The Ditmarsch Tale of Wonders — Dynamics of Lying

Hans van Ditmarsch

University of Sevilla, Sevilla, Spain, hvd@us.es

1 Introduction

My favourite of Grimm's fairytales is 'Hans im Glück' (Hans in luck). A close second comes 'The Ditmarsch Tale of Wonders'. In German this is called a 'Lügenmärchen', a 'Liar's Tale'. It contains the passage "A crab was chasing a hare which was running away at full speed; and high up on the roof lay a cow which had climbed up there. In that country the flies are as big as the goats are here." These are very obvious lies. Nobody considers it possible that this is true. Crabs are reputedly slow, hares are reputedly fast.

In the real world, if you lie, sometimes other people believe you and sometimes they don't. When can you get away with a lie? Consider the well-known consecutive numbers riddle (see [10], and the Appendix), where Anne has 2 and Bill has 3, and they only know that their natural numbers are one apart. Initially, Anne is uncertain between Bill having 3 or 1, and Bill is uncertain between Anne having 2 or 4. So both Anne and Bill do not initially know their number. Suppose Anne says to Bill: "I know your number." Anne is lying. Bill does not consider it possible that Anne knows his number, so he tells Anne that she is lying. However, Anne did not know that Bill would not believe her. She considered it possible that Bill had 1, in which case Bill would have considered it possible that Anne was telling the truth, and would then have drawn the incorrect conclusion that Anne had 0. I.e., if you are still following us... It seems not so clear how this should be formalized in a logic interpreted on epistemic modal structures, and this is the topic of our paper.

What is a lie? Let p be a Boolean proposition. You lie that p if you believe that $\neg p$ while you say that p and with the intention that the addressee believes p. This definition seems standard since Augustine [12]. A *believed lie* therefore is one that, when told, is believed by the addressee to be truthful. We abstract from the intentional aspect and model the believed lie. (Similarly, in AGM belief revision, we incorporating new information, abstracting from the process that made it acceptable.)

What are the modal preconditions and postconditions of a lie? Let i be the speaker (assumed female) and let j be the addressee (assumed male). Then the precondition of 'i is lying that p to j' is $B_i\neg p$, and the postcondition is B_jp. Also, the precondition should be preserved. More refined preconditions are conceivable, e.g., that the addressee consider it possible that the lie is true, or believes that the speaker knows the truth about p. Those are plausible additional conditions rather than rock-bottom requirements. Concerning the postcondition: the liar does not merely intend the speaker to believe p, but also wants him to believe that the speaker believes p. It is obvious that the postcondition should not be merely $B_j p$, but $B_j C_{ij} p$: after a lie that p, the addressee believes that speaker has shared knowledge with him about p. The modellings we propose satisfy this, but we restrict our discussion to logics without common knowledge.

In a dynamic setting, what we want, so far, is: *Lying that p is the epistemic action transforming information states satisfying $B_i \neg p$ into information states satisfying $B_j p$ and preserving $B_i \neg p$.* We need to make a choice concerning: information state, epistemic action, epistemic modal operator, and, finally, how to generalize lying about Booleans to lying about modal formulae. As *information state* we propose a multi-agent Kripke model. We consider one agent lying to one other agent; or one agent lying to the group of all other agents. A Kripke model transformation calls for a *dynamic modality*. As lying is the opposite of telling the truth, a variation of public announcement logic seems obvious. First, we model lying announcements by an external observer, comparable to truthful announcements by that observer. Then, we model lying of agent i to agent j, where both agents are modelled in the Kripke model.

Clearly, our epistemic modality cannot be knowledge. If the liar correctly believes that p is false and lies that p after which the addressee j believes p, then j holds a false belief. In AI, the next best thing to knowledge is belief, i.e., $KD45$ belief. We will *aim* for that, and therefore have to address the problem that consistency of belief is not necessarily preserved after update.

When generalizing from 'lying that p' to 'lying that φ' for epistemic propositions, we have to change to postcondition. The addressee j believes that φ is true when announced. It may no longer be true after the liar and the addressee have processed the information contained in the lie. We should require that j believes that φ *was* true before the lie, not that it still is true after the lie. This is because of Moorean phenomena: if I am lying to you, agent j, that $p \land \neg B_i p$, after the lie you believe p, not that you are ignorant about it. Lying in the consecutive number riddle is of that kind.

