and raising the stakes when her toss is heads, and lying in one third of the cases when her toss is tails, and $b$’s best response to this is to Pass in one third of all cases and Challenge two thirds of the time.

The game-theoretic analysis yields that lying pays off for player $a$, and that player $b$, knowing this, may reasonably expect to catch player $a$ on a lie in one sixth of all cases. The value of the game is $\frac{1}{3}$ euro, and the solution is $\frac{1}{3}$ ($(\mathbb{N} \text{Head}, \mathbb{N} \text{Head})$, $\frac{2}{3}$ ($(\mathbb{N} \text{Head}, \mathbb{N} \text{Tail})$) as player $a$’s optimal strategy, and $\frac{1}{3}$ Pass, $\frac{2}{3}$ Challenge as player $b$’s optimal strategy. It is clear that the honest strategy $(\mathbb{N} \text{Head}, \mathbb{N} \text{Tail})$ is not the optimal one for player $a$: given that player $b$ plays $\frac{1}{3}$ Pass and $\frac{2}{3}$ Challenge, the expected payoff for player $a$ is only $\frac{1}{6}$ if she sticks to the honest strategy. Lying indeed pays off sometimes.
If I modify the game so that player $a$ cannot lie anymore, by refusing her the privilege of having a peek at the toss outcome, the game immediately becomes a lot less interesting. In the extensive game form for this version, an extra dotted line indicates that player $a$ cannot distinguish the outcome Heads from the outcome Tails. See Figure 9.2.

![Figure 9.2: Modified game where player $a$ has no information advantage.](image)

Player $a$ has just two strategies left, $\text{HHead}$ and $\text{HTail}$, and the strategic form of the game becomes:

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{HHead}$</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
<tr>
<td>$\text{HTail}$</td>
<td>-1,1</td>
<td>-1,1</td>
</tr>
</tbody>
</table>

The strategy $\text{HTail}$ for player $a$ is weakly dominated by $\text{HHead}$, so it can be eliminated, and we are left with:

<table>
<thead>
<tr>
<th></th>
<th>Pass</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{HHead}$</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

The strategy pair ($\text{HHead}$, Challenge) is a Nash equilibrium. The game-theoretic analysis predicts that a rational player $a$ will always play $\text{HHead}$, and a rational player $b$ will always Challenge, and the game becomes a pure zero-sum game of chance. Surely, it is the possibility of lying that makes Liar’s Dice an interesting game.

### 9.4 Liar’s Dice — Doxastic Analysis

In the game of Liar’s Dice, when player $a$ announces Heads while she actually saw that the outcome of the toss was Tails, she is announcing something which
she believes to be false with the intent to be believed. This certainly seems to be a lie. However, we usually do not condemn people who tell such a lie in a game as untruthful. In fact, in this game player $a$ is supposed to lie sometimes, or she would never win. This is an important point: player $a$ intends player $b$ to believe her, but she probably does not expect it, because player $b$ may very well expect player $a$ to lie sometimes. As I have already shown, it is completely immaterial in Liar’s Dice whether an announcement is true or false: the only reasons for one or the other are strategic, and in view of winning the game. In this section I will analyze the game of Liar’s Dice from a doxastic viewpoint in order to answer the question: is lying really lying, when one is actually supposed to lie?

For my analysis I will use the doxastic model checker DEMO [van Eijck, 2007]. Using DEMO, I can automatically check the truth of formulas in a doxastic model. I have extended DEMO with factual changes to allow action models with substitutions and also with the possibility to store integer values in my Bachelor’s Thesis [Sietsma, 2007]. I will use this extended model checker. The code of this model checker is available from http://www.cwi.nl/~jve/software/demolight0/. I show how the game of Liar’s Dice can be modeled using DEMO, and I demonstrate the doxastic models that I get if I trace a particular run of the game. For full details, see Section 9.7.

The conclusion of this analysis is that, even though in the game of Liar’s Dice lying takes place according to the definition of Augustine, no misleading is taking place and the players are never duped into believing a falsehood. This is shown by the fact that all updates in the games, as modeled in the Appendix, are S5 updates: instead of unquestioningly taking for granted what they are being told, all players consider the opposite of what they are being told equally likely. In the resulting models there are no false beliefs, only true knowledge.

