Journal of Logic, Language and Information (2022) 31:423–450
https://doi.org/10.1007/s10849-022-09351-4

Public Announcements, Public Lies and Recoveries

Kai Li[1] · **Jan van Eijck**[2]

Accepted: 26 January 2022 / Published online: 17 March 2022

Abstract
The paper gives a formal analysis of public lies, explains how public lying is related to public announcement, and describes the process of recoveries from false beliefs engendered by public lying. The framework treats two kinds of public lies: simple lying update and two-step lying, which consists of suggesting that the lie may be true followed by announcing the lie. It turns out that agents' convictions of what is true are immune to the first kind, but can be shattered by the second kind. Next, recovery from public lying is analyzed. Public lies that are accepted by an audience cannot be undone simply by announcing their negation. The paper proposes a recovery process that works well for restoring beliefs about facts but cannot be extended to beliefs about beliefs. The formal machinery of the paper consists of KD45 models and conditional neighbourhood models, with various update procedures on them. Completeness proofs for a number of reasoning systems (converse belief logic, public lies logic, lying and recovery logic, conditional neighbourhood logic, plus its dynamic version) are included.

Keywords Dynamic epistemic logic · Multi-agent systems · Lying · Recovery · Conditional beliefs

1 Introduction

> 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

K. Li, The authors wish to thank Hans van Ditmarsch and Malvin Gattinger for their help and advice. We are also grateful to an anonymous reviewer for his/her comments on an earlier version of the paper.

✉ Kai Li
likaiedemon@gmail.com

[1] China University of Political Science and Law, Beijing, China

[2] ILLC, Amsterdam, The Netherlands

> 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.
> Hannah Arendt, Truth and Politics ((Arendt(1967(Penguin Classics Edition, 2006)))

The effect of public lies, according to Hannah Arendt, is that it destroys our bearings in the world. In this paper, we will make an attempt to explain this formally. We will also model how to recover from public lies. For this, we model *public lies* along the same lines as *public announcements*.

Our starting point is the representation of knowledge, ignorance and belief by means of Kripke models, more specifically KD45 models. Further on in the paper, we will also use conditional neighbourhood models, to model conditional beliefs.

We will model both public announcements and public lies as maps from Kripke models to Kripke models. The results of public lies are Kripke models where Bayesian conditioning gives *wrong* results, in the sense that agents can be 100% sure of things that are not true. The effect of public lies cannot be detected from the inside: agents still have fully consistent world views. The only thing is that they can be out of touch with reality. But the agents have no means of knowing this.

In order to explain *recoveries from false beliefs* one has to invoke the effects of acting on false beliefs. The results or utilities of our actions are not determined by our beliefs but by the real world.

To see how rational investigation and approach to the truth should ideally proceed, consider the following quote from MacKay (2003):

> Denote the proposition 'the suspect and one unknown person were present' by S. The alternative, $\bar{S}$, states 'two unknown people from the population were present'. The prior in this problem is the prior probability ratio between the propositions S and $\bar{S}$. This quantity is important to the final verdict and would be based on all other available information in the case. Our task here is just to evaluate the contribution made by the data D, that is, the likelihood ratio, $P(D|S, H)/P(D|\bar{S}, H)$. In my view, a jury's task should generally be to multiply together carefully evaluated likelihood ratios from each independent piece of admissible evidence with an equally carefully reasoned prior probability. [This view is shared by many statisticians but learned British appeal judges recently disagreed and actually overturned the verdict of a trial because the jurors had been taught to use Bayes' theorem to handle complicated DNA evidence.]

The core principles of rational belief seem to rely heavily on conditional reasoning. Suppose ϕ (a proposition that is not in contradiction with anything you know) is true. Would you then believe ψ? In other words, if the world would turn out to be ϕ, would you still believe ψ? It is important that the condition is not a counterfactual. If one knows that a condition does not hold, then speculating about what one would believe if it were otherwise is usually not fruitful for getting closer to the truth. We interpret belief in ψ conditional on ϕ in the information-theoretic sense. The inspiration for this is Bayesian update, with the following very useful notion of belief: Belief as willingness to bet on ψ, given information ϕ.

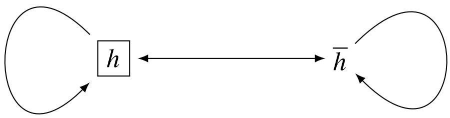
In Sect. 3 we introduce the converse belief operator, we show how knowledge can be expressed in terms of belief and converse belief, and we model public lies along the lines of Steiner (2006), Kooi and Renne (2011). Then we provide a recovery operation in Sect. 4. We show that if truth tellers are able to sequentially perform a recovery operation and a public announcement, then the audience can recover from false beliefs. However this observation also indicates that liars can use the same tactic; if they do then their alleged "truth" becomes a mutual conviction (mutual KD45 belief).

In Sect. 5 we discuss the effect of public lies and recoveries on a slightly different notion of belief: propositions that one assigns a probability greater than 0.5, under some condition. Because an audience may use prior probabilities different from those of truth tellers, and truth tellers are bound to announce truth, liars may have an advantage over truth tellers. We provide logic systems and completeness proofs.

2 Epistemic Logic and Public Announcements

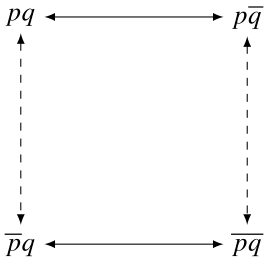
At the core of epistemic logic is the representation of uncertainty by means of a set of current options for what the actual world could turn out to be like (cf. Stalnaker (2006)).

Consider the case of a single fact, let us say the outcome of a coin toss, where the coin has landed, but is hidden under a cup. Let h represent the situation where the coin has landed heads up, and $\overline{h}$ the situation where the coin has landed tails up. Ignorance of some individual about this situation can be represented as follows:



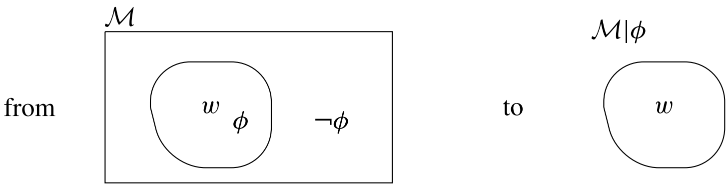
The actual world is $\boxed{h}$, but this indication of what is actually the case is invisible to the agent a. In general, if a representation for a knowledge situation contains a pointer to the actual world, then this pointer is always invisible to the knowing agents.

A situation where I know one thing and you know another thing has at least four possible states of affairs. Suppose a knows the status of p and b the status of q. Say they both toss a coin, and p denotes heads for a, q denotes heads for b. We now need to distinguish the four possible outcomes, as follows, where the solid arrows represent the uncertainty of a, and the dashed arrows are for b. For convenience, we leave out the self-loops.



Note that the accessibility relation of a (and b resp.) is an equivalence relation that is reflexive, transitive and symmetric. The models where all the accessibility relations are equivalence relations are called S5 models.

Public announcement logic was pioneered in Plaza (1989). Intuitively, a public announcement would make an agent restrict her belief to the announced case. A natural way to implement this is by restricting every belief-cell to ϕ-worlds after announcing ϕ. Observe that the public announcement update can be viewed as a restriction of the model and its accessibility relations to the set of worlds where the announcement is true. In a picture:



Alternatively, we can model public announcements by means of cutting accessibility links. The public announcement of ϕ results in cutting the links between ϕ and $\neg\phi$ situations. The precondition for publicly announcing ϕ is that ϕ is true in the real world. Viewed as a relational change, we can model this as the change from a to $(?\phi; a; ?\phi)$. Notice that this change maps equivalence relations to equivalence relations.

The key validity for public announcement is:

$$[!\phi]K_a\psi \leftrightarrow (\phi \rightarrow K_a(\phi \rightarrow [!\phi]\psi)).$$

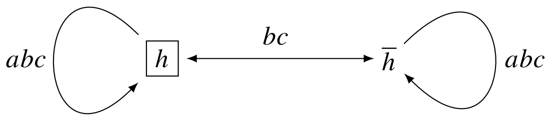
It expresses the equivalence of the following two statements: (1) a knows ψ after publicly announcing ϕ, and (2) if ϕ is true, then a knows that ϕ implies that ψ holds after a public announcement ϕ. We make our assertion on the right (the assertion about the model after the update) conditional on $!\phi$ being executable, i.e., on ϕ being true. Note that the consequent simplifies to the equivalent formula $\phi \rightarrow K_a[!\phi]\psi$.

Lying Announcements

As public announcements can be regarded as changes of epistemic models, it is natural to consider whether lies can be treated in a similar way.

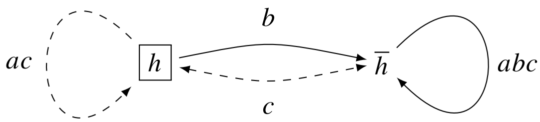
Following Augustine, a lie is a statement that the liar disbelieves, and intends to make the listener believe (cf. Mahon (2016)). Thus a lie can be seen as an action with two preconditions: (1) the liar disbelieves his statement and (2) the liar intends to deceive the listener. In dynamic epistemic logic (DEL), the first precondition can be embedded in action models (cf. van Ditmarsch (2014)). The second precondition about the liar's intention can be modelled by introducing new modal operators for intentions (Sakama et al. (2010)), but this requires introducing new relations for each agent in the model, and is omitted in most of the works using dynamic epistemic logic.

To model a speaker who lies to a listener, let us reconsider the coin tossing example. Suppose the coin has landed heads up, and is hidden by agent a under a cup. Both b and c are ignorant of the situation, but know that a is aware of the status of the coin:



Note that it is common knowledge that neither b nor c can distinguish $\boxed{h}$ and $\overline{h}$. In order to deal with the effects of lies one has to shift from S5 models to KD45 models (models where the epistemic state of an agent gets represented as a relation that is serial, transitive and euclidean, and is usually interpreted as 'belief').

Now suppose a privately lies to b that the coin has landed tails up. The updated model is given below, where the uncertainty of b is emphasized by the solid arrows:



Agent b is deceived by a, and as a result b falsely believes the actual world is $\overline{h}$. Note that all arrows of b pointing to $\boxed{h}$ are eliminated and all those to $\overline{h}$ remain unchanged. Also note that the relation represented by b's arrows is serial, transitive and euclidean.

In this picture the arrows of c are unchanged, which means she knows b's belief update, and everyone knows what she knows. This way of modelling is proposed by Steiner (2006). Steiner studied belief change on KD45 models, where the statement is announced to a subgroup of the population. Those outside the subgroup are unaffected, except for the fact that they will notice the belief changes of the subgroup. Thus an explanation of c's unaffectedness is that c noticed that a lied to b about h, but was suspicious about a's statement. Furthermore c's overhearing and suspicion are common knowledge to all three of them. This seems a bit too strong. To get around this, it is convenient to assume that all agents in our models are either speakers or listeners.