We conclude this introduction with an overview of the literature. Lying has been a thriving topic in the philosophical community for a long, long time [15, 5, 11, 12]— indeed, almost any analysis starts with quoting Augustine on lying (check!). The precision of the belief preconditions and postconditions is illuminating. E.g., emphasis that the addressee should not merely believe the lie but believe it to be believed by the speaker. Indeed, ... and even believed to be commonly believed, would the modal logician say. Interesting scenarios involving eavesdroppers (can you lie to an eavesdropper?) clearly are relevant for logic and multi-agent system design, and also claims that you can only lie if you really *say* something: an omission is not a lie [12]. Wrong, says the computer scientist: if the protocol is common knowledge, you can lie by *not* acting when you should; say, by not stepping forward in the muddy children problem although you know that you are muddy. The philosophical literature also clearly distinguishes between false propositions and propositions believed to be false but in fact true, so that when you lie about them, in fact you tell the truth. Interesting Gettier-like scenarios are discussed. Also, much is said on the morality of lying and on its intentional aspect. As said, we abstract from the intentional aspect of lying. We also abstract from its moral aspect.

In the modal logical community, papers on lying include [2, 16, 4, 19, 14, 9, 20]. They (almost) all model lying as an epistemic action, inducing a transformation of an epistemic model. Lying has been discussed by Baltag *et al.* from the inception of BMS onward [2, 4]; the latter also discusses lying in logics with knowledge and plausible belief (AGM belief revision with lying, so to speak), as does [19]. In [20] (dating from 2007) the conscious update in [7] is applied to model lying by an external observer to the public (of agents). The recent [14] gives a modal logic of lying, bluffing and (after all) intentions—they do not model lying as an epistemic action, and do not seem to realize the trouble this gets you into when you lie about a Moore-sentence. In [16, 9] the unbelievable lie is considered; this is the issue consistency preservation in $KD45$ updates.

2 Logical preliminaries

The logic of *lying* public announcements complements the well-known logic of *truthful* public announcements [13, 3], that is an extension of multi-agent epistemic logic. Its language, structures, and semantics are as follows.

Given a finite set of agents N and a countable set of propositional variables P, the *language* $\mathcal{L}(!)$ *of public announcement logic* is inductively defined as

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid B_i\varphi \mid [!\varphi]\psi$$

where $p \in P$, $i \in N$. For $B_i\varphi$, read 'agent i believes formula φ'. For $[!\varphi]\psi$, read 'after truthful announcement of φ, formula ψ (is true)'.

An *epistemic model* $M = \langle S, R, V \rangle$ consists of a *domain* S of *states* (or 'worlds'), an *accessibility function* $R : N \to \mathcal{P}(S \times S)$, where each R_i is an accessibility relation, and a *valuation* $V : P \to \mathcal{P}(S)$. For $s \in S$, (M, s) is an *epistemic state*, also known as a pointed Kripke model. The class of models where all accessibility relations are serial, transitive and euclidean is called $\mathcal{KD}45$. Without any restrictions we call the model class $\mathcal{K}$.

Assume an epistemic model $M = \langle S, \sim, V \rangle$.

$$\begin{array}{ll}
M, s \models p & \text{iff } s \in V_p \\
M, s \models \neg\varphi & \text{iff } M, s \not\models \varphi \\
M, s \models \varphi \wedge \psi & \text{iff } M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models B_i\varphi & \text{iff for all } \ t \in S : R_i(s,t) \text{ implies } M, t \models \varphi \\
M, s \models [!\varphi]\psi & \text{iff } M, s \models \varphi \text{ implies } M|\varphi, s \models \psi
\end{array}$$

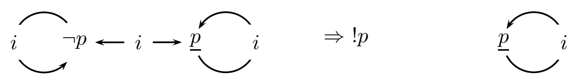
where the model restriction $M|\varphi = \langle S', R', V' \rangle$ is defined as $S' = \{s' \in S \mid M, s' \models \varphi\}$ $(= [\![\varphi]\!]_M)$, $R'_i = R_i \cap \ (S' \times S')$ and $V'(p) = V(p) \cap S'$. A complete proof system for this logic (for class $\mathcal{S}5$, originally) is presented in [13]. The interaction between announcement and belief is

$$[!\varphi]B_i\psi \leftrightarrow \varphi \to B_i[!\varphi]\psi$$

The interaction between announcement and other operators we assume known. It changes predictably in the other logics we present. The class $\mathcal{KD}45$ is not closed under public announcements: given $\neg p \wedge B_i p$, and new information $!\neg p$, agent i's accessibility relation becomes empty: she believes everything.

In the coming sections, we will only vary the dynamic part of the logic.