9.5 Conclusion

First of all, I will compare the approach presented here to that of Chapter 8. There, the only constraint on the basic relations is that they are linked and from these basic relations four different notions of belief are constructed using PDL. Here, all relations satisfy the KD45 axioms and I only use one notion of belief. The notion used here is probably closest to the notion of strong belief discussed there, although the relations in my model do not need to be reflexive while strong belief is constructed as the reflexive transitive closure of the basic relations. Using one single notion of belief allowed me to focus on the effects of lies on an agent’s belief. The update discussed here differs from the one proposed in Chapter 8 because it results in “stronger” belief of the formula that is communicated. This is appropriate for the interpretation as a lie that is believed by the agents who hear it. In Chapter 8 the agents’ relations represent preference or a “softer” form of belief, that allows for different levels of plausibility or preference. Such
an interpretation is more appropriate for the modeling of belief revision and judgment aggregation.

There are still two discrepancies that I have to address. The first one is between my treatment of lying in public discourse and my treatment of lying in games. As I have shown, lying in public discourse can lead to KD45 models, which illustrates the fact that genuine misleading takes place. I argued that the players in a game like Liar’s Dice are never actually misled, so in a sense no real lying takes place here at all. But one might also say that lying is attempted, but due to the smartness of the opponent, these attempts are never really believed. So lying in public discourse and lying in games are connected after all.

The difference between the two settings could be seen as a difference in the protocol the agents are following. In public discourse, the agents usually assume that they are following the protocol “only speak the truth”. Therefore, when one of them deviates from the protocol by telling a lie, the others believe him and are misled. In the game of Liar’s Dice, the protocol is “say anything in order to improve your payoff”. Since all agents know that the others are following the protocol, under the assumption of common knowledge of rationality, they do not believe each other’s lies. The issue of protocol dynamics in epistemic modeling is explored further in [Wang, 2010].

The second discrepancy is between the game-theoretical analysis of lying in games in terms of mixed strategies that use probabilities, and the logical analysis in terms of truth values. To see that these perspectives still do not quite match, consider the game situation where player $a$ tosses the coin, observes the result, and announces ‘heads’. In my logical analysis this does not lead to the false belief of player $b$ that the coin has landed heads; it does not lead to a belief change at all. But the game-theoretical analysis reveals that a rational agent would have formed a belief about the probability that the claim is true. So it seems that the logical analysis is still too crude.

This defect could be remedied by using probabilistic beliefs and probabilistic updates, in the style of [van Benthem et al., 2009b], which would allow me to express the probability of actions in the game. With these, one can model the fact that the game-theoretical analysis in terms of mixed strategies is common knowledge. For if this is the case, it is common knowledge that if the toss is tails, then player $a$ will announce ‘heads’ with probability $\frac{1}{3}$ and ‘tails’ with probability $\frac{2}{3}$.

Interestingly, this is also relevant for the first discrepancy. For why are the players not duped into believing falsehoods, in the game of Liar’s Dice? Because they look further than a single run of the game, and they know that as the game gets repeated they can adhere to mixed strategies. Therefore, an analysis in terms of manipulative probabilistic updates might work for both lying in public discourse and lying in games.
9.6 Appendix: The Full Logic of Manipulative Updating

The full logic of manipulative updating extends the logic of lies and individual beliefs from Section 9.2 to doxastic PDL. It consists of doxastic PDL extended with manipulative updates, lies and announcements:

\[
\begin{align*}
\alpha & ::= i | ?\phi; \alpha_1; \alpha_2 | \alpha_1 \cup \alpha_2 | \alpha^* \\
\phi & ::= p | \neg \phi | \phi_1 \land \phi_2 | [\alpha] \phi | [\‡ \phi_1] \phi_2 | [i \phi_1] \phi_2 | [! \phi_1] \phi_2
\end{align*}
\]

There is a complete axiomatisation: the axioms and rules of PDL, the axioms of KD45, necessitation for \[\‡ \phi\], \[i \phi\], \[! \phi\], and the following reduction axioms for the three update modalities.

