Public Lies and How to Recover From Them

Jan van Eijck[1,2]

[1]Centrum Wiskunde & Informatica (CWI)
Amsterdam

[2]Institute for Logic, Language and Computation (ILLC)
Amsterdam

Jan 24, 2017 — Workshop on Rationality, Logic, and Decisions

University of Amsterdam

Abstract

The talk gives a formal analysis of public lies, explains how public lying is related to public announcement, and describes the process of recovery from public lies. The aim is to give a formal picture of the effects of brainwashing by a repeated stream of public lies.

Outline

Truth, Knowledge and Rational Belief

Outline

Outline

Outline

Outline

Truth, Knowledge and Rational Belief

Hannah Arendt on Truth

It has frequently been noted that the surest result of brainwashing in the long run is a peculiar kind of cynicism, the absolute refusal to believe in the truth of anything, no matter how well it may be established. In other words, the result of a consistent and total substitution of lies for factual truth is not that the lie will now be accepted as truth, and truth be defamed as lie, but that the sense by which we take our bearings in the real world – and the category of truth versus falsehood is among the mental means to this end – is being destroyed.
Hannah Arendt, Truth and Politics, 1967 [Are06]

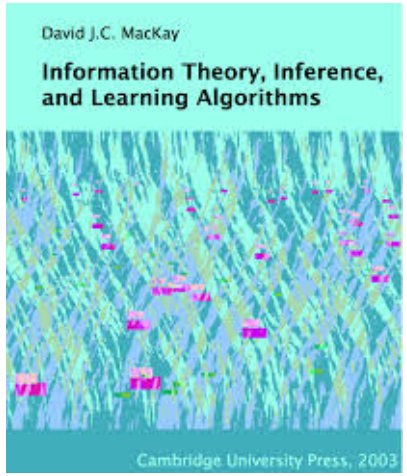
Knowledge, Ignorance, False Belief

- The effect of public lies, according to Hannah Arendt, is that it destroys our bearings in the world.
- Can we explain this formally? For this, we model *public lies* along the same lines as *public announcements*.
- Starting point is the representation of knowledge, ignorance and belief by means of Kripke models.
- For a more refined analysis we can use Kripke models with weights for the various possibilities for what the world could be like.
- Still more refined: different agents may assign different weights to the various possibilities.

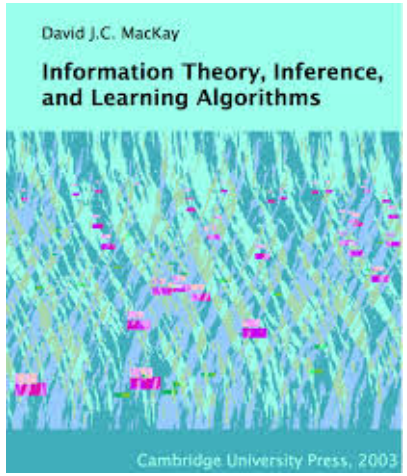
The Effect of Public Lies

- Public announcements and public lies are 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 *recovery from public lies* 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.

Rational Crime Scene Investigation



Rational Crime Scene Investigation



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. [Mac03]

Core Principles of Rational Belief

- Suppose Bernie Sanders had won the Democratic nomination.
- Who would then be president of the USA now?
- No idea, for I *know* that Sanders did not win.
- Suppose ϕ (not in contradiction with anything you know).
- Would you then believe ψ?
- If the world would turn out to be ϕ, would you still believe ψ?
- $B_a(\phi, \psi)$.

- Inspiration: Bayesian update
- Belief as willingness to bet on ψ, given information ϕ.