For an example of the semantics of public announcement, consider a situation wherein the agent is uncertain about p, and receives the information that p. In view of the continuation, we draw all access. A state has been given the value of the atom there as its name. The actual state is underlined.



3 Logic of truthful and lying public announcements

We expand the language of truthful public announcement logic with another inductive construct $[¡\varphi]\psi$, for 'after lying public announcement of φ, formula ψ (is true)'; in short 'after the lie that φ, ψ'. This is the language $\mathcal{L}(!, ¡)$.

Truthful public announcement logic is the logic to model the revelations of a benevolent god, taken as the truth without questioning. The announcing agent is not modelled in public announcement logic, but only the effect of her announcements on the audience, the set of all agents. Consider a *false* public announcement, made by a malevolent entity, the devil. Everything he says is false. Everything is a lie. Not surprisingly, god and the devil are inseparable and should be modelled simultaneously. This is as in religion.

An alternative for the semantics of public announcements is the semantics of *conscious updates* [7]. (In fact, [7] and [13] were independently proposed.) When announcing φ, instead of eliminating states where φ does not hold, one eliminates *access* to states where φ does not hold. The effect of the announcement of φ is that only states where φ is true are accessible for the agents. It is not a model restricting transformation but an arrow restricting transformation. We see this as the logic of *believed* public announcements. There is no relation between the agent accepting new information and the truth of that information.

In [20], this believed announcement of φ is called manipulative update with φ. The original proposal there is to view this as non-deterministic choice $!\varphi \cup ¡\varphi$ between truthful announcement and lying announcement, with the following semantics

$$M, s \models [!\varphi]\psi \text{ iff } M, s \models \varphi \text{ implies } M^{\varphi}, s \models \psi$$
$$M, s \models [¡\varphi]\psi \text{ iff } M, s \models \neg\varphi \text{ implies } M^{\varphi}, s \models \psi$$

where epistemic model M^{φ} is as M except that (with S the domain of M)

$$R_i^{\varphi} \;:=\; R_i \cap \; (S \times [\![\varphi]\!]_M).$$

We can keep writing $!\varphi$ for 'arrow elimininiting' truthful announcement without risk of ambiguity with 'state eliminating' truthful announcement, because on the states s where φ is true in M we have that

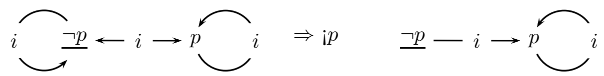
$$(M^\varphi, s) \underline{\leftrightarrow} (M|\varphi, s).$$

The axioms for truthful announcement remain what they were and the axiom for the reduction of belief after lying is

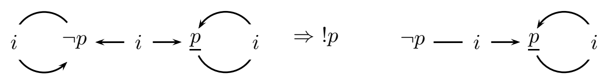
$$[¡\varphi]B_i\psi \leftrightarrow \neg\varphi \to B_i[!\varphi]\psi.$$

After the lying announcement that φ, agent i believes that ψ, if and only if, on condition that φ is false, agent i believes that ψ *after truthful announcement* that φ. To the credulous person who believes the lie, the lie appears to be the truth. This proposal to model lying has been investigated in detail in [20].

For an example, we show the effect of truthful and lying announcement of p in the model with uncertainty about p. The actual state must be different in these models: when lying, p is (believed) false, and when being truthful, p is (believed) true. For lying we get



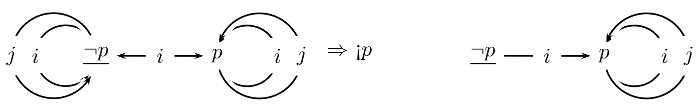
whereas for truthtelling we get



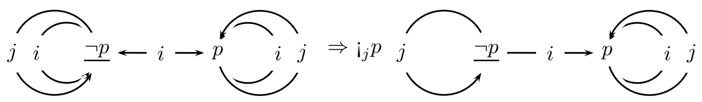
4 Agent announcement logic

In the logic of lying and truthful public announcements, the outside observer is implicit. Therefore, it is also implicit that she believes that the announcement is false or true. In multi-agent epistemic logic, it is common to formalize ‘agent i truthfully announces φ’ as ‘the outside observer truthfully announces $B_i\varphi$’. However, ‘agent i lies that φ’ cannot be modelled as ‘the outside observer lies that $B_i\varphi$’.

For a counterexample, consider an epistemic state where i does not know whether p, j knows whether p, and p is true. Agent j is in the position to tell i the truth about p. The reader can check that a truthful public announcement of B_ip indeed simulates that i truthfully announces p. Now suppose p is false, and that j lies that p. A lying public announcement of B_ip does not result in the desired information state, because this makes agent j believe his own lie. In fact, as he already knew $\neg p$, this makes j’s beliefs inconsistent.