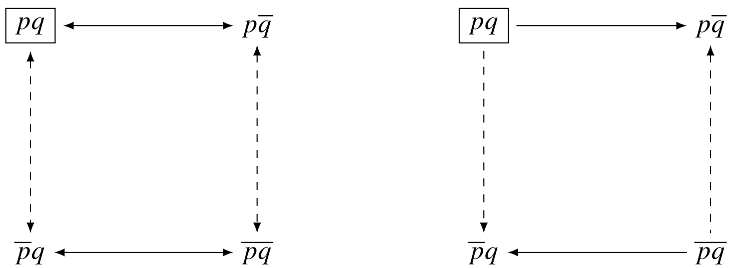
The key validity for this kind of lies is (van Ditmarsch et al. (2012)):

$$[¡_a\phi]B_b\psi \leftrightarrow (B_a\neg\phi \rightarrow B_b[!_a\phi]\psi).$$

It expresses that b believing ψ after a lie that ϕ amounts to the following: if the liar does not believe ϕ, then b believes that after a truthful announcement that ϕ, ψ holds.

If we focus on the effects of lies, we can further assume all agents are listeners. Thus liars can be regarded as outside speakers/observers, and the two preconditions (liars' disbeliefs and intentions) of lying can be removed from our framework.

Let us reconsider the coin tossing example with two coins. This time an outsider speaker lies to a and b that the two coins landed differently ($p \leftrightarrow \neg q$), which is illustrated in the following picture:



Before lie ($p \leftrightarrow \neg q$) After lie ($p \leftrightarrow \neg q$)

The statement $p \leftrightarrow \neg q$ makes a and b falsely believe that coins tossed by them respectively landed differently. Furthermore it becomes a common belief.

Note that $p \leftrightarrow \neg q$ would be a public announcement if the evaluation point were $p\overline{q}$. van Ditmarsch et al. (2012) suggest this kind of updates is a generalization of public announcements: if the announcement is true, it is a public announcement, and if it is false, the public is deceived into taking the announcement as true. This kind of lying announcement is referred to as public lies, and the outside liar is considered by van Ditmarsch (2014) as a malevolent agent who always tells falsehoods.

The key validity for public lies is:

$$[¡\phi]B_b\psi \leftrightarrow (\neg\phi \rightarrow B_b[!\phi]\psi).$$

It expresses that believing ψ after a lie that ϕ is equivalent to the following implication: $\neg\phi$ entails the belief that a public announcement of ϕ implies ψ.

However, this kind of public lie assumes that the audience is credulous enough to accept any statement, even statements that contradict what the audience believes. Thus, by this axiom, after an unbelievable public lie, the audience will believe anything. Since this is not very realistic, in the next section we will turn our attention to more cautious audiences.

3 Public Lies and KD45 Beliefs

In this section we extend the framework to public lying. As mentioned earlier, in order to deal with the effects of lies one has to shift from belief based on S5 models to KD45 models. The epistemic state now means unshakable consistent conviction. A KD45 model is a model where all accessibility relations are serial, transitive and euclidean. A KD45 model looks like a octopus, for a KD45 relation R can be viewed as a union of two relations R_1 and R_2 where $R_1 = \{(x, y) \in R | (y, x) \notin R\}$ and $R_2 = R - R_1$. The R_1 part is the set of tentacles into the body of the octopus, and the R_2 part (the body of the octopus) is an equivalence relation. (Someone might say that a KD45 model looks more like a coronavirus – a body with spikes – but we prefer the less scary octopus image.)

Definition 1 A **belief model** is a tuple $\mathcal{M} = (W, R, V)$ where W is a nonempty set of worlds, each accessibility relations R_a is serial, transitive and euclidean (D45 relations), and V is a valuation on W.

A belief model is a model where the knowledge-cell of w given ϕ need not include w. Given world w and agent a, the belief-cell of a, written as $R_a(w)$, is the set of worlds that are a-accessible from w. Note that w need not be in $R_a(w)$. If R_a is serial and euclidean, then $R_a(w)$ is an equivalence. Indeed, it follows from seriality of R_a that $R_a(w)$ is non-empty. Next, assume (w, u) and (w, v) are both in R_a. We have to show that $(u, v) \in R_a$. But this follows immediately from euclideanness of R_a.

In order to incorporate public lies and their effects, we will turn our attention to belief operators. Our basic language includes belief operators and their reverse operators. Let $\mathcal{A}g$ be a finite set of agents, and let $Prop$ be a set of propositional variables. The basic language is given by the following BNF-form:

$$\phi ::= p | \neg\phi | \phi \wedge \phi | B_a\phi | B_a^{\smile}\phi$$

$\vee$, $\rightarrow$ and $\top$ are defined as usual. $B_a\phi$ can be interpreted as "a is convinced of the truth of ϕ" or "a is certain of ϕ", in the sense of no new information will change this conviction. $B_a^{\smile}$ is the reverse of B_a, and $B_a^{\smile}\phi$ can be read as "if a's conviction is true, then she knows ϕ". We introduce these reverse operators mainly for technical reasons, to ensure that the language is expressive enough to describe belief-cells. $\hat{B}_a$ is the abbreviation of $\neg B_a \neg$.

Definition 2 Let $\mathcal{M} = (W, R, V)$ be a belief model. Key causes of truth conditions for reverse operators $B_a^{\smile}$ are:

$$\mathcal{M}, w \models B_a^{\smile}\phi \text{ iff } \mathcal{M}, u \models \phi \text{ for each } u \in W \text{ s.t. } uR_a w.$$

We use $[\![\phi]\!]_{\mathcal{M}}$ for the set $\{w \in W | \mathcal{M}, w \models \phi\}$ as usual, and omit the index $\mathcal{M}$ if it is clear in the context.

Note that we can use $B_a B_a^{\smile}$ as the context or background knowledge operator K_a, which, as we will see later, cannot be affected by public lies nor by recoveries. Similar to typical S5 knowledge operators, we can derive the truth condition for K_a:

$$\mathcal{M}, w \models K_a\phi \text{ iff } \mathcal{M}, u \models \phi \text{ for each } u \in W \text{ s.t. } wR_a \circ R_a^{-1} v,$$

and we use $\hat{K}_a\phi$ for $\neg K_a \neg\phi$. Note that if B_a would satisfy weaker conditions this would also affect the conditions for knowledge. For instance, if B_a only satisfies D, then the background knowledge K_a would satisfy only principle T. The reader can also check that if the system for B_a is D4.2, that is serial, transitive and convergent, then K_a satisfies the knowledge principles S4.2 proposed by Stalnaker (2006).

It is also worth noting that our language is more expressive than the language of standard epistemic and doxastic logic, because, for instance, $B_a^{\smile}\bot$-worlds are outside belief-cells of a, which basically says that a's conviction is false. We can also define knowledge-cells of a as follows:

Definition 3 Let $\mathcal{M} = (W, R, V)$ be a belief model, and let $w \in W$. The knowledge-cell of a containing w is given by $[w]_a = \{v \in W | wR_a \circ R_a^{-1} v\}$.

It is easy to check that $R_a \circ R_a^{-1}$ is reflexive and symmetric. Before proving transitivity, we show the following lemma that worlds in a knowledge-cell share the same belief-cell, which parallels the positive introspection principle $B_a\phi \rightarrow K_a B_a\phi$ discussed by Stalnaker (2006)).

Lemma 4 *Let $\mathcal{M} = (W, R, V)$ be a belief model and let $a \in \mathcal{A}g$. Then for each $u \in [w]_a$, $R_a(u) = R_a(w)$.*

Proof Let $u \in [w]_a$. Then there is a $v \in W$ such that $wR_a v R_a^{-1} u$, i.e., $v \in R_a(w) \cap R_a(u)$.

Suppose $w' \in R_a(w)$. Because R_a is euclidean, we have $vR_a w'$, and then by transitivity of R_a and $uR_a v$ we have $uR_a w'$, i.e., $w' \in R_a(u)$.

Thus follows that $R_a(u) \subseteq R_a(w)$. Similarly we can prove $R_a(w) \subseteq R_a(u)$. Therefore $R_a(w) = R_a(u)$. □

To show the transitivity of $R_a \circ R_a^{-1}$, suppose $wR_a \circ R_a^{-1} u$ and $uR_a \circ R_a^{-1} v$. Then both w and v are in $[u]_a$. Using the above lemma we have $R_a(w) = R_a(v)$, which implies the transitivity of $R_a \circ R_a^{-1}$. Recall that $R_a \circ R_a^{-1}$ is reflexive and symmetric. It follows that $R_a \circ R_a^{-1}$ is an equivalence relation, and $[w]_a$ is a knowledge-cell for our background knowledge operators K_a, as in standard epistemic logic.

The calculus CBL (for converse belief logic) is given by the following axioms and rules:

(Taut)	All instances of propositional tautologies
(Dist-B)	$B_a(\phi \to \psi) \to B_a\phi \to B_a\psi$
(BD)	$\hat{B}_a\top$
(B4)	$B_a\phi \to B_aB_a\phi$
(B5)	$\neg B_a\phi \to B_a\neg B_a\phi$
(Dist-C)	$B_a^{\smile}(\phi \to \psi) \to B_a^{\smile}\phi \to B_a^{\smile}\psi$
(BC)	$\phi \to B_a\neg B_a^{\smile}\neg\phi$
(CB)	$\phi \to B_a^{\smile}\hat{B}_a\phi$

RULES:

$$\frac{\phi \to \psi \quad \phi}{\psi}(\text{MP}) \qquad \frac{\phi}{B_a\phi}(\text{Nec-B}) \qquad \frac{\phi}{B_a^{\smile}\phi}(\text{Nec-C})$$

Axioms (BC) and (CB) are the usual converse axioms.

Theorem 5 *The calculus CBL is sound and complete for belief models.*

Proof Note that the KD45 system for standard doxastic logic is a sub-logic of this calculus, and is complete for belief models. Thus this theorem immediately follows from the completeness result of the converse axioms for bidirectional frames (Corollary 4.36, Blackburn et al. (2001), Chapter 4). □

Using this result it is easy to check the validity of the following formulas.

1. $\phi \wedge B_a\neg\phi \to B_a^{\smile}\psi$
2. $\neg B_a^{\smile}\bot \wedge B_a^{\smile}\phi \to B_a\phi$

(1) says that if agent a's convictions are false, then the real world is not a-accessible. (2) says that if a's beliefs are true and ϕ is her converse belief, then she is certain of ϕ. However, note that $B_a^{\smile}\phi \to B_a\phi$ is not valid, because the real world may not be in a's belief cell and a may not believe ϕ, which implies $B_a^{\smile}\phi$ is true and $B_a\phi$ is false.