Representing Uncertainty

- Uncertainty is the set of current options for the actual world.
- Focus on a single fact: the outcome of a coin toss, where the coin 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 i about this situation can be represented as follows:

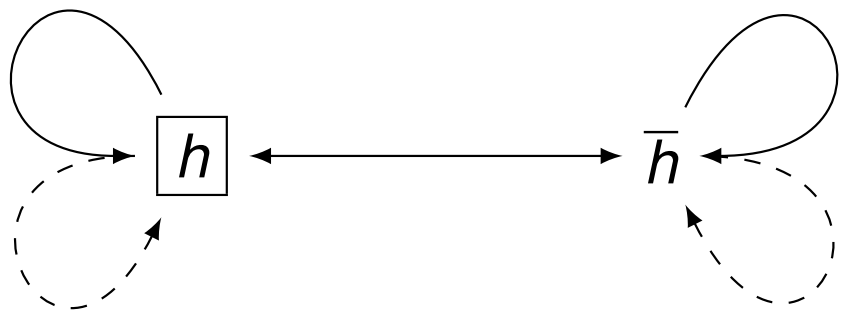
Uncertainty and the Real World

Suppose we also want to represent that *actually*, the coin has landed heads up, but i does not know this yet. Then we can indicate the actual world in the picture, as follows:

- This indication of what is actually the case is invisible to the agent i.
- If a representation for a knowledge situation contains a pointer to the actual world, then this pointer is always invisible to the knowing agents.

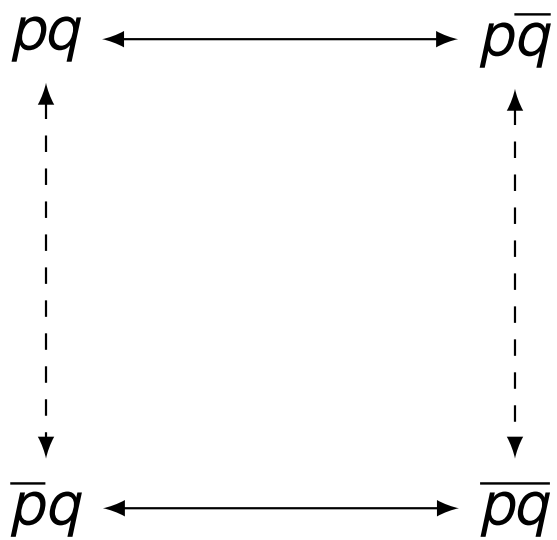
I Know What You do Not Know

There are two agents *i* and *j*, with *i* ignorant about the coin toss outcome, but *j* informed about it.
We use different *accessibility relations* for the two agents, say solid lines for *i* and dashed lines for *j*:

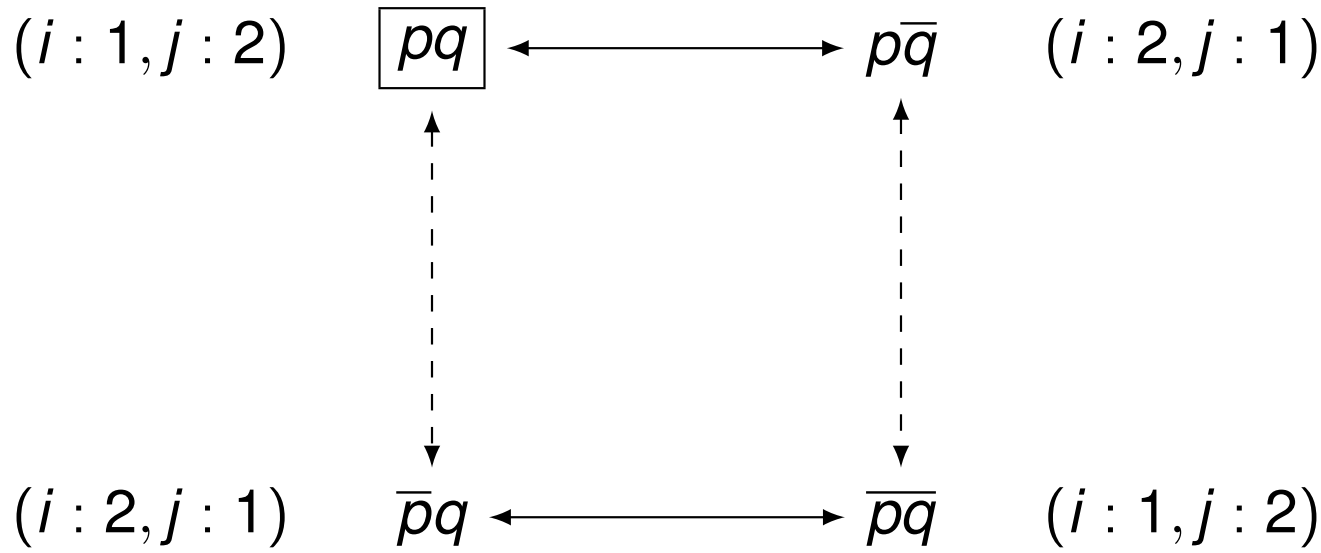


I Know One Thing and You Know Another Thing

We want to picture a situation where i knows the status of p and j the status of q.
Say they both toss a coin, and p denotes heads for i, q denotes heads for j.
We need to distinguish the four possible outcomes, as follows.
For convenience, we leave out the self-loops.



Adding Beliefs



In the real situation:

- i knows p
- i does not believe q (i-subjective probability of q is $\frac{1}{3}$)
- j knows q
- j believes p (j-subjective probability of p is $\frac{2}{3}$)

From Weighted Knowledge Models to Conditional Neighbourhood Models

- Given world w and agent i, $[w]_i$ is the set of worlds that are i-accessible from w.
- The neighbourhoods of w for i, given ϕ, are the subsets X of $Y = [\![\phi]\!] \cap [w]_i$ with the property that the i-weight of X is larger than the i-weight of $Y - X$.

CNL Calculus for Conditional Neighbourhood Logic

(Taut)	All instances of propositional tautologies
(Def-K)	$K_a\phi \leftrightarrow \neg B_a(\neg\phi, \top)$
(Dist-K)	$K_a(\phi \to \psi) \to K_a\phi \to K_a\psi$
(T)	$K_a\phi \to \phi$
(PI-KB)	$B_a(\phi, \psi) \to K_a B_a(\phi, \psi)$
(NI-KB)	$\neg B_a(\phi, \psi) \to K_a \neg B_a(\phi, \psi)$
(M)	$K_a(\phi \to \psi) \to B_a(\chi, \phi) \to B_a(\chi, \psi)$
(EC)	$K_a(\phi \leftrightarrow \psi) \to B_a(\phi, \chi) \to B_a(\psi, \chi)$.
(D)	$B_a(\phi, \psi) \to \neg B_a(\phi, \neg\psi)$
(N)	$B_a(\phi, \psi) \to B_a(\phi, \phi \wedge \psi)$

$$\frac{\phi \to \psi \quad \phi}{\psi}\ \text{(MP)} \qquad \frac{\phi}{K_a\phi}\ \text{(Nec-K)}$$

CNL Calculus for Conditional Neighbourhood Logic

(Taut)	All instances of propositional tautologies
(Def-K)	$K_a\phi \leftrightarrow \neg B_a(\neg\phi, \top)$
(Dist-K)	$K_a(\phi \to \psi) \to K_a\phi \to K_a\psi$
(T)	$K_a\phi \to \phi$
(PI-KB)	$B_a(\phi, \psi) \to K_a B_a(\phi, \psi)$
(NI-KB)	$\neg B_a(\phi, \psi) \to K_a \neg B_a(\phi, \psi)$
(M)	$K_a(\phi \to \psi) \to B_a(\chi, \phi) \to B_a(\chi, \psi)$
(EC)	$K_a(\phi \leftrightarrow \psi) \to B_a(\phi, \chi) \to B_a(\psi, \chi)$.
(D)	$B_a(\phi, \psi) \to \neg B_a(\phi, \neg\psi)$
(N)	$B_a(\phi, \psi) \to B_a(\phi, \phi \wedge \psi)$