The definition of \[\‡ \phi\] in terms of \[i \phi\] and \[! \phi\] is as in Section 9.2:

\[ [\‡ \phi] \psi \leftrightarrow [i \phi] \psi \land [! \phi] \psi \]

Reduction axioms for public announcement are as follows:

\[
\begin{align*}
[! \phi] p & \leftrightarrow \phi \rightarrow p \\
[! \phi] \neg \psi & \leftrightarrow \phi \rightarrow \neg [! \phi] \psi \\
[! \phi] (\psi_1 \land \psi_2) & \leftrightarrow [! \phi] \psi_1 \land [! \phi] \psi_2 \\
[! \phi] [a] \psi & \leftrightarrow [? \phi; a] [! \phi] \psi \\
[! \phi] [? \chi] \psi & \leftrightarrow [? \phi; ? \chi] [! \phi] \psi \\
[! \phi] [\alpha_1; \alpha_2] \psi & \leftrightarrow [! \phi] [\alpha_1] [\alpha_2] \psi \\
[! \phi] [\alpha_1 \cup \alpha_2] \psi & \leftrightarrow [! \phi] ([\alpha_1] \psi \land [\alpha_2] \psi) \\
[! \phi] [\alpha^*] \psi & \leftrightarrow [\alpha^*] [! \phi] \psi
\end{align*}
\]

where \(\alpha^*\) such that \([! \phi] [\alpha] \psi \leftrightarrow [\alpha^*] [! \phi] \psi\).

It can be shown by an inductive argument that for every doxastic program \(\alpha\), every announcement \(! \phi\), and every postcondition \(\psi\) a doxastic program \(\alpha'\) exists such that \([! \phi] [\alpha] \psi \leftrightarrow [\alpha'] [! \phi] \psi\). This \(\alpha'\), which does not have to be unique, can be found by applying the above reduction axioms.
Reduction axioms for public lies:

\[
[!\phi]p \leftrightarrow \neg\phi \rightarrow p
\]
\[
[!\phi]\neg\psi \leftrightarrow \neg\phi \rightarrow [!\phi]\psi
\]
\[
[!\phi](\psi_1 \land \psi_2) \leftrightarrow [!\phi]\psi_1 \land [!\phi]\psi_2
\]
\[
[!\phi][a]\psi \leftrightarrow [?\neg\phi; a][!\phi]\psi
\]
\[
[!\phi][?\chi]\psi \leftrightarrow [?\neg\phi; ?\chi][!\phi]\psi
\]
\[
[!\phi][\alpha_1; \alpha_2]\psi \leftrightarrow [!\phi][\alpha_1][\alpha_2]\psi
\]
\[
[!\phi][\alpha_1 \cup \alpha_2]\psi \leftrightarrow [!\phi][\alpha_1]\psi \land [\alpha_2]\psi
\]
\[
[!\phi][\alpha^*]\psi \leftrightarrow [\alpha'; \alpha''^*][!\phi]\psi
\]

where \(\alpha'\) such that \([!\phi][\alpha]\psi \leftrightarrow [\alpha'][!\phi]\psi\)

and \(\alpha''\) such that \([!\phi][\alpha]\psi \leftrightarrow [\alpha''][!\phi]\psi\).

Again, it can be shown by an inductive argument that for every doxastic program \(\alpha\), every lie \([!\phi]\), and every postcondition \(\psi\), a doxastic programs \(\alpha'\) exists such that

\([!\phi][\alpha]\psi \leftrightarrow [\alpha'][!\phi]\psi\).

The \(\alpha'\) in the axioms for \(\alpha^*\) can be viewed as the transformed versions of the programs \(\alpha\), where the update operator acts as a doxastic program transformer. To give an example, suppose \(\alpha = a \cup b\), and I want to calculate the way common belief of \(a\) and \(b\) is transformed by a public lie that \(\phi\). Then the transformed program for \(a \cup b\) becomes \(?\neg\phi; a \cup b\), i.e., I have:

\([!\phi][a \cup b]\psi \leftrightarrow [?\neg\phi; a \cup b][!\phi]\psi\).

Similarly for the way common belief of \(a\) and \(b\) is transformed by a public announcement: the transformed program for \(a \cup b\) becomes \(?\phi; a \cup b\), and I have:

\([!\phi][a \cup b]\psi \leftrightarrow [?\phi; a \cup b][!\phi]\psi\).

Using these transformed programs, one can see that the reduction axiom for \((a \cup b)^*\) takes the shape:

\([!\phi][((a \cup b)^*)\psi \leftrightarrow [?\neg\phi; a \cup b; (?\phi; a \cup b)^*][!\phi]\psi]\).