Public Lies

In this subsection we model the effects of public lies on the convictions of an audience. As mentioned earlier, the traditional definition of lying requires (1) the liar disbelieves the statement and (2) the liar intends to make the listener believe the statement. We will briefly discuss (1) at the end of Sect. 5, but since our formal machinery does not allow us to model intentions, we have no way of giving an account of (2).

Because credulous listeners of public lies at the end of the previous section are unrealistic, we turn our attention to cautious audiences. Our *cautious audience assumption* (Kooi and Renne (2011)) for conviction change is that an agent accepts statements consistent with her convictions, and rejects those she is certain of being false.

Not all public lies have such effects, and we focus on those that have. Brainwashing, for instance, can be regarded as a tactic of public lying that influences the convictions of its target audience by means of a repeated stream of public lies accompanied by suppression of diverging opinions. We will not look into the complex structure of brainwashing techniques in this paper. We simply treat brainwashing as public lying that has an effect on the convictions of its audience.

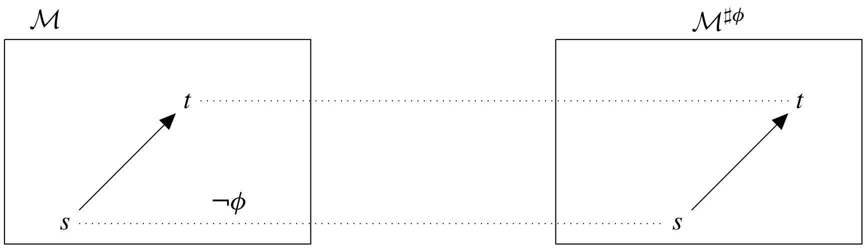
Steiner (2006) studied belief change of the audience who react to announcements in accordance with the cautious audience assumption. Kooi and Renne (2011) showed that Steiner's system is a special case of their theory, and they call this kind of audience *cautious*. The effects of lying to a cautious audience, along with other two types of agents, is also modelled by van Ditmarsch (2014), also by means of action models. Instead of using the word "cautious", he calls them *skeptical* and adds another precondition for lying: the skeptical listener "considers it possible that the speaker believes [the statement]". However since public lies and especially brainwashing are usually performed by powerful (and perhaps insane) people, it is hard for the audience to understand what is going on in their heads, let alone to keep skeptical if the statement looks authentic. Therefore we will restrict our attention to cautious audiences.

A successful public lie $\neg\phi$ will cut the accessibility links of the audience to the real world (where ϕ is true), which is modelled as relation change: from c to $(?\phi; c; ?\neg\phi) \cup (?\neg\phi; c; ?\neg\phi)$.

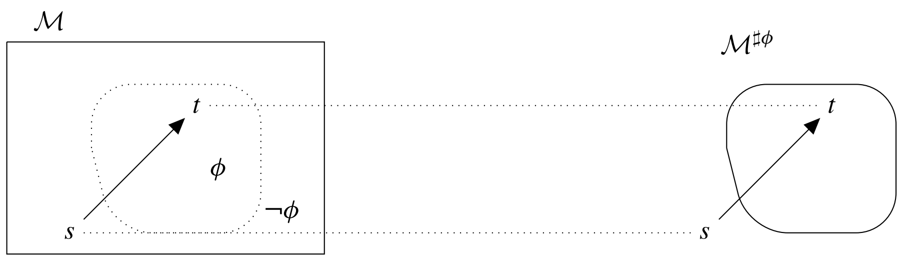
We use a dynamic update operator "$[\sharp\phi]$" for public lying about ϕ. "$[\sharp\phi]\psi$" can be interpreted as "after public lying ϕ, ψ becomes true". The key validity for public lying is:

$$[\sharp\phi]B_a\psi \leftrightarrow (B_a\neg\phi \rightarrow B_a[\sharp\phi]\psi) \wedge (\hat{B}_a\phi \rightarrow B_a(\phi \rightarrow [\sharp\phi]\psi))$$

This formula is a reformulation of axioms (A4) and (A5) for KD45 belief change given by Steiner (2006). The corresponding picture is given below, where $\mathcal{M}^{\sharp\phi}$ is the model updated after public lie ϕ:



(Case 1:$\mathcal{M}, s \vDash B_a\neg\phi$)



(Case 2:$\mathcal{M}, s \vDash \hat{B}_a\phi$)

The formula $[\sharp\phi]B_a\psi$ says that, in $\mathcal{M}^{\sharp\phi}$, all worlds t that are a-accessible from s satisfy ψ.

However as this update was originally treated as belief change in Steiner (2006), $[\sharp\phi]$ can also be interpreted as a truthful public announcement if ϕ is believed by the speaker. Because the precondition for the speaker's epistemic state is abstracted in our modelling (until the end of Sect. 5), public lies and truthful public announcement become the same update mechanism, which should not be surprising as it is a reflection of the difficulty of detecting lies. Even though we will use $[\sharp\phi]$ for both public lying about ϕ and truthful announcement ϕ, we will call it public lying update for convenience. Nevertheless we may still define public announcement consistent with the cautious audience assumption (with a slight abuse of notation):

$$[!\phi]\psi ::= \phi \to [\sharp\phi]\psi$$

The Language $\mathcal{L}_{PL}$ for public lies is the basic language plus operators $[\sharp\phi]$, which is given by the following BNF-form:

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid B_a\phi \mid B_a^{\smile}\phi \mid [\sharp\phi]\phi.$$

In a next move, we can define products of public lying updates for belief models formally.

Definition 6 Let $\mathcal{M} = (W, R, V)$ be a belief model and let ϕ be a formula in $\mathcal{L}_{PL}$-language. Model $\mathcal{M}^{\sharp\phi} = (W, R^{\sharp\phi}, V)$ is the model updated by public lying ϕ iff

- $R_a^{\sharp\phi} = \{(w, u) \in R_a \mid \mathcal{M}, u \vDash \phi \text{ or } \mathcal{M}, u \vDash B_a\neg\phi)\}$.

The key causes of truth conditions for public lying operators:

$$\mathcal{M}, w \vDash [\sharp\phi]\psi \text{ iff } \mathcal{M}^{\sharp\phi}, w \vDash \psi.$$

It is easy to verify that the updated relation $R_a^{\sharp\phi}$ is also serial, transitive and euclidean. However public lies cannot influence one's background knowledge, as is shown in the following lemma.

Lemma 7 *Let $\mathcal{M} = (W, R, V)$ be a belief model, let $w \in W$ and let ϕ be any formula. Then $[w]_a = [w]_a^{\sharp\phi}$.*

Proof By the above definition, $R_a^{\sharp\phi} \subseteq R_a$. Thus it suffices to show that for each $u \in W$, $u \in [w]_a$ implies $u \in [w]_a^{\sharp\phi}$, i.e., $uR_a v$ and $wR_a v$ for some $v \in W$ only if $uR_a^{\sharp\phi} v$ and $wR_a^{\sharp\phi} v$ for some $v \in W$.

Consider any $u \in [w]_a$. By Lemma 4 we have $R_a(u) = R_a(w)$. Clearly either $R_a(w) \cap [\![\phi]\!]_{\mathcal{M}} = \emptyset$ or $R_a(w) \cap [\![\phi]\!]_{\mathcal{M}} \neq \emptyset$. Suppose $R_a(w) \cap [\![\phi]\!]_{\mathcal{M}} = \emptyset$, which means there is no $v \in W$ such that $wR_a v$ and $\mathcal{M}, v \models \phi$. Using the definition above, we have $wR_a v$ iff $wR_a^{\sharp\phi} v$ for each $v \in W$, i.e., $R_a(w) = R_a(w)^{\sharp\phi}$. Similarly we can obtain $R_a(u) = R_a(u)^{\sharp\phi}$, which implies that $R_a(u)^{\sharp\phi} = R_a(u) = R_a(w) = R_a(w)^{\sharp\phi}$.

Suppose $R_a(w) \cap [\![\phi]\!]_{\mathcal{M}} \neq \emptyset$. Then by the definition above, for each $v \in W$, $wR_a^{\sharp\phi} v$ iff $v \in R_a(w)$ and $\mathcal{M}, v \models \phi$. Similarly for each $v \in W$, $uR_a^{\sharp\phi} v$ iff $v \in R_a(u)$ and $\mathcal{M}, v \models \phi$. Recall that we already have $R_a(u) = R_a(w)$. Thus we can also obtain $R_a(u)^{\sharp\phi} = R_a(w)^{\sharp\phi}$. Therefore either implies $R_a(u)^{\sharp\phi} = R_a(w)^{\sharp\phi}$, and because $R_a^{\sharp\phi}$ is serial, u is also in the knowledge-cell $[w]_a^{\sharp\phi}$, which completes our proof. □

The calculus PLL (public lies logic) is CBL plus the following reduction axioms for $[\sharp\phi]$:

($\sharp$1)	$[\sharp\phi]p \leftrightarrow p$
($\sharp$2)	$[\sharp\phi]\neg\psi \leftrightarrow \neg[\sharp\phi]\psi$
($\sharp$3)	$[\sharp\phi](\psi \wedge \chi) \leftrightarrow [\sharp\phi]\psi \wedge [\sharp\phi]\chi$
($\sharp$4)	$[\sharp\phi]B_a\psi \leftrightarrow (B_a\neg\phi \rightarrow B_a[\sharp\phi]\psi) \wedge (\hat{B}_a\phi \rightarrow B_a(\phi \rightarrow [\sharp\phi]\psi))$
($\sharp$5)	$[\sharp\phi]B_a^{\smile}\psi \leftrightarrow (\phi \vee B_a\neg\phi \rightarrow B_a^{\smile}[\sharp\phi]\psi)$

RULES:

$$\frac{\phi}{[\sharp\psi]\phi}\ (\text{Nec-}\sharp) \qquad \frac{\phi \leftrightarrow \psi}{[\sharp\phi]\chi \leftrightarrow [\sharp\psi]\chi}\ (\text{Rep-}\sharp)$$

The intuition of Axiom ($\sharp$5) is that if after $[\sharp\phi]$ agent a's convictions will still be true, then it is necessary that either ϕ is true, or a is certain of not ϕ.

Theorem 8 *The calculus PLL is sound and complete for belief models.*

Proof Because the new axioms from RL are reduction axioms, it suffices to show that these reduction axioms are sound. It is easy to check that axiom ($\sharp$4) express public lying update at the syntactic level. We only have to illustrate the soundness of axiom ($\sharp$5). Let $\mathcal{M} = (W, R, V)$ be a belief model, let $w \in W$ and let ϕ, ψ be any $\mathcal{L}_{PL}$-formula.