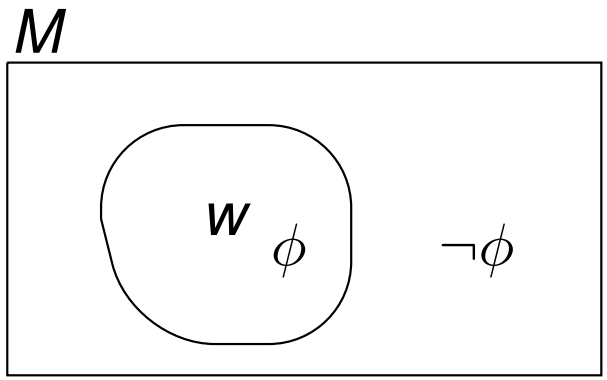
$$\frac{\phi \to \psi \quad \phi}{\psi}\ \text{(MP)} \qquad \frac{\phi}{K_a\phi}\ \text{(Nec-K)}$$

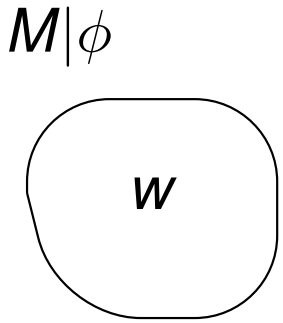
Theorem

This calculus is sound and complete for conditional neighbourhood models, and sound but not complete for weighted knowledge models (S5 Kripke models with weights) [FL16]

Public Announcements

Public Announcements [Pla89, BvDvEJ13]

from 

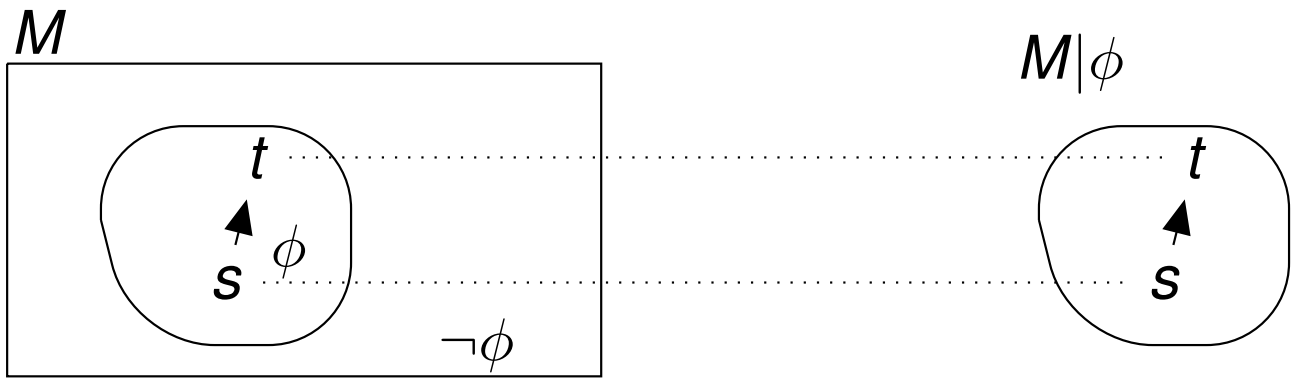
to 

Public Announcement by Cutting Links

- Public announcement of ϕ: cut links between ϕ and $\neg\phi$ situations.
- Precondition: ϕ is true in the real world.
- Relational change: from a to $(?\phi; a; ?\phi)$.
- Maps equivalence relations to equivalence relations.