Instead, a lie from j to i should have the following effect:



After this lie we have that j still believes that $\neg p$, i believes that p, and i believes that i and j have common belief of p. We satisfied the requirements of a truthful and lying agent announcement.

Apart from lying and telling the truth, another form of announcement is *bluffing*. You are bluffing that φ, if you say that φ but are uncertain about φ. The precondition for bluffing is therefore $\neg(B_i\varphi \vee B_i\neg\varphi)$. If belief is explicit there are always three preconditions for announcing φ: $B_i\varphi$, $B_i\neg\varphi$, and $\neg(B_i\varphi \vee B_i\neg\varphi)$, the preconditions for truthtelling, lying, and bluffing. If belief is implicit there are only two preconditions for announcing φ: φ and $\neg\varphi$, for truthtelling and lying. God and the devil are omniscient, and bluffing is therefore inconceivable for them. More prosaically, they can be considered an agent with an accessibility relation that is the identity on the model.

The logical language $\mathcal{L}(!_j, \text{¡}_j, \text{¡!}_j)$ of *agent announcement logic* is defined by adding inductive constructs

$$[!_j\varphi]\psi \mid [\text{¡}_j\varphi]\psi \mid [\text{¡!}_j\varphi]\psi$$

to the epistemic language, for, respectively, j truthfully announces φ, j is lying that φ, and j is bluffing that φ; where agent j addresses all other agents i.

The preconditions of these three types of announcement are all different, but their effect on the speaker and on the listeners are the same: States where φ was believed by j, if any (none, if j is lying), remain accessible for j (*i*); states where $\neg\varphi$ was believed by j, if any (none, if j is truthful), remain accessible for j (*ii*); states where φ was believed by i, if any (if there are none, i will 'go mad'), remain accessible for i (*iii*); and states where $\neg\varphi$ was believed by i, if any, are no longer accessible for i (*iv*). This is embodied by the following semantics.

$$\begin{array}{l} M, s \models [!_j\varphi]\psi \text{ iff } M, s \models B_i\varphi \text{ implies } M_j^\varphi, s \models \psi \\ M, s \models [\text{¡}_j\varphi]\psi \text{ iff } M, s \models B_i\neg\varphi \text{ implies } M_j^\varphi, s \models \psi \\ M, s \models [\text{¡!}_j\varphi]\psi \text{ iff } M, s \models \neg(B_i\varphi \vee B_i\neg\varphi) \text{ implies } M_j^\varphi, s \models \psi \end{array}$$

where M_j^φ is as M except that a new accessibility relation R' is defined as (S is the domain of M, and $i \neq j$)

$$\begin{array}{l} R'_j := R_j \\ R'_i := R_i \cap \ (S \times [\![\varphi]\!]_M) \end{array}$$

If φ is believed by j in state s in M we have that

$$(M_j^\varphi, s) \leftrightarrows (M|B_j\varphi, s).$$

This justifies that there is no difference between agent j truthfully announcing that φ and the truthful public announcement of $B_j\varphi$.

The principles for j lying to i are as follows:

$$[¡_j\varphi]B_i\psi \leftrightarrow B_j\neg\varphi \rightarrow B_i[!_j\varphi]\psi$$
$$[¡_j\varphi]B_j\psi \leftrightarrow B_j\neg\varphi \rightarrow B_j[¡_j\varphi]\psi$$

In other words, the liar knows that he is lying, but the dupe he is lying to, believes that the liar is telling the truth. The principles for truthtelling and bluffing are similar, but with (the obvious) different conditions on the right hand side. With these principles, the logic is completely axiomatizated. (This is, because it is a logic for a specific action model. See the next section.)

The Appendix illustrates agent lying in the consecutive numbers riddle. In the continuation we discuss consequences and variations of public lying and agent lying: an action model perspective, how to address the issue of unbelievable lies, lying about beliefs, and lying and plausible beliefs.

5 Action models and lying

Whether I am telling the truth to you, am lying, or am bluffing, to you it all appears as the same annoucement. A familiar way to formalize uncertainty about actions are *action models* [3]. We can view truthful and lying public announcement as the two points of an action model, and we can also view truthful, lying and bluffing agent announcement as the three different points in another action model.