This expresses that after a lie with \(\phi\), \(a\) and \(b\) have a common belief that \(\psi\) iff in the model before the lie it holds that along all \(a \cup b\) paths that start from a \(\neg\phi\) world and that pass only through \(\phi\) worlds, \([!\phi]\psi\) is true. Note that this is a 'relativized common belief' similar to the relativized common knowledge that is needed to get a reduction style analysis going of public announcement in the presence of common knowledge.

In fact, the style of axiomatisation that I have adopted is borrowed from the reduction axioms formulated in terms of program transformations, in [van Benthem et al., 2006]. In the same manner as in [van Benthem et al., 2006] I can derive (with the restriction to multi-K models, not to multi-KD45 models):

9.6.1. Theorem. The calculus of manipulative updating is complete.
9.7 Appendix: Liar’s Dice in DEMO

First I will closely examine the different actions that take place in the game and their representations as action models. Let $p$ represent the value of a coin, with 1 signifying heads, and 0 signifying tails. Let agents $a$ and $b$ represent the two players, and let $C_1$ represent the contents of the purse of player $a$ ($C$ for cash), and $C_2$ that of player $b$, with natural number values representing the amounts in euros that each player has in her purse. These natural number registers are available in the new extension of DEMO that was presented in [Sietsma, 2007]. Let $S_1, S_2$ represent the money at stake for each player. Factual change can be thought of as assignment of new values to variables. This is an essential ingredient of the various actions in the game:

**Initialisation** Both players put one euro at stake, and they both know this. $S_1 := 1, C_1 := C_1 - 1, S_2 := 1, C_2 := C_2 - 1$, together with public announcement of these factual changes.

**Heads** Factual change of the propositional value of a coin $p$ to 1, with private communication of the result to player $a$ ($p = 1$ signifies heads).

**Tails** Factual change of the propositional value of a coin $p$ to 0, with private communication of the result to player $a$. ($p = 0$ signifies tails).

**Announce** Player $a$ announces either $\&Head$ or $\&Tail$. There are several ways to model this and I will come back to this later.

**Pass** Player $b$ passes and loses, player $a$ gets the stakes. $C_1 := C_1 + S_1 + S_2, S_1 := 0, S_2 := 0$.

**Challenge** Public setting of $C_2 := C_2 - 1, S_2 := S_2 + 1$, followed by public announcement of the value of $p$. If the outcome is $p$ then $C_1 := C_1 + S_1 + S_2$, otherwise $C_2 := C_2 + S_1 + S_2$ and in any case $S_1 := 0, S_2 := 0$.

I will show how these actions can be defined as doxastic action models in Haskell code using DEMO.

```haskell
module Lies
where
import ModelsVocab hiding (m0)
import ActionVocab hiding (upd,public,preconditions,
    vocProp,vocReg)
import ChangeVocab
import ChangePerception
import Data.Set (Set)
import qualified Data.Set as Set
```
I first define the cash and stakes of each player as integer registers.

\[
\begin{align*}
c_1, c_2, s_1, s_2 :: & \text{Reg} \\
c_1 = (\text{Rg 1}); & c_2 = (\text{Rg 2}) \\
s_1 = (\text{Rg 3}); & s_2 = (\text{Rg 4})
\end{align*}
\]

This declares four integer registers, and gives them appropriate names. The initial contents of the purses of the two players must also be defined. Let us assume both players have five euros in cash to start with.

\[
\begin{align*}
\text{initCash1, initCash2 ::} & \text{Int} \\
\text{initCash1} = 5 & \\
\text{initCash2} = 5
\end{align*}
\]

Initialisation of the game: both players put one euro at stake. This is modeled by the following factual change: \( S_1 := 1, C_1 := C_1 - 1, S_2 := 1, C_2 := C_2 - 1 \). Representing this in my modeling language is straightforward. I just represent the contents of the registers at startup.

\[
\begin{align*}
\text{initGame :: EM} \\
\text{initGame} = (\text{Mo} \\
[0] \\
[a,b] \\
[\text{}]) \\
[s_1, s_2, c_1, c_2] \\
[(0,[])] \\
\quad [(0,[s_1,1),(s_2,1), \\
(c_1,\text{(initCash1-1)}),(c_2,\text{(initCash2-1)}))] \\
\quad [(a,0,0),(b,0,0)] \\
\quad [0])
\end{align*}
\]