From left to right. Suppose $\mathcal{M}, w \models [\sharp\phi]B_a^{\smile}\psi$ and $\mathcal{M}, w \models \phi \vee B_a\neg\phi$. Then either $\mathcal{M}, w \models \phi$ or $\mathcal{M}, w \models B_a\neg\phi$. Consider any $v \in W$ such that $vR_a w$. If $\mathcal{M}, w \models \phi$, then using Definition 11 we can obtain $vR_a^{\sharp\phi} w$. If $\mathcal{M}, w \models B_a\neg\phi$, then there is no $u \in R_a(w)$ such that $\mathcal{M}, u \models \phi$, and hence using Definition 11 again we have $vR_a^{\sharp\phi} w$. Either implies $vR_a^{\sharp\phi} w$, and since $\mathcal{M}, w \models [\sharp\phi]B_a^{\smile}\psi$, we know that $\mathcal{M}^{\sharp\phi}, v \models \psi$. It follows that $\mathcal{M}, v \models [\sharp\phi]\psi$ for each $v \in W$ such that $vR_a w$. Therefore $\mathcal{M}, v \models B_a^{\smile}[\sharp\phi]\psi$.

From right to left. Suppose $\mathcal{M}, w \models \phi \vee B_a \neg\phi \to B_a^{\smile}[\sharp\phi]\psi$. Then either $\mathcal{M}, w \models \neg\phi \wedge \neg B_a \neg\phi$ or $\mathcal{M}, w \models B_a^{\smile}[\sharp\phi]\psi$.

Suppose the $\mathcal{M}, w \models \neg\phi \wedge \neg B_a \neg\phi$. Consider any $v \in W$ such that $vR_a w$. Then $\mathcal{M}, w \models \neg\phi$ and there is a $u \in R_a(w)$ such that $\mathcal{M}, u \models \phi$, which implies, using Definition 11, that not $vR_a^{\sharp\phi} w$. If follows that there is no $v \in W$ such that $vR_a^{\sharp\phi} w$, and hence $\mathcal{M}, w \models [\sharp\phi]B_a^{\smile}\psi$.

Suppose $\mathcal{M}, w \models B_a^{\smile}[\sharp\phi]\psi$. Consider any $v \in W$ such that $vR_a^{\sharp\phi} w$. By Definition 11 we have $vR_a w$, and hence $\mathcal{M}, v \models [\sharp\phi]\psi$. It follows that $\mathcal{M}^{\sharp\phi}, v \models \psi$ for each $v \in W$ such that $vR_a^{\sharp\phi} w$, which implies $\mathcal{M}, w \models [\sharp\phi]B_a^{\smile}\psi$. □

The case for a cautious audience accepting statements consistent with their convictions is formally given by the following proposition.

Proposition 9 *Let $a \in \mathcal{A}g$ be any agent and let ϕ be any boolean formula. Then $\vdash_{PLL} \hat{B}_a\phi \to [\sharp\phi]B_a\phi$.*

Proof Straightforward, by the completeness of calculus PLL and Definition 6.

The following proposition shows that if a statement ϕ has been accepted, any further statement of $\neg\phi$ cannot retract it.

Proposition 10 *Let $a \in \mathcal{A}g$ be any agent, let ϕ be any boolean formula. Then $\vdash_{PLL} \hat{B}_a\phi \to [\sharp\phi][\sharp\neg\phi]B_a\phi$.*

Proof By induction on the construction of ϕ and axioms ($\sharp$1-3), we have $\vdash_{PLL} \phi \leftrightarrow [\sharp\neg\phi]\phi$. Since $\vdash_{PLL} B_a\phi \to B_a\phi$, we can imply $\vdash_{PLL} B_a\phi \to B_a[\sharp\neg\phi]\phi$. Using axiom ($\sharp$4) we have $\vdash_{PLL} B_a\phi \to [\sharp\neg\phi]B_a\phi$. Therefore $\vdash_{PLL} \hat{B}_a\phi \to [\sharp\phi][\sharp\neg\phi]B_a\phi$. □

Using Proposition 9 it is trivial to show that for each boolean formula ϕ:

$$\vdash_{PLL} \bigwedge_{a \in \mathcal{A}g} (\hat{B}_a\phi \wedge \hat{B}_a\neg\phi) \to [\sharp\phi] \bigwedge_{a \in \mathcal{A}g} B_a\phi$$

That is, if the audience all consider both ϕ and $\neg\phi$ are possible, then after announcing ϕ, ϕ will become a mutual conviction. By proposition 10, further announcements cannot retract this mutual belief. Thus it really matters which is announced first, a public lie or a truthful public announcement. Note that this does not hold anymore if instead of mutual conviction we consider common belief, for some of the audience may consider it possible that someone else is certain of $\neg\phi$ before the announcement ϕ and will not be affected by the operation $\sharp\phi$.

4 Recoveries from False Beliefs

In this section, instead of addressing the important but difficult question how public lies can be detected, we focus on a simpler question. What does the process of recovery from false beliefs look like?

When Donald Trump stated that the practice of voting by mail could lead to voter fraud, an FBI official testified that "there is no evidence of a coordinated mail-in voting fraud effort". In this example, the key to recovery is the statement "there is no evidence of $\neg\phi$". We wish to model this retraction from the false belief $\neg\phi$, the public opening of the mind for ϕ again, or, in still other words, the common realization that ϕ might be true.

Adopting the notation $\natural\phi$ for this, the sequence $\natural\phi; !\phi$ models the recovery from the false belief $\neg\phi$ followed by public update with ϕ. Compared to the contraction in AGM belief revision (see Gärdenfors (2003)), $\natural\phi$ does not contract $\neg\phi$ from the audience's belief sets, but contracts the evidence supporting $\neg\phi$, which leaves both ϕ and $\neg\phi$ possibly true. This effect is similar to forgetting (cf. van Ditmarsch et al. (2009)), with the difference that our operation does not erase all evidence either way. Erasure of all evidence for or against ϕ would result in the belief $\hat{B}_a\phi \wedge \hat{B}_a\neg\phi$. The effect of $\natural\phi$, however, is to convince a that $\hat{B}_a\phi$ is true. Thus if a is certain of ϕ, $\natural\phi$ would not make her believe $\hat{B}_a\neg\phi$.

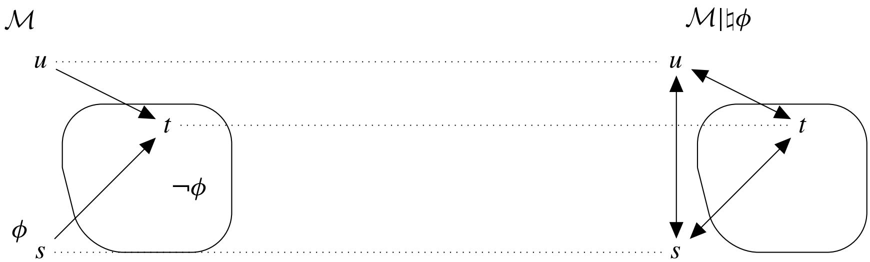
Another related topic is the discussion of reverse public announcement in Balbiani et al. (2016) and Haney (2018). The converse announcement update proposed by Balbiani et al. can be seen as a kind of recovery from public announcement, but it is not deterministic. Haney uses worlds in a canonical model to expand epistemic models in reverse public announcement update. In contrast with this, we use the worlds in an agent's background knowledge to expand belief-cells in recoveries.

Suppose a current belief state is given by relation c. Then the act of recovering from the false belief $\neg\phi$ is given by the relational change

$$c := c \cup (c; c^{\smile}; ?\phi).$$

Explanation: we need to put ϕ situations back into an agents' consideration. If you are in a situation s that ϕ is disbelieved, then you can recover the connection to any ϕ-situation t that is disbelieved by first taking a c step inside the body of the octopus, and next taking a reverse c step out of the octopus again to t. This relational change also affects those who are certain of ϕ, for this act implicitly suggests people to examine all ϕ-situations.

The operation of recovering from the lie that $\neg\phi$ can be pictured as follows:



We have the following key formula for the recovery from the false belief that $\neg p$:

$$[\natural p]B_a q \leftrightarrow B_a q \wedge K_a(p \rightarrow q).$$

What this says is this. After the recovery from the false belief that $\neg p$, $B_a q$ is true if a is certain of q and knows that if p holds then q is true. Note that this opens the way for an axiomatization with reduction axioms.

We can also define public recovery $!\natural\phi$ using the operation $\natural\phi$ with the precondition that ϕ is actually true:

$$[!\natural\phi]\psi ::= \phi \rightarrow [\natural\phi]\psi$$

A truthful recovery ϕ is an operation $\natural\phi$ with the precondition that the speaker is certain of ϕ, and a lying recovery ϕ is an operation $\natural\phi$ with the precondition that the speaker is certain of $\neg\phi$. Since the speaker's belief state is abstracted in our modelling, we use operation $\natural\phi$ for both truthful and lying recovery ϕ, and simply call it recovery.

A successful recovery ϕ for agent a means that a was previously convinced of $\neg\phi$, and after the recovery she believes that ϕ possibly holds. However using axiom $B4$ we know that $B_a\neg\phi$ implies $B_a\neg\hat{B}_a\phi$, and since we assume agents are always cautious about influence on their conviction, how can a accept $\hat{B}_a\phi$? An explanation is that the use of axiom ($B4$) invokes introspection, which is perhaps too strong in real life, and even if someone is aware that $\neg\hat{B}_a\phi$ is in his conviction, he may occasionally let it slip away. Thus if "there is no evidence of ϕ" is announced when he is in a less conscious state, he might think that since $\hat{B}_a\phi$ is consistent with $\neg\phi$, he is better to accept it. In any case, recovery is never an easy task. It is usually performed by authentic people or those trusted by the audience, and it requires either explanation, persuasion or repetition, which are abstracted in our framework.

The language $\mathcal{L}_{LR}$ for public lies and recoveries is language $\mathcal{L}_{PL}$ plus operators $[\natural\phi]$, which gives the following BNF definition:

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid B_a\phi \mid B_a^{\smile}\phi \mid [\sharp\phi]\phi \mid [\natural\phi]\phi.$$

$[\natural\phi]$ can be read as "there is no evidence that ϕ is false", and $[\natural\phi]\psi$ can be interpreted as "after the announcement that there is no evidence that ϕ is false, ψ becomes true". The key clauses of truth conditions for recovery operators $[\natural\phi]$ is given below.