Key Validity for Public Announcement

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



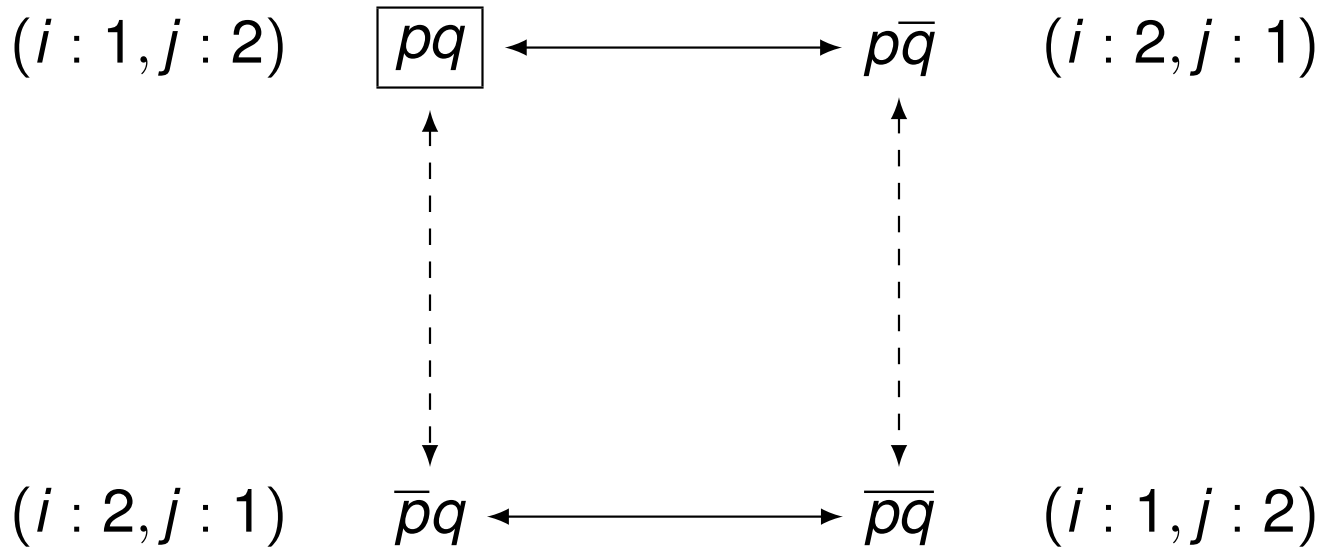
The formula $[!\phi]K_a\psi$ says that, in $M|\phi$, all worlds t that are i-accessible from s satisfy ψ.
The corresponding worlds t in M are those i-accessible from s which satisfy ϕ.
We make our assertion on the right (the assertion about the model after the update) conditional on $!\phi$ being executable, i.e., on ϕ being true.

Public Lies

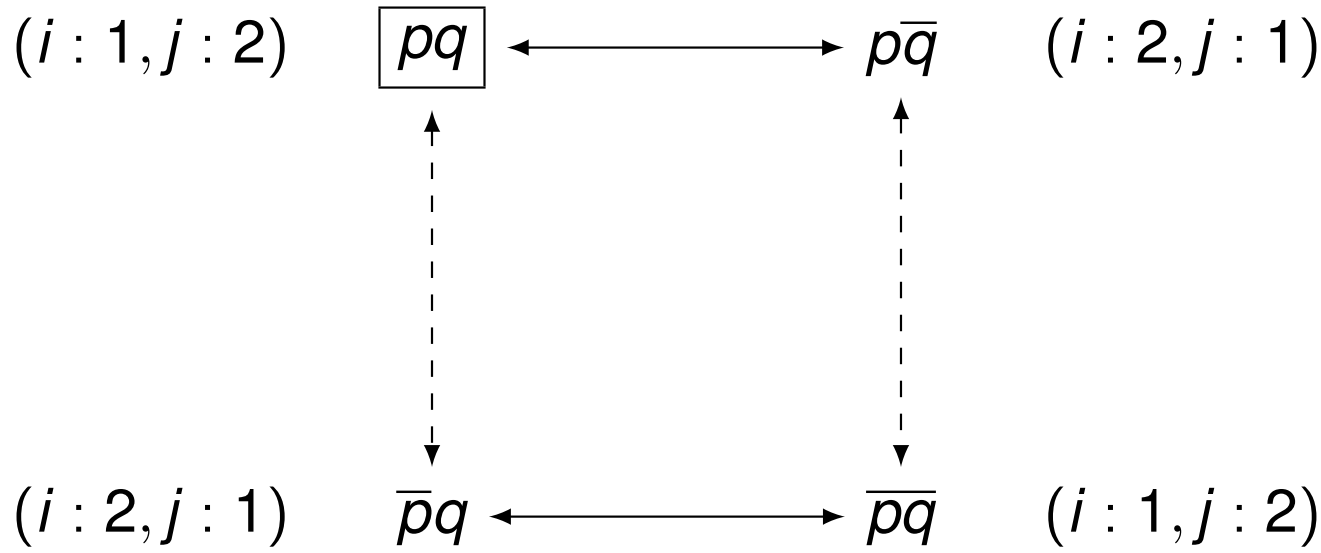
From Belief Based on S5 to Belief Based on D45