An *action model* $\mathsf{M} = \langle \mathsf{S}, R, \mathsf{pre} \rangle$ consists of a *domain* S of *actions*, an *accessibility function* $R : N \rightarrow \mathcal{P}(\mathsf{S} \times \mathsf{S})$, where each R_i is an accessibility relation, and a *precondition function* $\mathsf{pre} : \mathsf{S} \rightarrow \mathcal{L}$, where $\mathcal{L}$ is a logical language. A pointed action model is an *epistemic action.* Performing an epistemic action in an epistemic state means computing their restricted modal product—restricted to state/action pairs (t, t) such that $M, t \models \mathsf{pre}(\mathsf{t})$. With such an epistemic action (M, s) we can associate a dynamic modal operator $[\mathsf{M}, \mathsf{s}]$ in the usual way.

The action model M' for truthful and lying public announcement consists of two actions suggestively named ! and ¡ with preconditions φ and $\neg\varphi$ in $\mathcal{L}(!, ¡)$, respectively, and for all agents only action ! is accessible. *Truthful public announcement of* φ is the epistemic action $(\mathsf{M}', !)$. Given that $\mathsf{pre}(!) = \varphi$, $[!\varphi]\psi$ corresponds to $[\mathsf{M}', !]\psi$. *Lying that* φ is the epistemic action $(\mathsf{M}', ¡)$.

The action model M'' for agent announcement consists of three actions named $¡!_j$, $!_j$, and $¡_j$ with preconditions $\neg(B_j\varphi \vee B_j\neg\varphi)$, $B_j\varphi$, and $B_j\neg\varphi$, respectively (all in $\mathcal{L}(!_i, ¡_i, ¡!_i)$). The announcing agent j has identity access on the action model and to the other agents only action $!_j$ is accessible. Agent j truthfully announcing φ to all other i is is the epistemic action $(\mathsf{M}'', !_j)$—with precondition $B_j\varphi$, therefore—and similarly lying and bluffing are the action models $(\mathsf{M}'', ¡_j)$ and $(\mathsf{M}'', ¡!_j)$. Action models M' and M'' are depicted in Figure 1.

The action model representations validate the axioms for announcement and belief, for all versions shown; and they justify that these axioms form part of

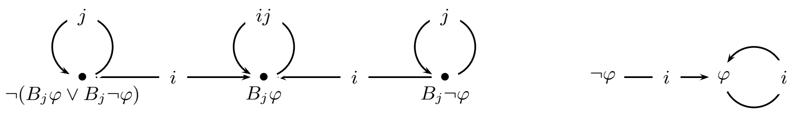


Fig. 1. Action models for lying, truthtelling and bluffing

complete axiomatizations.[1] These axioms are simply instantations of a more general axiom for an epistemic action followed by a belief. Note that M' and M'' are both in class $\mathcal{KD}45$ but nevertheless, as we have seen, executing a $\mathcal{KD}45$ epistemic action in a $\mathcal{KD}45$ epistemic state does not guarantee a $\mathcal{KD}45$ updated epistemic state.

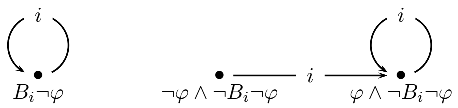
5.1 Unbelievable lies

The class of $\mathcal{S}5$ epistemic models is closed under update with $\mathcal{S}5$ epistemic actions, such as truthful public announcements, but the class of $\mathcal{KD}45$ models is *not* closed under update with $\mathcal{KD}45$ epistemic actions such as a lying public announcement. (It is not even closed under update with correct information.) The problem is that beliefs may be mistaken and that new information may be incorrect. Either way, if you tell me that p but I already believe the opposite, then I 'go mad' if I accept the new information without discarding the old information. My accessibility relation has become empty: I lose the D in $\mathcal{KD}45$.

$\mathcal{KD}45$-preserving updates have been investigated in [16, 1, 9]. Aucher [1] defines a language fragment that makes you go mad ('crazy formulas'). Steiner [16] proposes that the agent does not incorporate the new information if she already believes to the contrary. In that case, nothing happens. Otherwise, access to states where the information is not believed is eliminated, just as for believed public announcements. This solution to model unbelievable lies (and unbelievable truths!) is similarly proposed in the elegant and promising [9], where it is called *cautious update*—a suitable term.

Steiner gives a useful parable for the case where you do not accept new information. Someone is calling you and is telling you something that you don't want to believe. What do you do? You start shouting through the phone: 'What did you say? Is there anyone on the other side? The connection is bad!' And then you hang up, quickly, before the caller can repeat his message. Thus you create common knowledge that the message has been received but its content not accepted.