Definition 11 Let $\mathcal{M} = (W, R, V)$ be a belief model, and let ϕ be a $\mathcal{L}_{LR}$-formula. $\mathcal{M}^{\natural\phi} = (W, R^{\natural\phi}, V)$ is the model recovered from $\mathcal{M}$ by $\natural\phi$ if:

$$R_a^{\natural\phi} = R_a \cup \{(w, u) \in R_a \circ R_a^{-1} \mid \mathcal{M}, u \models \phi\}.$$

The key causes of truth conditions for $\mathcal{L}_{LR}$:

$$\mathcal{M}, w \models [\natural\phi]\psi \text{ iff } \mathcal{M}^{\natural\phi}, w \models \psi.$$

Note that the updated relations $R_a^{\natural\phi}$ are still serial, transitive and euclidean. Note that neither public lies nor recoveries change one's background knowledge. This is borne out by the following lemma.

Lemma 12 *Let $\mathcal{M} = (W, R, V)$ be a belief model, let $w \in W$ and let ϕ be any formula. Then $[w]_a = [w]_a^{\natural\phi}$.*

Proof By the above definition, $R_a \subseteq R_a^{\natural\phi}$. Thus it suffices to show that $[w]_a^{\natural\phi} \subseteq [w]_a$, i.e., for each $u \in W$, $uR_a^{\natural\phi}v$ and $wR_a^{\natural\phi}v$ for some $v \in W$ only if uR_av and wR_av for some $v \in W$.

Consider any $u \in [w]_a^{\natural\phi}$. Clearly there is a $v' \in W$ such that $uR_a^{\natural\phi}v'$ and $wR_a^{\natural\phi}v'$. Using Definition 11, we have either wR_av' (that implies $v' \in [w]_a$) or $v' \in [w]_a$, and similarly v' must be in $[u]_a$. It follows that w and u are in the same knowledge cell in $\mathcal{M}$. □

The calculus LRL (for Public Lying and Recovery Logic) is PLL plus the following reduction axioms and rules for $[\natural\phi]$:

$$
\begin{array}{ll}
(\natural 1) & [\natural\phi]p \leftrightarrow p \\
(\natural 2) & [\natural\phi]\neg\psi \leftrightarrow \neg[\natural\phi]\psi \\
(\natural 3) & [\natural\phi](\psi \wedge \chi) \leftrightarrow [\natural\phi]\psi \wedge [\natural\phi]\chi \\
(\natural 4) & [\natural\phi]B_a\psi \leftrightarrow B_a[\natural\phi]\psi \wedge K_a(\phi \rightarrow [\natural\phi]\psi) \\
(\natural 5) & [\natural\phi]B_a^{\smile}\psi \leftrightarrow (\phi \vee \neg B_a^{\smile}\bot \rightarrow K_a[\natural\phi]\psi)
\end{array}
$$

RULES:

$$
\frac{\phi}{[\natural\psi]\phi}\ (\text{Nec-}\natural) \qquad \frac{\phi \leftrightarrow \psi}{[\natural\phi]\chi \leftrightarrow [\natural\psi]\chi}\ (\text{Rep-}\natural)
$$

$(\natural 4)$ parallels the recovery update $\natural\phi$ for R_a in Definition 11. $(\natural 5)$ describes the equivalent condition for $[\natural\phi]B_a^{\smile}\psi$ that is: a will know ψ if a's convictions will become true after recovery ϕ. This condition can be read as "if either ϕ is true or a's convictions are true before the recovery, then a knows that ψ will be true after recovery ϕ".

Theorem 13 *Calculus LRL is sound and complete for belief models.*

Proof Since calculus LRL is PLL plus reduction axioms ($\natural$1-5), it suffices to show that these axioms are sound. We only prove the soundness of $(\natural 4)$ and $(\natural 5)$. Let $\mathcal{M} = (W, R, V)$ be a belief model, let $w \in W$ and let ϕ, ψ be any $\mathcal{L}_{LR}$-formula.

First consider $(\natural 4)$. $\mathcal{M}, w \models [\natural\phi]B_a\psi$ iff

– for each $v \in W$, $wR_a^{\natural\phi}v$ implies $\mathcal{M}^{\natural\phi}, v \models \psi$,

iff, by Definition 11,

– for each $v \in W$, wR_av only if $\mathcal{M}, v \models [\natural\phi]\psi$, furthermore $(w, v) \in R_a \circ R_a^{-1}$ and $\mathcal{M}, v \models \psi$ entails $\mathcal{M}, v \models [\natural\phi]\psi$, iff $\mathcal{M}, w \models B_a[\natural\phi]\psi \wedge K_a(\phi \to [\natural\phi]\psi)$.

Next consider ($\natural$5). From left to right. Suppose $\mathcal{M}, w \models [\natural\phi]B_a^{\smile}\psi$ and $\mathcal{M}, w \models \phi \vee \neg B_a^{\smile}\bot$. We have either $\mathcal{M}, w \models \phi$ or $w \in R_a(w)$. Consider any $v \in [w]_a$. If $\mathcal{M}, w \models \phi$, then using Definition 11 we can obtain that $vR_a^{\natural\phi}w$. If $w \in R_a(w)$, then by Lemma 4 we know that vR_aw, and hence using Definition 11 again we have $vR_a^{\natural\phi}w$. Since either implies $vR_a^{\natural\phi}w$, by $\mathcal{M}, w \models [\natural\phi]B_a^{\smile}\psi$ we can get $\mathcal{M}^{\natural\phi}, v \models \psi$, which implies $\mathcal{M}, v \models [\natural\phi]\psi$. It follows that $\mathcal{M}, v \models K_a[\natural\phi]\psi$.

From right to left. Suppose either $\mathcal{M}, w \models \neg\phi \wedge B_a^{\smile}\bot$ or $\mathcal{M}, w \models K_a[\natural\phi]\psi$. If $\mathcal{M}, w \models \neg\phi \wedge B_a^{\smile}\bot$, then neither $\mathcal{M}, w \models \phi$ nor there is a $v \in W$ such that vR_aw, and hence using Definition 11 there is no $u \in W$ such that $uR_a^{\natural\phi}w$, which vacuously implies that $\mathcal{M}^{\natural\phi}, w \models B_a^{\smile}\psi$. Suppose $\mathcal{M}, w \models K_a[\natural\phi]\psi$. Consider any $u \in W$ such that $uR_a^{\natural\phi}w$. By Definition 11 u is in $[w]_a$. Thus we have $\mathcal{M}, u \models [\natural\phi]\psi$, which implies that $\mathcal{M}^{\natural\phi}, u \models \psi$. It follows that $\mathcal{M}^{\natural\phi}, w \models B_a^{\smile}\psi$. Therefore either implies that $\mathcal{M}^{\natural\phi}, w \models B_a^{\smile}\psi$, and hence $\mathcal{M}, w \models [\natural\phi]B_a^{\smile}\psi$. □

The following proposition illustrates that the sequence $\natural p; !p$ indeed helps the audience recover from the lie that $\neg p$ and convinces them that p.

Proposition 14 *Let $a \in \mathcal{A}g$ be any agent and let ϕ be any boolean formula. Then*

1. $\vdash_{LRL} \hat{K}_a\phi \to [\natural\phi][\sharp\phi]B_a\phi$,
2. $\vdash_{LRL} \hat{K}_a\phi \to [\sharp\neg\phi][\natural\phi][\sharp\phi]B_a\phi$

Proof (1) :

1.	$\vdash_{LRL} \hat{B}_a\phi \to [\sharp\phi]B_a\phi$	Proposition 9
2.	$\vdash_{LRL} [\natural\phi]\hat{B}_a\phi \to [\natural\phi][\sharp\phi]B_a\phi$	1, Axioms ($\natural$ 2,3), Nec-$\natural$
3.	$\vdash_{LRL} \hat{B}_a\phi \vee \hat{K}_a\neg(\phi \to \neg\phi) \to [\natural\phi]\hat{B}_a\phi$	Axioms ($\natural$ 2,4)
4.	$\vdash_{LRL} \hat{K}_a\phi \to \hat{K}_a\neg(\phi \to \neg\phi)$	propositional logic
5.	$\vdash_{LRL} \hat{K}_a\phi \to [\natural\phi][\sharp\phi]B_a\phi$	mp. 2,3,4

(2) immediately follows from (1) if we can prove $\vdash_{LRL} \hat{K}_a\phi \to [\sharp\neg\phi]\hat{K}_a\phi$. This is straightforward from Lemma 7 and Theorem 13.

□

Note that this proposition cannot be generalized for arbitrary formulas. For instance

$$\hat{K}_aB_a^{\smile}\bot \to [\natural B_a^{\smile}\bot][\sharp B_a^{\smile}\bot]B_aB_a^{\smile}\bot$$

is not valid in LRL, for $B_aB_a^{\smile}\bot$ (which is $K_a\bot$) is always false in belief models.

As an example of the effects of recoveries, suppose a tribe is facing the coordination game Stag Hunt[1]. Everyone in the tribe has two options: to hunt for a stag together (STAG) or to capture a hare by themselves (HARE). Those who go hunting for hares can each get one hare, but it is better for the tribe to hunt stag together, as a stag can

[1] Stag Hunt is a stock example from game theory, about the coordination required in hunting a stag together. This requires conventions, common knowledge and common beliefs. We will not discuss these notions in the paper, but we will use the example to illustrate the effects of public lies and recoveries on mutual beliefs.

provide much more food for everyone. The snag is that for successful stag-hunting everyone has to join in. If anyone abandons the joint task, the hunt would fail. Thus there are two equilibria for the tribe: STAG or HARE.

Now, to twist this into our own story, suppose that a priestess has decided that the omens are auspicious for stag hunting, and she has ordered two elders to convey her message. As it turns out, one of the elders is dishonest and the other one is honest. The dishonest elder publicly lies that the decision is HARE.

Let p and $\neg p$ be "the decision is STAG" and "the decision is HARE" respectively. Suppose everyone is fooled by the dishonest elder into believing that the priestess has decided HARE. Then a cure (the sequence $\natural p$; $\sharp p$) is to first claim that there is no evidence that the priestess has made her mind to hunt for hares, and then to announce that actually the priestess' decision is STAG. By means of these two steps the tribe can reach a mutual conviction of STAG, as the above proposition implies. We get:

$$\vdash_{LRL} \bigwedge_{a \in \mathcal{Ag}} \hat{K}_a p \to [\sharp\neg p][\natural p][\sharp p] \bigwedge_{a \in \mathcal{Ag}} B_a p.$$

However, the proposition also suggests a more vicious form of "public lying" executed in the very same sequence $\natural\neg p$; $\sharp\neg p$. This tactic can be seen in real life, for instance in cults proclaiming that "science is just another religion" before preaching their own doctrines.