Tossing the coin is a factual change of \( p \) to 0 or 1. The coin is tossed secretly and before player \( a \) looks both players are unaware of the value of the coin. Therefore there are two worlds, one where \( p \) is set to 0 and one where \( p \) is set to 1, and neither of the two players can distinguish these worlds.
9.7. Appendix: Liar’s Dice in DEMO

\[
toss :: \text{Integer} \rightarrow \text{FACM State}
toss c \text{ ags} = (\text{Acm [0,1] ags})
\]

\[
[(0, (\text{Top}, ([([P 0, \neg \text{Top}]), []]), (1, (\text{Top}, ([([P 0, \text{Top}]), []])))]

[(ag, w, w') | w \leftarrow [0,1], w' \leftarrow [0,1], ag \leftarrow \text{ags}]
\]
\]

Note that the action model has a list that assigns to each world a precondition, a change to the propositions, and a change to the registers. In world 0, the precondition is \(\top\) and the change is to set \(p\) to value \(\neg\top\), i.e., \(\bot\) (and there is no change to the registers), and in world 1, the precondition is again \(\top\) and the change is to set \(p\) to value \(\top\) (and again, there is no change to the registers).

After the coin has been tossed player \(a\) looks under the cup without showing the coin to player \(b\). I define a generic function for computing the model of the action where a group of agents looks under the cup. These models consist of two worlds, one where \(p\) is true (heads) and one where \(p\) is false (tails), the agents in the group can distinguish these two worlds and the other agents cannot.

\[
look :: [\text{Agent}] \rightarrow \text{FACM State}
look \text{ group ags} = (\text{Acm [0,1] ags})
\]

\[
[(0, (p, (\text{[]}, [[]])), (1, (\neg p), ([[]], [[]])))]

\[
[(ag, w, w') | w \leftarrow [0,1], w' \leftarrow [0,1],
\quad ag \leftarrow \text{ags}, \text{notElem ag group}] ++
\quad [(ag, w, w) | w \leftarrow [0,1], ag \leftarrow \text{group}]
\]
\]

In this case, there are no changes to propositions or registers, but world 0 has precondition \(p\), and world 1 has precondition \(\neg p\).

Now I define the models of the situation after the coin has been tossed and player \(a\) has looked at the outcome, distinguishing the two outcomes of the toss:
headsg :: EM
headsg = upd (upd initGame (toss 1)) (look [a])

tailsg :: EM
tailsg = upd (upd initGame (toss 0)) (look [a])

Before looking at the way to model the announcement of an outcome of the toss by player \(a\) I will first define the action models for passing and challenging.

When player \(b\) passes, the stakes are added to player \(a\)'s cash: \(C_2 := C_2 + S_1 + S_1, S_1 := 0, S_2 := 0\). Player \(b\) never gets to see the actual value of the coin so there are no changes in the knowledge of the agents about \(p\). The model for this has only one world that indicates the changes in the stakes and cash.

\[
pass :: FACM State
\]
\[
\begin{array}{l}
pass ags = (Acm [0] ags \\
\quad [(0, \langle \text{Top}, \langle [] , \\
\quad \quad \langle (s1, (I 0)) , \\
\quad \quad \quad (s2, (I 0)) , \\
\quad \quad \quad \quad (c1, \text{ASum [Reg c1, Reg s1, Reg s2]}) \rangle )]) \\
\quad [(ag, 0, 0) | ag <- ags] \\
\quad [0])
\end{array}
\]

Note that here for the first time there are changes of the registers.

When player \(b\) decides to challenge player \(a\), the cup is lifted and both players get to know the value of \(p\). Then the stakes are added to the cash of player \(a\) in case of heads and player \(b\) in case of tails, together with one extra euro from the cash of player \(b\) that player \(b\) added to the stakes while challenging player \(a\). So instead of \(S_2 := S_2 + 1, C_2 := C_2 - 1\) and after that \(C_1 := C_1 + S_1 + S_2\) in case of heads and \(C_2 := C_2 + S_1 + S_2\) in case of tails, I use \(C_1 := C_1 + S_1 + S_2 + 1, C_2 := C_2 - 1\) in case of heads and \(C_2 := C_2 + S_1 + S_2\) in case of tails. The action model for this has one world for the case of heads and one world for the case of tails. Both players can distinguish these worlds because the cup was lifted, and the stakes are divided differently in the two worlds.
challenge :: FACM State
challenge ags =
  Acm
  [0, 1]
ags
[(0, (Neg (p), ([],
   [(s1, (I 0)),
    (s2, (I 0)),
    (c2, ASum [Reg c2, Reg s1, Reg s2]]))]),
 (1, (p, ([],
   [(s1, (I 0)),
    (s2, (I 0)),
    (c2, ASum [Reg c2, I (-1)],
     (c1, ASum [Reg c1, Reg s1, Reg s2, I 1]]))])]
[(ag, w, w) | w <- [0, 1], ag <- ags]
[0, 1]