A three-point action model for cautious update is as follows. The difference with the action model for truthful and lying public announcement is that those alternatives now have an additional precondition $\neg B_i\neg\varphi$, meaning that the announcement is 'believable'.



[1] The logic of believed announcements was originally axiomatized in [7]. The redescription of these operations with an action model, providing the alternative axiomatization, was suggested in [17, 8].

We have explored this modelling of lying in more depth. We consider these mere variations, and move on. Note that for agent announcements, the addressee does not go mad if she already believes $\neg p$ and the speaker is lying that p. The addressee then merely concludes that $\neg p \wedge B_j p$: the speaker must be mistaken in his truthful belief of p. Of course the addressee will *still* go mad if she believed $\neg B_j p$.

For believed announcements we mentioned the problem that the agent believes new information whether it is true or not. For cautious update it still is the case that the agent can process (although maybe not believe) new information whether it is true or not, and even whether she already believed it or not. Going mad is too strong a response, but not changing contradictory beliefs is too weak. The next section presents a solution in between.

5.2 Lying and plausible belief

Suppose that we also have a preference relation, expressing which states are more and less plausible. We then can distinguish degrees of belief. For example, suppose states s and t are indistinguishable for agent i but she considers s more plausible than t; and proposition p is true in s and false in state t. The agent (defeasibly) *believes* φ if φ is true in all preferred states, and the agent *knows* (or, strongly believes) φ if φ is true in all accessible states. We keep writing B for belief and we write K for knowledge. Given that, $B_i p$ is true in t, because p is true in the preferred state s, but $K_i p$ is not true in t. When presented with evidence that $\neg p$, in t, i will eliminate s from consideration; t is now the most preferred state, and $B_i \neg p$ is now true. Such a distinction between epistemic access and preference can also be made in the action models, where agents may consider more and less plausible actions. We will refrain from details, see [18, 17, 4]. How to model lying with plausibility models was summarily discussed in [4, 19].

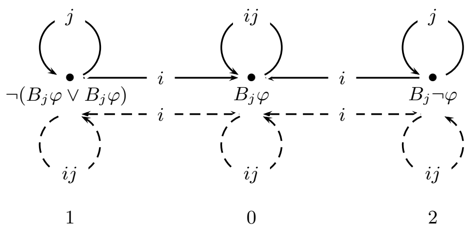


Fig. 2. Belief and preference

The action model of Figure 1 enriched with plausibility is depicted in Figure 2. The adressee i is most inclined to believe that j is telling the truth (0), less inclined to believe that he is bluffing (1), and least inclined to believe that he is lying (2). Agent i's accessibility relation is the dashed relation. (This is the universal relation. We assume transitivity.) She cannot excluse any of the three types of announcement. From that and her preference the solid accessibility relation is what she considers most likely. This will determine her plausible beliefs. (A third, intermediate degree of belief, is implicit in the figure.) Now consider an epistemic state wherein i has hard evidence that $\neg B_j p$, and let j announces p, thus suggesting $B_j p$. In the first place, i will now not go mad, the problem discussed before. She will merely eliminate truthtelling from the alternatives, and from the two remaining alternatives she considers it more likely that j is bluffing than that he is lying. If she had also hard evidence that j is not bluffing, she will still not go mad, and finally conclude that he is a liar.

5.3 Lying about beliefs

If I lie to you that "you don't know that I will fly to Amsterdam tomorrow", something of the form $p \wedge \neg B_j p$, the lie succeeds if you believe p afterwards, i.e., if $B_j p$ is then true, not if the contradictory sentence $B_j(p \wedge \neg B_j p)$ is true. This is not merely some theoretical boundary case. I can very well lie to you about the knowledge or ignorance of other agents or about my own knowledge. In fact, I do that all the time.

Agents may announce factual propositions but also modal propositions, and thus be lying and bluffing about them. For example, in the consecutive number riddle, both i and j may lie about their knowledge or ignorance of the other's number.

In social interaction, untruthfully announcing modalities is not always considered lying (with the moral connotation). Suppose we work in the same department and one of our colleagues, X, is having a divorce. I know this. I also know that you know this. But we have not discussed the matter between us. I can bring up the matter in conversation by saying 'You know that X is having a divorce!'. But this is unwise. You may not be willing to admit your knowledge, because X's husband is your friend, which I have no reason to know; etc. A better strategy for me is to say 'You may not know that X is having a divorce'. This is a lie. I do not consider it possible that you do not know that. But, unless we are very good friends, you will not laugh in my face to that and respond with 'Liar!'.