5 Lockean Beliefs and Conditional Beliefs

In the stag hunt game, cooperation requires conventions or common knowledge, which one may assume to be acquired by a public announcement or a public event. But things are more complicated in real life. Chwe (2013) emphasizes the importance of common experience, where everyone is seeing the reactions of the rest of the audience. This makes it crucial to have public gatherings. Monderer and Samet (1989) consider cases where it is probable that not everyone is hearing a communication, and prove that in those situations common p-beliefs can approximate common knowledge, where p is a probability. Binmore (2008) suggests that "most conventions arise gradually and acquire force by a slow progression", and thus that not all conventions need to be common knowledge, and some of them may be the product of social evolution. Because of these considerations it becomes important to investigate the effects of public lies and recoveries on subjective probabilities instead of on KD45 beliefs. In this section, we take steps in that direction.

We will focus on a kind of very simple belief operators $\mathcal{P}_a$. $\mathcal{P}_a\phi$ is true if a's subjective probability of ϕ is greater than 0.5 or a is willing to bet ϕ. These 0.5-beliefs (Monderer and Samet (1989)) are related to what Foley (1992) calls the *Lockean thesis*, and we will call them *Lockean beliefs*. For related work, see Hamblin (1959), Burgess (1969) and Herzig (2003). Herzig and Longin (2003) give a system of Lockean beliefs and KD45 beliefs on neighbourhood models. Ghosh and de Jongh (2013) present, among many other systems, a logic of these two kind of beliefs for plausibility models.

Lockean belief operators and probability models are discussed by van Eijck and Renne (2016).

Many of our daily decisions are based on this kind of belief. For instance if it has been raining for days, Bob may think that it is likely to rain tomorrow. If the weather forecast says that it will not rain, he may believe that there is no need to take an umbrella to work tomorrow. Neither of these beliefs is a KD45 belief. Since the weather forecast only provides predictions, the belief based on the forecast should be interpreted as a conditional belief, just like the belief based on the observation of rain today, which is also a conditional belief, with the observation of the recent state of the weather as an implicit condition.

Conditional beliefs can be interpreted on plausibility models (cf. Baltag and Smets (2006), Pacuit (2013)). Demey (2013) studies public announcement on such models. Another way to represent conditional beliefs is by means of neighbourhood functions, as in van Eijck and Li (2017) and Marianna et al. (2018). If A is the proposition that it will rain tomorrow, then Bob's belief about A can be represented as A in his neighbourhood N_b. If B is the proposition that the weather forecast says it will not rain tomorrow, we can assign the proposition that Bob will not take an umbrella (C) in his neighbourhood with condition B ($C \in N_b(B)$).

Assume p ranges over a set of proposition letters P, and $a \in \mathcal{Ag}$. The language for conditional neighbourhood logic $\mathcal{L}_{CN}$ is our basic language plus binary operators $\mathcal{C}_a$, which is given by the following BNF definition:

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid B_a\phi \mid B_a^{\smile}\phi \mid \mathcal{C}_a(\phi, \phi)$$

$\mathcal{C}_a(\phi, \psi)$ can be read as "assuming ϕ, agent a is willing to bet ψ against $\neg\psi$".

Definition 15 Let $\mathcal{Ag}$ be a finite set of agents. A *conditional neighbourhood model* $\mathcal{M}$ is a tuple (W, R, N, V) where

- (W, R, V) is a belief model;
- $N : \mathcal{Ag} \times W \times \mathscr{P}W \to \mathscr{P}\mathscr{P}W$ is a function that assigns to every agent $a \in \mathcal{Ag}$, every world $w \in W$ and set of worlds $X \subseteq W$ a collection $N_a^w(X)$ of sets of worlds–each such set called a neighbourhood of X–subject to the following conditions:

  (c) $\forall Y \in N_a^w(X) : Y \subseteq X \cap [w]_a$.
  (ec) $\forall Y \subseteq W$: if $X \cap [w]_a = Y \cap [w]_a$, then $N_a^w(X) = N_a^w(Y)$.
  (d) $\forall Y \in N_a^w(X)$, $X \cap [w]_a - Y \notin N_a^w(X)$.
  (sc) $\forall Y, Z \subseteq X \cap [w]_a$: if $X \cap [w]_a - Y \notin N_a^w(X)$ and $Y \subsetneq Z$, then $Z \in N_a^w(X)$.

We call N a neighbourhood function; a neighbourhood $N_a^w(X)$ for agent a in w, conditioned by X is a set of propositions each of which agent a believes more likely to be true than its complement.

Property (c) expresses that what is believed is also known; (ec) expresses **equivalence of conditions**, i.e., if an agent knows that two conditions are equivalent, then the agent's beliefs are the same under both conditions; (d) expresses "determinacy": an agent does not believe both a proposition and its complement; (sc) expresses a form of "strong commitment": if the agent does not believe the complement of Y then she must believe any weaker Z implied by Y. It was proved in van Eijck and Li (2017) that neighbourhood functions also satisfy the following, for any $a \in A$, $w \in W$, $X \subseteq W$:

(m) $\forall Y \subseteq Z \subseteq X \cap [w]_a$: if $Y \in N_a^w(X)$, then $Z \in N_a^w(X)$;
(ni) $\emptyset \notin N_a^w(X)$;
(n)* if $X \cap [w]_a \neq \emptyset$, then $X \cap [w]_a \in N_a^w(X)$;
($\emptyset$) if $X \cap [w]_a = \emptyset$, then $N_a^w(X) = \emptyset$;

where (m) and (ni) expresses **monotonicity** and **no-inconsistency** (an agent does not hold an inconsistent belief) respectively. ($\emptyset$) expresses that conditioning with information that contradicts what the agent knows will cause an agent to believe nothing anymore.

Let $\mathcal{M} = (W, R, N, V)$ be a conditional neighbourhood model, let $w \in W$. Then the key clauses of truth conditions are given by:

$$\mathcal{M}, w \models \mathcal{C}_a(\phi, \psi) \text{ iff for some } Y \in N_a^w([\![\phi]\!] \cap [w]_a), Y \subseteq [\![\psi]\!] .$$

As the Lockean belief $\mathcal{P}_a\phi$ is interpreted as "a is willing to bet ϕ", this decision on ϕ should be based on what a is certain of, namely her conviction. Thus using the above definition, Lockean belief operators $\mathcal{P}_a$ can be given by:

$$\mathcal{P}_a\phi ::= \mathcal{C}_a(\neg \breve{B}_a \bot, \phi).$$

$\mathcal{P}_a\phi$ expresses our intuition that "given what a is certain of, a is willing to bet ϕ", and we can establish the following equivalence:

$$\mathcal{M}, w \models \mathcal{P}_a\phi \text{ iff for some } Y \in N_a^w(R_a(w)), Y \subseteq [\![\psi]\!] .$$

We can also give a complete calculus CNL (conditional neighbourhood logic) for conditional neighbourhood models, which is calculus CBL plus the following axioms.

(5B) $\mathcal{C}_a(\phi, \psi) \to K_a\mathcal{C}_a(\phi, \psi)$
(4B) $\neg\mathcal{C}_a(\phi, \psi) \to K_a\neg\mathcal{C}_a(\phi, \psi)$
(D) $\mathcal{C}_a(\phi, \psi) \to \neg\mathcal{C}_a(\phi, \neg\psi)$
(EC) $K_a(\phi \leftrightarrow \psi) \to \mathcal{C}_a(\phi, \chi) \to \mathcal{C}_a(\psi, \chi)$
(M) $K_a(\phi \to \psi) \to \mathcal{C}_a(\chi, \phi) \to \mathcal{C}_a(\chi, \psi)$
(C) $\mathcal{C}_a(\phi, \psi) \to \mathcal{C}_a(\phi, \phi \wedge \psi)$
(SC) $\neg\mathcal{C}_a(\chi, \neg\phi) \wedge \hat{K}_a(\neg\phi \wedge \psi) \to \mathcal{C}_a(\chi, \phi \vee \psi)$

Axiom (D) guarantees the truth of neighbourhood condition (d), (EC) would correspond to (ec), (M) to (m), (C) to (c) and (SC) to (sc). Using Theorem 5 and the completeness result in van Eijck and Li (2017), we can easily derive the following completeness theorem.

Theorem 16 *The calculus CNL for Conditional Neighbourhood logic given above is sound and complete for conditional neighbourhood models.*

After adding public lying operators and recovery operators into $\mathcal{L}_{CN}$, we can also define public lies and recoveries for conditional neighbourhood models.

Definition 17 let $\mathcal{M} = (W, R, N, V)$ be a conditional neighbourhood model, and let ϕ be any formula. $\mathcal{M}^{\sharp\phi} = \left(W^{\sharp\phi}, R^{\sharp\phi}, N^{\sharp\phi}, V^{\sharp\phi}\right)$ is the model updated from $\mathcal{M}$ by the public lie that ϕ if

- $\left(W^{\sharp\phi}, R^{\sharp\phi}, V^{\sharp\phi}\right)$ is the model updated from (W, R, V) by the public lie ϕ,
- $N^{\sharp\phi} = N$.

$\mathcal{M}^{\natural\phi} = \left(W^{\natural\phi}, R^{\natural\phi}, N^{\natural\phi}, V^{\natural\phi}\right)$ is the model updated from $\mathcal{M}$ by recovery ϕ if

- $\left(W^{\natural\phi}, R^{\natural\phi}, V^{\natural\phi}\right)$ is the model recovered from (W, R, V) by ϕ,
- $N^{\natural\phi} = N$.

Note that this definition relies on the fact that public lies and recoveries have no effect on an agent's background knowledge. Also note that the definition of the neighbourhood function does not depend on KD45 beliefs.

The truth conditions for our two kinds of dynamic operators are defined as usual. This "dynamic version" of CNL (let's call it calculus CND) is CNL plus PLL and the following two reduction axioms for $\mathcal{C}_a$.

$$\begin{array}{ll} (\sharp\mathcal{C}) & [\sharp\phi]\mathcal{C}_a(\psi, \chi) \leftrightarrow \mathcal{C}_a([\sharp\phi]\psi, [\sharp\phi]\chi) \\ (\natural\mathcal{C}) & [\natural\phi]\mathcal{C}_a(\psi, \chi) \leftrightarrow \mathcal{C}_a([\natural\phi]\psi, [\natural\phi]\chi) \end{array}$$

Theorem 18 *Calculus CND is sound and complete for conditional neighbourhood models.*

Proof First consider axiom $(\sharp\mathcal{C})$. $\mathcal{M}, w \models [\sharp\phi]\mathcal{C}_a(\psi, \chi)$ iff

- $([w]_a^{\sharp\phi} \cap [\![\chi]\!]_{\mathcal{M}^{\sharp\phi}}) \in N_a^{w\sharp\phi}([w]_a^{\sharp\phi} \cap [\![\psi]\!]_{\mathcal{M}^{\sharp\phi}})$,

iff, by Lemma 7 and Definition 17,

- $([w]_a \cap [\![[\sharp\phi]\chi]\!]_{\mathcal{M}}) \in N_a^w([w]_a \cap [\![[\sharp\phi]\psi]\!]_{\mathcal{M}})$,

iff $\mathcal{M}, w \models \mathcal{C}_a([\sharp\phi]\psi, [\sharp\phi]\chi)$.