When player a announces $\aleph$Head or $\aleph$Tail the stakes change. In case of $\aleph$Head $C_1 := C_1 - 1, S_1 := S_1 + 1$ and in case of $\aleph$Tail $C_2 := C_2 + S_1 + S_2, S_1 := 0, S_2 := 0$.

announceStakes :: Integer -> FACM State
announceStakes 0 ags =
  Acm
  [0]
ags
[(0, (Top, ([], [(s1, (I 0)),
   (s2, (I 0)),
   (c2, ASum [Reg c2, Reg s1, Reg s2]]))])
[(ag, 0, 0) | ag <- ags]
[0]
announceStakes 1 ags =
  Acm
  [0]
ags
[(0, (Top, ([], [(s1, ASum [Reg s1, I 1]),
   (c1, ASum [Reg c1, I (-1)]])])
[(ag, 0, 0) | ag <- ags]
[0]
Now the only thing I have to decide is how I will model the announcement of $\aleph_{\text{Head}}$ or $\aleph_{\text{Tail}}$. Suppose I would use the manipulative update $\frac{1}{2}p$ or $\frac{1}{2}\neg p$ for this. This would imply that the other player believes the claims that are made.

I first define a generic function that computes the model for any manipulative update. This is the model with two worlds, one where the formula that is announced is true and one where it is false, and relations from the world where it is false to the world where it is true for the agents that believe the announcement.

```haskell
manipulative :: Form -> [Agent] -> FACM State
manipulative f group ags =
  (Acm
   [0,1] ags
   [(0,(Neg f,([],[])),(1,(f,([],[])))],
    [[(ag,w,w') | w <- [0,1], w' <- [0,1],
       ag <- ags, notElem ag group ] ++
     [(ag,w,1) | w <- [0,1], ag <- group ])
   [0,1])
```

Now when player $a$ announces $\aleph_{\text{Head}}$ or $\aleph_{\text{Tail}}$ two things happen: the manipulative update is made to player 2, and player 1 adds one euro to the stakes in case of $\aleph_{\text{Head}}$ or player $b$ wins the stakes in case of $\aleph_{\text{Tail}}$. I first model the manipulative update. In case of announcement of $\aleph_{\text{Head}}$ this is the manipulative update with $p$, otherwise it is the manipulative update with $\neg p$.

```haskell
announceManip :: Integer -> FACM State
announceManip c = manipulative (fct c) [b]
  where fct 0 = Neg (Prp (P 0))
         fct 1 = (Prp (P 0))
```

I can combine these action models in a function on doxastic models:

```haskell
announce' :: Integer -> EM -> EM
announce' c m =
  upd (upd m (announceManip c)) (announceStakes c)
```

Now I have a complete way to model any game of Liar’s Dice. However, though this way to model things seems correct, it is not. When I model player
a’s announcement with manipulative updates player b will actually believe player a’s announcement. I can use the model checker to show this:

*Lies> isTrue (announce’ 0 headsg) (K b (Neg p))  
True  
*Lies> isTrue (announce’ 0 tailsg) (K b (Neg p))  
True  
*Lies> isTrue (announce’ 1 headsg) (K b p)  
True  
*Lies> isTrue (announce’ 1 tailsg) (K b p)  
True

However, in a real game of Liar’s Dice player b knows that player a might very well be bluffing and she does not really believe player a’s claim at all. So to correctly model the game I should not use the manipulative update. When player a makes an announcement this does not even change player b’s knowledge and beliefs because player b does not believe player a.