It is also strange that I may be *bluffing* if I tell you p, given that in fact I don't know if p, but I would be *lying* if I tell you that I believe that p. This is because I believe that I don't believe p: $\neg B_i p$ entails by negative introspection $B_i \neg B_i p$, where $\neg B_i p$ is now the negation of the announced formula $B_i p$!

6 Conclusions and further research

Lying is an epistemic action inducing a transformation of an epistemic model. We presented logics for public lying and truthtelling, and logics for agent lying, bluffing, and truthtelling. These logics abstract from the moral and intentional aspect of lying, and only consider the effect of lies that are believed by the addressee. We also presented versions that treat unbelievable lies differently, and lying in the presence of plausible (defeasible) belief.

There are many topics for further research. **1.** Explicit agency is missing in our approach (as so often in dynamic epistemic logics). **2.** We only summarily discussed common knowledge—this seems a straightforward enough generalization, that also allows for more more refined preconditions then merely requiring that lies are believable for the addressee. A good (and possibly strongest?) precondition seems:

$$B_i\neg\varphi \wedge \neg B_j\neg\varphi \wedge C_{ij}((B_i\varphi \vee B_i\neg\varphi) \wedge \neg(B_j\varphi \vee B_j\neg\varphi))$$

3. One problem with lying to some and telling the truth to others is that you have to keep track of who knows the truth and who not, and that you should carefully consider what you can still say and in whose company. In everyday communication, this (logical) computational cost of lying seems a strong incentive against lying. Can this intuition be formalized? We are inspired by results on the computational cost of insincere voting in social choice theory [6]: in well-designed voting procedures this is intractable, so that sincere voting is your best strategy. **4.** In multi-agent systems with several agents on may investigate how robust certain communication procedures are in the presence of few liars; and results might be compared to those for signal analysis with 'intentional' noise. **5.** Finally, we would like to model a liar's paradox in a dynamic epistemic logic.

Acknowledgement

I thank the workshop reviewers for their comments.

References

1. G. Aucher. Consistency preservation and crazy formulas in BMS. In S. Hölldobler, C. Lutz, and H. Wansing, editors, *Logics in Artificial Intelligence, 11th European Conference, JELIA 2008. Proceedings*, pages 21–33. Springer, 2008. LNCS 5293.
2. A. Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54(1):1–45, 2002.
3. A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pages 43–56, 1998.
4. A. Baltag and S. Smets. The logic of conditional doxastic actions. In K.R. Apt and R. van Rooij, editors, *New Perspectives on Games and Interaction*, Texts in Logic and Games 4. Amsterdam University Press, 2008.
5. S. Bok. *Lying: Moral Choice in Public and Private Life.* Random House, New York, 1978.
6. V. Conitzer, J. Lang, and L. Xia. How hard is it to control sequential elections via the agenda? In *IJCAI'09: Proceedings of the 21st international joint conference on Artifical intelligence*, pages 103–108. Morgan Kaufmann Publishers Inc., 2009.

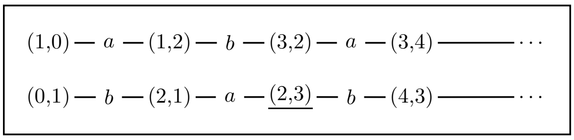
7. J.D. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.
8. B. Kooi. Expressivity and completeness for public update logics via reduction axioms. *Journal of Applied Non-Classical Logics*, 17(2):231–254, 2007.
9. B. Kooi and B. Renne. Arrow update logic. Manuscript, 2010.
10. J.E. Littlewood. *A Mathematician's Miscellany*. Methuen and company, 1953.
11. J.E. Mahon. Two definitions of lying. *Journal of Applied Philosophy*, 22(2):21–230, 2006.
12. J.E. Mahon. The definition of lying and deception. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2008. `http://plato.stanford.edu/archives/fall2008/entries/lying-definition/`.
13. J.A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216. Oak Ridge National Laboratory, 1989.
14. C. Sakama, M. Caminada, and A. Herzig. A logical account of lying. In *Proceedings of JELIA 2010, LNAI 6341*, pages 286–299, 2010.
15. F.A. Siegler. Lying. *American Philosophical Quarterly*, 3:128–136, 1966.
16. D. Steiner. A system for consistency preserving belief change. In *Proceedings of the ESSLLI Workshop on Rationality and Knowledge*, pages 133–144, 2006.
17. J. van Benthem. Dynamic logic of belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.
18. H. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese (Knowledge, Rationality & Action)*, 147:229–275, 2005.
19. H. van Ditmarsch. Comments on 'the logic of conditional doxastic actions'. In K.R. Apt and R. van Rooij, editors, *New Perspectives on Games and Interaction*, Texts in Logic and Games 4, pages 33–44. Amsterdam University Press, 2008.
20. H. van Ditmarsch, J. van Eijck, F. Sietsma, and Y. Wang. On the logic of lying. In J. van Eijck and R. Verbrugge, editors, *Games, Actions and Social Software*. Springer, 2011. FoLLI-LNCS series 'Texts in Logic and Games'. To appear.