For axiom $(\natural\mathcal{C})$ similarly using Lemma 12 and Definition 17, we can obtain that $\mathcal{M}, w \models [\natural\phi]\mathcal{C}_a(\psi, \chi)$ iff $\mathcal{M}, w \models \mathcal{C}_a([\natural\phi]\psi, [\natural\phi]\chi)$. □

Our next proposition shows that every conviction ϕ is also a Lockean belief.

Proposition 19 *For each agent $a \in \mathcal{A}g$ and each formula $\phi \in \mathcal{L}_{CN}$, $\vdash_{CND} B_a(\phi) \to \mathcal{P}_a(\phi)$.*

Proof By (N)* we know that for each world w in the domain, $R_a(w)$ is always in the neighbourhood $N_a^w(R_a(w))$. Thus the formula, which is equivalent to $B_a(\phi) \to \mathcal{C}_a(\neg B_a^{\smile} \bot, \phi)$, is valid. Use the completeness of CND to obtain its derivability. □

Effects of Public Lies and recoveries on Lockean Beliefs

Using the completeness result of CND, we can show that the effects of public lying on Lockean beliefs are similar to the effects on convictions, i.e., if a's convictions are consistent with a boolean formula ϕ, then publicly lying that ϕ would make a become willing to bet ϕ against its negation, and it cannot be undone by announcing not ϕ. To see why it holds, first notice that by Proposition 9 and 10 we have $\vdash_{CND} \hat{B}_a\phi \rightarrow [\sharp\phi]B_a(\phi) \wedge [\sharp\phi][!\neg\phi]B_a(\phi)$, and using Proposition 19 we can establish $\vdash_{CND} \hat{B}_a\phi \rightarrow [\sharp\phi]\mathcal{P}_a(\phi) \wedge [\sharp\phi][!\neg\phi]\mathcal{P}_a(\phi)$.

Can we also deduce that if the truth $\neg\phi$ is the first to be publicly announced, then public lying ϕ will not affect one's Lockean beliefs? Yes, if $\neg\phi$ can truly be publicly announced. However there is a difference between the facts we can observe and the propositions inferred from those facts. What is inferred from a fact may not be a fact, and thus such inferred propositions may not be announced by truth tellers. But usually it is the inferred propositions that really matter for one's decision, and people may infer differently from truth tellers.

The importance of "carefully reasoned prior probability" was already mentioned by MacKay (2003). The movie *The Big Short* that tells the story of the unfolding of the financial crisis provides another illustration of this. Bear Stearns stock has fallen more than 38% and everyone is pessimistic. Well, almost everyone is, for Bruce Miller, a bullish investor, still believes that he should buy more stock. Because truth tellers cannot announce their conclusions like "Bear Stearns will go bankrupt" as facts, they will not be able to persuade Miller. But liars have no such restriction. They can announce "Bear Stearns will not go bankrupt" as a fact. Thus liars can directly affect their audience's Lockean beliefs, while truth tellers can only hope that when presenting the facts, their audience will draw the appropriate conclusion. This gives liars an advantage over truth tellers.

To model this, we treat the announcements by liars and truth tellers as agent announcements, as in van Ditmarsch (2014). We assume the liar and truth teller have the same epistemic status, and use e (the elders in the tribe) for either of them. Truthful public announcement $[!\sharp\phi]\psi$ (with the precondition that the announced proposition is what the truth teller is convinced of) is given by:

$$[!\sharp\phi]\psi ::= B_e\phi \rightarrow [\sharp\phi]\psi.$$

Thus the ideal process of rational investigation described by MacKay can be approximated as follows. After announcing the evidence D, each agent a will examine whether $\mathcal{P}_a S$ is true.[2] If the truth teller is certain that from the evidence D it is rational and scientific to infer S being more likely, then naturally by the act $!\sharp D$ she is also expecting that each agent a will endorse $\mathcal{P}_a S$ (until being confronted by people like Miller who do not accept the evidence).

[2] Lockean beliefs are not representing likelihood ratios, but posterior odds: $P(S|D, H)/P(\bar{S}|D, H)$, which is more related to the topic of beliefs.

Untruthful announcement $[¡\sharp\phi]\psi$ (which is either lying or bluffing in van Ditmarsch (2014)) is defined as:

$$[¡\sharp\phi]\psi ::= \neg B_e\phi \rightarrow [\sharp\phi]\psi.$$

Note that the precondition $\neg B_e\phi$ differs from the requirement of the traditional definition of lying, namely that the liar disbelieves the statement, which is expressed by $B_e\neg\phi$.

Proposition 20 *Let ϕ be any boolean formula, and let $a \in \mathcal{A}g$ be any agent different from e. Then $\vdash_{CND} \neg B_e\phi \wedge \hat{B}_a\phi \rightarrow [¡\sharp\phi]\mathcal{P}_a\phi$, and if $\neg B_e\phi$ is true, $!\sharp\phi$ is not executable.*

Proof Immediate from Propositions 9 and 19. □

To represent the invariance of R_e under the updates of public lies and recoveries, we need two reduction axioms:

- $[\sharp\phi]B_e\psi \leftrightarrow B_e[\sharp\phi]\psi$
- $[\natural\phi]B_e\psi \leftrightarrow B_e[\natural\phi]\psi$

The completeness proof is just routine. Proofs for the first axiom can be found in both Steiner (2006) and van Ditmarsch (2014).

As an example, let us go back to the tribe that is about to decide on a stag hunt, where there is also a gap to be bridged between "the decision (of the priestess) is STAG" (p) and "the tribe will perform STAG" (q). As q is still not settled, $\neg B_e q \wedge \hat{B}_e q$ holds for both elders. Suppose a tribesman, say Bob (b), believes that the tribe is so disorganized that even if the decision is STAG, they are still very likely to hunt hares instead. Thus if p is announced by the honest elder, Bob would still bet $\neg q$.

Usually, when trying to persuade our audience, we either state our own judgement or that of other people. So what if the honest elder announces her bet that q is true (i.e., $\mathcal{P}_e q$)? This will work if $\mathcal{C}_b(\mathcal{P}_e q, q)$ holds, namely Bob accepts her Lockean belief that q, but it will not work if Bob is certain of $p \rightarrow \mathcal{P}_e q$, that if p holds, the elders are willing to bet q. The following proposition illustrates the effects of announcing other agent's beliefs.

Proposition 21 *Let $a \in \mathcal{A}g \cup \{e\}$ and $b \in \mathcal{A}g$ be two distinct agents, let $\mathcal{O}$ be either B or $\mathcal{P}$, and let ϕ be any boolean formula. Then $\vdash_{CND} \hat{B}_b\mathcal{O}_a\phi \rightarrow ([\sharp\mathcal{O}_a\phi]\mathcal{P}_b\phi \leftrightarrow \mathcal{C}_b(\mathcal{O}_a\phi \wedge \neg B_b^{\smile}\bot, \phi))$.*

Proof Because $\mathcal{P}_b\phi$ is $\mathcal{C}_b(\neg B_b^{\smile}\bot, \phi)$, using Definition 6, 17 and Theorem 18 we can obtain the conclusion. □

Next, consider the dishonest elder who can announce $\neg q$, which will result in every tribesmen being convinced of $\neg q$. It follows that, to their best interests, they all should hunt for hares, and thus $\neg q$ will become true. Compare this with the *true lies* in Agotnes et al. (2018), where a true lie is a formula ϕ satisfying $\neg\phi \rightarrow [¡\sharp\phi]\phi$, which may suggest $\neg q$ is a true lie. However there is a slight difference: from a deterministic perspective, $\neg q$ may not be false at the true history (past, present and future); it can only be false if we cannot remember our history. In our perspective, a more suitable candidate for a true lie would be the statement "it is inevitable that the tribe will hunt for hares", but the formal analysis of that is beyond our current scope.

As for recovery updates, we will show that the lying sequence $\natural\phi; \sharp\phi$ for Lockean beliefs is still as deleterious as for conviction.

Proposition 22 *Let $a \in \mathcal{Ag}$ be any agent and ϕ be any boolean formula. Then $\vdash_{CND} \hat{K}_a\phi \rightarrow [\natural\phi][\sharp\phi]\mathcal{P}_a\phi$.*

Proof Since Proposition 14 also holds for calculus CND, we know that $\vdash_{CND} \hat{K}_a\phi \rightarrow [\natural\phi][\sharp\phi]B_a\phi$. Then by Proposition 19, we have $\vdash_{CND} [\natural\phi][\sharp\phi]B_a\phi \rightarrow [\natural\phi][\sharp\phi]\mathcal{P}_a\phi$. Therefore $\vdash_{CND} \hat{K}_a\phi \rightarrow [\natural\phi][\sharp\phi]\mathcal{P}_a\phi$. □

With an abuse of notation, the truthful recovery $[!\natural\phi]$ and untruthful recovery $[¡\natural\phi]$ can be introduced as follows:

- $[!\natural\phi]\psi ::= B_e\phi \rightarrow [\natural\phi]\psi$
- $[¡\natural\phi]\psi ::= \neg B_e\phi \rightarrow [\natural\phi]\psi$

Using the above proposition we can easily verify that:

Proposition 23 *Let ϕ be any boolean formula, and let $a \in \mathcal{Ag}$ be any agent different from e. Then $\vdash_{CND} \neg B_e\phi \wedge \hat{K}_a\phi \rightarrow [¡\natural\phi][¡\sharp\phi]\mathcal{P}_a\phi$, and if $\neg B_e\phi$ is true, $!\natural\phi$ is not executable.*

Again consider the tribal stag hunt. Recall p and q are "the decision is STAG" and "the tribe will do STAG". Using Proposition 23, it is easy to check that the lying sequence $¡\natural\neg q; ¡\neg q$ makes everyone be willing to bet $\neg q$. Because $!\natural q$ is not viable for the honest elder, the best she can do is to execute the sequence $!\natural p; !p$. Then there will be a mutual conviction of p. However after this Bob, the pessimistic tribesman who holds $\mathcal{C}_a(p, \neg q)$, will still bet $\neg q$.

Executing $!\natural\mathcal{P}_e q$ will still be in vain if Bob knows that the elder is willing to bet q only if p holds. Our final proposition illustrates the effect of recovery sequence $\natural; \sharp$ when announcing the speaker's beliefs, that is to make the recovery sequence successful, the listener should agree with the speaker's belief.