So instead of the manipulative update I should only use the model for changing the stakes to model the announcement:

```haskell
announce :: Integer -> FACM State  
announce = announceStakes
```

Now player b does not know whether p is true, but she knows she doesn’t know:

```haskell
bKnows :: Form  
bKnows = Disj [(K b (Neg p)), (K b p)]
```

*Lies> isTrue (upd tailsg (announce 0)) bKnows  
False  
*Lies> isTrue (upd tailsg (announce 0)) (K b (Neg bKnows))  
True  
*Lies> isTrue (upd headsg (announce 0)) bKnows  
False  
*Lies> isTrue (upd headsg (announce 0)) (K b (Neg bKnows))  
True  
*Lies> isTrue (upd tailsg (announce 1)) bKnows  
False  
*Lies> isTrue (upd tailsg (announce 1)) (K b (Neg bKnows))
*Lies* > isTrue (upd headsg (announce 1)) bKnows
False
*Lies* > isTrue (upd headsg (announce 1)) (K b (Neg bKnows))
True

Note that since I did not use the manipulative update to model player a’s announcement the resulting models are still S5-models.

Lies> isS5Model (upd headsg (announce 1))
True
Lies> isS5Model (upd headsg (announce 0))
True
Lies> isS5Model (upd tailsg (announce 1))
True
Lies> isS5Model (upd tailsg (announce 0))
True

This means that no actual misleading is taking place at all! This is actually very plausible because player b knows that player a’s announcement might very well be false. This shows that lying only creates false belief if the person who lies is believed to be telling the truth.

Now I can use these action models to do a doxastic analysis of a game of Liar’s Dice. The different possible games are:

1. Player a tosses tails and announces \(\forall \text{Tail}\)
2. Player a tosses heads and announces \(\forall \text{Tail}\)
3. Player a tosses tails and announces \(\forall \text{Head}\) and player b passes
4. Player a tosses tails and announces \(\forall \text{Head}\) and player b challenges
5. Player a tosses heads and announces \(\forall \text{Head}\) and player b passes
6. Player a tosses heads and announces \(\forall \text{Head}\) and player b challenges

The models for these games are:

```
 game1, game2, game3, game4, game5, game6 :: EM
game1 = gsm (upd tailsg (announce 0))
game2 = gsm (upd headsg (announce 0))
game3 = gsm (upd (upd tailsg (announce 1)) pass)
game4 = gsm (upd (upd tailsg (announce 1)) challenge)
game5 = gsm (upd (upd headsg (announce 1)) pass)
game6 = gsm (upd (upd headsg (announce 1)) challenge)
```
9.7. Appendix: Liar’s Dice in DEMO

I will now consider these six different cases in turn.

Game 1 is the game where player 1 tosses tails and admits this.

In this case both players stake one euro and player b wins the stakes, so in the end player a lost one euro and player b won one euro. This can be checked with DEMO:

* Lies> isTrue game1 (Eq (Reg c1) (ASum [I initCash1, I (-1)]))
  True
* Lies> isTrue game1 (Eq (Reg c2) (ASum [I initCash2, I 1]))
  True

Player b does not get to know what the value of the coin was:

* Lies> isTrue game1 bKnows
  False

The model for game 1 is:

* Lies> displayS5 game1

  [0, 1]
  [p]
  [R1, R2, R3, R4]
  [(0, []), (1, [p])]
  [(0, [(R1, 4), (R2, 6), (R3, 0), (R4, 0)]),
   (1, [(R1, 4), (R2, 6), (R3, 0), (R4, 0)])]
  (a, [[0], [1]])
  (b, [[0, 1]])
  [0]

A picture of this model is below. There are two worlds, one where the toss was heads and one where it was tails. Player a can distinguish these worlds, but player b cannot because player b never got to see the coin. In both worlds the cash of player a is 4 and that of player b is 6 euros, because the division of the stakes does not depend on the value of the coin. Reflexive arrows are not shown.

Game 2 is the game where player a falsely announces \& Head. Just like in game 1, player a loses one euro and player b wins one euro, and player b does not get to know the value of the coin.
Chapter 9. The Logic of Lying

*Lies> isTrue game2 (Eq (Reg c1) (ASum [I initCash1,I (-1)]))
True
*Lies> isTrue game2 (Eq (Reg c2) (ASum [I initCash2,I 1]))
True
*Lies> isTrue game2 bKnows
False

The model for this game is almost the same as for game 1: the difference is that now the world where \( p \) is true is actual instead of the world where \( p \) is false.