Appendix: Lying about consecutive numbers

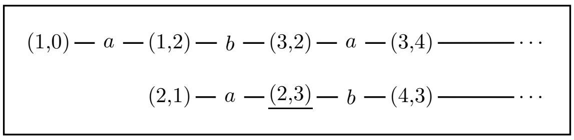
The consecutive numbers riddle is often attributed to Littlewood [10]. It is as follows.

> *Anne and Bill are each going to be told a natural number. Their numbers will be one apart. The numbers are now being whispered in their respective ears. They are aware of this scenario. Suppose Anne is told 2 and Bill is told 3.*
> *The following truthful conversation between Anne and Bill now takes place:*
> – *Anne: "I do not know your number."*
> – *Bill: "I do not know your number."*
> – *Anne: "I know your number."*
> – *Bill: "I know your number."*
>
> *Explain why is this possible.*

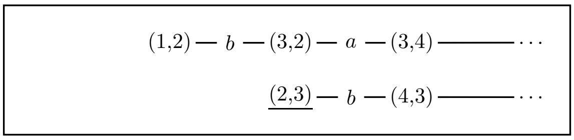
First, the standard analysis of the informative consequences of these four announcements.



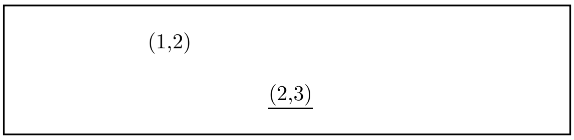
– Anne: "I do not know your number."



– Bill: "I do not know your number."



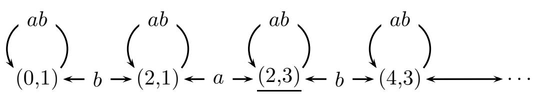
– Anne: "I know your number."



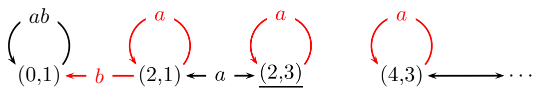
– Bill: "I know your number."

This last announcement does not make a difference anymore, as it is already common knowledge that Anne and Bill know each other's number.

Next, we show two different scenarios for the consecutive number riddle with lying. This is agent lying (and truthtelling), the actions we modelled as $!_i\varphi$ and $¡_i\varphi$. (Bluffing is not an option in this example, because the lying is about ignorance or knowledge, and introspective agents *know* their ignorance and *know* their knowledge.) As we are reasoning from the actual state $(2,3)$, we do not depict the top chain of possibilities any more. And as beliefs may now be incorrect, we show all arrows. Positions in the model where a change took place (i.e., where arrows have been removed) are shown in red. The first scenario consists of Anne lying in her first announcement. We do not model Bill's response that Anne is a liar! After Anne's lie, in the actual state $(2,3)$, Bill does not consider any state possible, and therefore believes everything. (Of course you have Bill say that he has gone mad—by way of truthfully announcing that $B_j(p \wedge \neg p)$.)

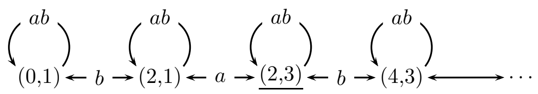


– Anne: "I know your number." Anne is lying

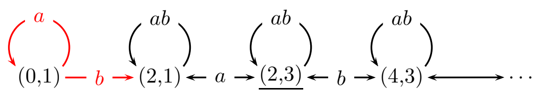


– Bill: "That's a lie."

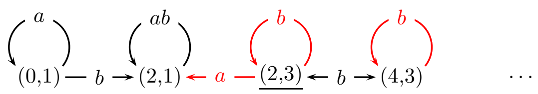
In the second scenario Anne initially tells the truth, after which Bill is lying, resulting in Anne mistakenly concluding (and announcing) that she knows Bill's number: observe that she believes it to be 1. This mistaken announcement by Anne is informative to Bill: he learns from it (correctly) that Anne's number is 3.



– Anne: "I do not know your number."



– Bill: "I know your number." Bill is lying



– Anne: "I know your number." Anne is mistaken.