Proposition 24 *Let $b \in \mathcal{Ag}$ be any agent, let $\mathcal{O}$ be either B or $\mathcal{P}$, and let ϕ be any boolean formula. Then $\vdash_{CND} \hat{K}_b\mathcal{O}_e\phi \rightarrow ([\natural\mathcal{O}_e\phi][\sharp\mathcal{O}_e\phi]\mathcal{P}_b\phi \leftrightarrow \mathcal{C}_b(\mathcal{O}_e\phi, \phi))$.*

Proof The proof is similar to Proposition 21: Use Definition 17,[3] Theorem 18 and the four reduction axioms for B_e. □

This suggests that liars continue to have an advantage over truth tellers. Is there a remedy for this at all? Would imperatives like "all hunt for stag, and you go *now*!" prevent people from adopting unreasonable prior probabilities? Maybe we should conclude with the truism that there is no substitute for education.

[3] A modification of Definition 17 has to be made that R_e does not change after either update, public lies or recoveries.

Comparison Between KD45 Beliefs and Lockean Beliefs

Liars also have an advantage over truth tellers in both KD45 models, as in conditional neighbourhood models. This is because the asymmetry between the preconditions of $!\sharp$ and $¡\sharp$ does not rely on neighbourhood semantics. Nevertheless, we conclude this section to show that neighbourhood semantics could be a better candidate to model the influence of authoritative opinions to public decisions.

Let us reconsider the previous stag hunt example informally. Suppose the message given by the priestess is too vague and ambiguous that it is common knowledge that no one knows the truth value of p (the decision of the priestess is STAG). The elders have more experiences in interpreting the priestess's message than tribesmen. They both tend to believe p is true, but are not certain, and they do not know each other's epistemic status. By tradition in this situation after the message is announced in the gathering, there will be a vote to decide whether to perform STAG or HARE. We use s for "the tribe should perform STAG". Thus both $B_a s$ and $\mathcal{P}_a s$ can express that a votes for STAG.

The dishonest elder wants to manipulate the vote so that the tribe will hunt for hares. While he is not in the position to tell what the tribesmen should do, his has two options: claiming $\neg p$ is a fact or announcing he is certain of $\neg p$. The former is not effective because everyone knows the message is too vague even for him. However the latter can hardly be opposed to even by the honest elder, for one should be free to speak his mind. Let a be a tribesman. We will examine on what conditions $[\sharp B_e\neg p]B_a\neg s$ and $[\sharp B_e\neg p]\mathcal{P}_a\neg s$ hold respectively.

First consider $[\sharp B_e\neg p]B_a\neg s$. It can be checked that this formula is entailed by two propositions: $B_a(B_e\neg p \to \neg p)$ and $[\sharp B_e\neg p](B_a\neg p \to B_a\neg s)$. The first expresses that after knowing e's conviction of $\neg p$, a is convinced of $\neg p$. The second can be interpreted as after the announcement of the elder's conviction, a is certain of $\neg p$ only if he is certain of $\neg s$. Thus in order to manipulate the vote, the dishonest elder should make sure the tribesmen will blindly follow his "conviction". This is a strong condition, especially when one realizes that the elder's conviction is only his interpretation, which may be wrong.

Next consider $[\sharp B_e\neg p]\mathcal{P}_a\neg q$. Using Proposition 23, it is entailed also by two propositions: $\mathcal{C}_a(B_e\neg p \wedge \neg \breve{B}_b \bot, \neg p)$ and $[\sharp B_e\neg p](\mathcal{P}_a\neg p \to \mathcal{P}_a\neg s)$. The first says that after knowing e's conviction of $\neg p$, a is willing to bet $\neg p$. The second expresses that after knowing e's conviction of $\neg p$, a is willing to bet $\neg p$ only if he is willing to bet $\neg s$. In other words, a takes the elder's conviction as an advice, which is common in real life. Whenever a piece of news, a policy or even a sign is important but hard to fully understand, we are likely to consult experts or read relevant analyses to form our own judgments, which in turn would guide our decisions or actions.

Comparison of the two cases shows that requirements for manipulating Lockean beliefs are weaker than those for manipulating KD45 beliefs. Since public lying has to involve forms of belief manipulation, this suggests that neighbourhood semantics is perhaps more suitable than KD45 semantics for modeling the effects of public lies on public opinions.

Conclusion and Further Work

We have modelled the effects of public lies on KD45 beliefs, and after introducing the reverse belief operators, we have also provided recoveries of false beliefs, and we have axiomatized these updates. By first executing recovery update then public announcement, an audience can be made to recover from false beliefs. However, similar tactics can be used by liars, so that liars still can deceive cautious audiences.

Next, we have investigated public lies and recoveries on conditional beliefs and Lockean beliefs. The reduction axioms for these updates turned out to be straightforward. Again, the analysis shows that those who do not stick to the truth have an advantage over truth tellers.

We end with some suggestions for further work. An obvious step would be to give a calculus for public lying and recovery in a language $\mathcal{L}_{LB}$ for probability models, and show soundness and completeness.

As was mentioned above, common knowledge and common beliefs play important roles in cooperation. What are sound axioms for public lying and recoveries from false beliefs for a language with common knowledge and common belief operators, and how can we show completeness?

The key effect of public lying is that the community of agents loses touch with reality. This is detrimental for all agents, because the utilities of our actions in the world are determined by properties of the world, not by what agents *believe* about the world. To work this out formally, we need to add agent-utilities, and use these to model the effects on individual agents when these agents act on false beliefs. This would allow us to connect up to Paolo Galeazzi's world (Galeazzi (2017)).

We did not present a very detailed analysis of Stag Hunt, for we only used the game to illustrate flawed communication. For a full analysis of how cooperation is achieved in the game, one has to take common knowledge and utilities into account. In the preparation for the stag hunt, individual tribesmen will not cooperate unless others do so. So it is natural to assume that the more people an agent *believes* will cooperate, the more she is willing to take part. All of this is yet beyond the scope of our framework.

For suppose Alice is the only one in the group who firmly believes the decision is STAG ($B_a p$), and is willing to bet STAG ($\mathcal{P}_a q$). However, after a public lie that $\neg p$, everyone else would become certain of $\neg p$, and most of these believers would be willing to bet $\neg q$. In our framework, Alice's convictions cannot be affected by the lie $\neg p$, and thus Alice will still bet q ($\mathcal{P}_a q$). However as Alice is aware most of others are willing to bet $\neg q$, actually she should become pessimistic about cooperation ($\mathcal{P}_a \neg q$). How can we extend our framework to represent this?

Declarations

Conflicts of interest The authors have no financial or proprietary interests in any material discussed in this article.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

Agotnes, T., van Ditmarsch, H., & Wang, Y. (2018). True lies. *Synthese, 195*(10), 4581–4615.

Arendt, H. (1967 (Penguin Classics Edition, 2006)). Truth and politics. In Between Past and Future—Six Exercises in Political Thought, Viking Press.

Balbiani, P., Van Ditmarsch, H., & Herzig, A. (2016). Before announcement. In: 11th conference on Advances in Modal logic (AiML 2016), Budapest, Hungary, pp. 58–77, https://hal.archives-ouvertes.fr/hal-01650180.

Baltag, A., & Smets, S. (2006). Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electron Notes Theory of Computer Science, 165,* 5–21. https://doi.org/10.1016/j.entcs.2006.05.034.

Binmore, K. (2008). Do conventions need to be common knowledge? *Topoi, 27*(1–2), 17.

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge University Press.

Burgess, J. P. (1969). Probability logic. *Journal of Symbol Log, 34*(2), 264–274. https://doi.org/10.2307/2271103.

Chwe, M. S. Y. (2013). *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.

Demey, L. (2013). Contemporary epistemic logic and the lockean thesis. *Foundations of Science, 18*(4), 599–610.

Foley, R. (1992). The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly, 29*(2), 111–124.

Galeazzi, P. (2017). Play without regret. PhD thesis, ILLC, University of Amsterdam.

Gärdenfors, P. (2003). *Belief revision* (Vol. 29). Cambridge University Press.

Ghosh, S., & de Jongh, D. (2013). Comparing strengths of beliefs explicitly. *Logic Journal of the IGPL, 21*(3), 488–514. https://doi.org/10.1093/jigpal/jzs050.

Hamblin, C. L. (1959). The modal 'probably'. Mind *68*(270), 234–240. http://www.jstor.org/stable/2251572.

Haney, R. S. (2018). Reverse public announcement operators on expanded models. *Journal of Logic, Language and Information, 27*(3), 205–224.

Herzig, A., & Longin, D. (2003). On modal probability and belief. In: European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Springer, pp 62–73.

Herzig, A. (2003). Modal probability, belief, and actions. *Fundamenta Informaticae, 57*(2–4), 323–344.

Kooi, B., & Renne, B. (2011). Arrow update logic. *Review of Symbol Log, 4*(4), 536–559.

MacKay, D. J. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press, available from http://www.inference.phy.cam.ac.uk/mackay/itila/.

Mahon, J. E. (2016). The Definition of Lying and Deception. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy, winter* (2016th ed.). Metaphysics Research Lab, Stanford University.

Marianna, G., Sara, N., Nicola, O., & Vincent, R. (2018). Conditional beliefs: from neighbourhood semantics to sequent calculus. Review of Symbolic Logic pp 1–44.

Monderer, D., & Samet, D. (1989). Approximating common knowledge with common beliefs. *Games and Economic Behavior, 1*(2), 170–190.

Pacuit, E. (2013). Dynamic epistemic logic i: Modeling knowledge and belief. *Philosophy Compass, 8*(9), 798–814.

Plaza, J. (1989). Logics of public communications. In: M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, Z. W. Ras (Eds.), In: Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, pp 201–216.

Sakama, C., Caminada, M., & Herzig, A. (2010). A logical account of lying. In T. Janhunen & I. Niemelä (Eds.), *Logics in Artificial Intelligence* (pp. 286–299). Berlin, Heidelberg: Springer Berlin Heidelberg.

Stalnaker, R. C. (2006). On logics of knowledge and belief. *Philosophical Studies, 128*(1), 169–199.

Steiner, D. (2006). A system for consistency preserving belief change. In Proceedings of the Workshop on Rationality and Knowledge, 18th European Summer School in Logic, Language, and Information (ESSLLI), pp 133–144.

van Ditmarsch, H., Herzig, A., Lang, J., & Marquis, P. (2009). Introspective forgetting. In: Australasian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence.

van Ditmarsch, H. (2014). Dynamics of lying. *Synthese, 191*(5), 745–777.

van Ditmarsch, H., van Eijck, J., Sietsma, F., & Wang, Y. (2012). *On the Logic of Lying* (pp. 41–72). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-29326-9_4.

van Eijck, J., & Li, K. (2017). Conditional belief, knowledge and probability. arXiv preprint arXiv:1707.08744.
van Eijck, J., & Renne, B. (2016). Update, probability, knowledge and belief. In L. Belkemishev, S. Demri, & A. Máté (Eds.), *Advances in Modal Logic* (Vol. 11, pp. 551–570). College Publications.