*Lies> displayS5 game2
[0,1]
[p]
[R1,R2,R3,R4]
[(0,[]), (1,[p])]
[(0,[(R1,4),(R2,6),(R3,0),(R4,0)]),
 (1,[(R1,4),(R2,6),(R3,0),(R4,0)])]
(a,[[0],[1]])
(b,[[0,1]])
[1]

The picture of this model (reflexive arrows not shown) is:

![Diagram](https://via.placeholder.com/150)

The third game is the case where player \( a \) tosses tails but falsely announces \( \not{H}ead \) and player \( b \) passes. In this case player \( a \) stakes two euros and player \( b \) stakes one euro, and player \( a \) gets to keep the stakes, so the final payoff is that player \( a \) wins one euro and player \( b \) loses one euro:

*Lies> isTrue game3 (Eq (Reg c1) (ASum [I initCash1,I 1]))
True
*Lies> isTrue game3 (Eq (Reg c1) (ASum [I initCash1,I 1]))
True

Player \( b \) passes, so the cup is never lifted and player \( b \) does not know the value of the coin:

*Lies> isTrue game3 bKnows
False
The model for this game is:

*Lies> displayS5 game3
[0,1]
[p]
[R1,R2,R3,R4]
[(0,[]),(1,[p])]
[(0,[(R1,6),(R2,4),(R3,0),(R4,0)]),
 (1,[(R1,6),(R2,4),(R3,0),(R4,0)])]
(a,[[0],[1]])
(b,[[0,1]])
[0]

This model has the same two worlds as the models for game 1 and 2 except for the changes in the player’s cash.

In the fourth game, player a tosses tails but falsely announces \(\#Head\) and player b challenges player a. This means that both players stake one extra euro and then the cup is lifted and player b gets the stakes.

In this case player b does know the value of the coin:

*Lies> isTrue game4 bKnows
True

The payoffs are \(-2\) euros for player a and \(2\) euros for player b:

*Lies> isTrue game4 (Eq (Reg c1) (ASum [I initCash1,I (-2)]))
True
*Lies> isTrue game4 (Eq (Reg c1) (ASum [I initCash1,I (-2)]))
True

The model for this game is:

*Lies> displayS5 game4
[0]
[p]
[R1,R2,R3,R4]
[(0,[])]
[(0,[(R1,3),(R2,7),(R3,0),(R4,0)])]
(a,[[0]])
(b,[[0]])
[0]

This model has only one world because none of the players consider any other world possible. This is because both players know the values of the coin. In this world \(p\) is false (because the toss was tails), player a’s cash is 3 euros and player b’s cash is 7 euros. A picture of this model is below.
Chapter 9. The Logic of Lying

The fifth game is the game where player \( a \) tosses heads and truthfully announces this and player \( b \) passes. In this case the cup is not lifted so player \( b \) does not know the value of the coin again:

\[
\text{\textit{Lies}}\text{> isTrue game5 bKnows False}
\]

The payoffs are 1 for player \( a \) and \(-1\) for player \( b \):

\[
\text{\textit{Lies}}\text{> isTrue game5 (Eq (Reg c1) (ASum [I initCash1,I 1])) True}
\]

\[
\text{\textit{Lies}}\text{> isTrue game5 (Eq (Reg c2) (ASum [I initCash2,I (-1)])) True}
\]

The model for game 5 has two worlds again because player \( b \) does not know the value of the coin.

\[
\text{\textit{Lies}}\text{> displayS5 game5 [0,1] [p] [R1,R2,R3,R4] [(0,[]),(1,[p])] [(0,[(R1,6),(R2,4),(R3,0),(R4,0)]), (1,[(R1,6),(R2,4),(R3,0),(R4,0)])] (a,[[0],[1]]) (b,[[0,1]]) [1]}
\]

In game 6 player \( a \) tosses heads and truthfully announces this and player \( b \) challenges player \( a \). In this case both players add one extra euro to the stakes, the cup is lifted and player \( a \) gets to keep the stakes. The model for this has one world where \( p \) is true, player \( a \) has 7 euros and player \( b \) has 3 euros.

\[
\text{\textit{Lies}}\text{> displayS5 game6 [0] [p] [R1,R2,R3,R4]}
\]
In this case player b knows the value of the coin and the payoffs are 2 euros for player 1 and −2 euros for player 2:

* Lies> isTrue game6 bKnows
  True
* Lies> isTrue game6 (Eq (Reg c1) (ASum [I initCash1, I 2]))
  True
* Lies> isTrue game6 (Eq (Reg c2) (ASum [I initCash2, I (-2)]))
  True