- Reinterpret $K_a\phi$ as "i is convinced of the truth of ϕ" (in the sense of no new information will change this conviction)
- A successful public lie $\neg\phi$ has as precondition that $\phi \wedge \neg K_a\phi$ is the case.
- A successful public lie $\neg\phi$ will cut the accessibility links of the audience to the real world (where ϕ is true).
- A D45 model is a model where all accessibility relations are serial, transitive and euclidean.
- A D45 model looks like a lollipop; for convenience we leave out the self-loops inside the lollipop from the pictures.
- Successful public lies are maps from D45 models to D45 models
- Successful public lie that $\neg\phi$ is modelled as relation change:
  from a to $(?\phi; a; ?\neg\phi) \cup (?\neg\phi; a; ?\neg\phi)$.
- See also [AvDW16], *True Lies*.

Effect of Public Lie that $p \leftrightarrow \neg q$

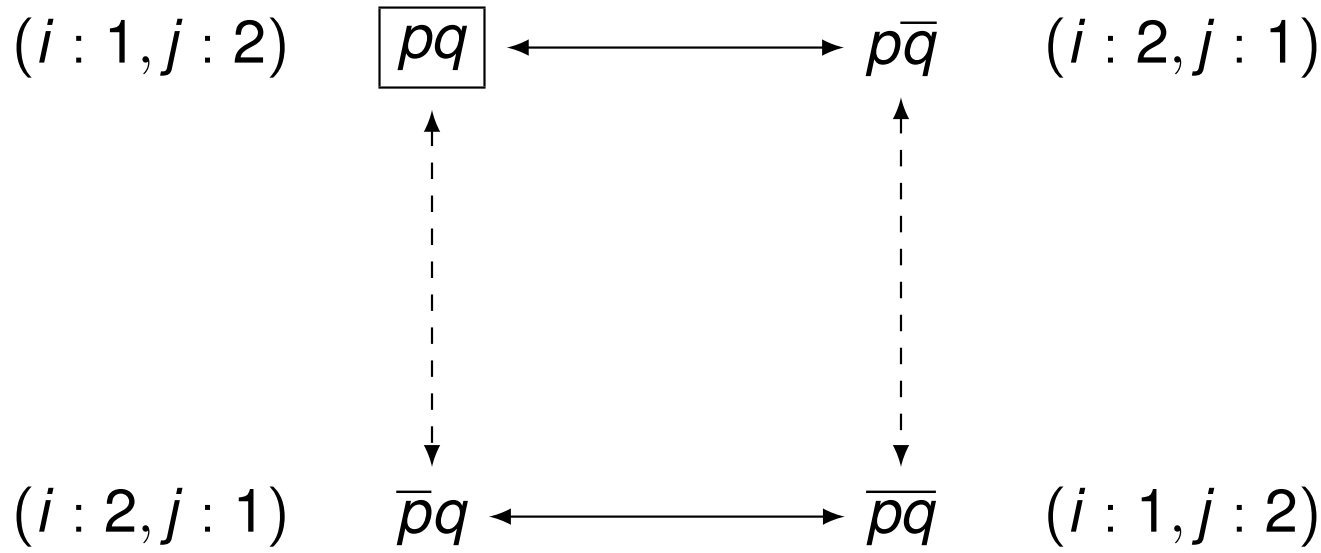


Effect of Public Lie that $p \leftrightarrow \neg q$

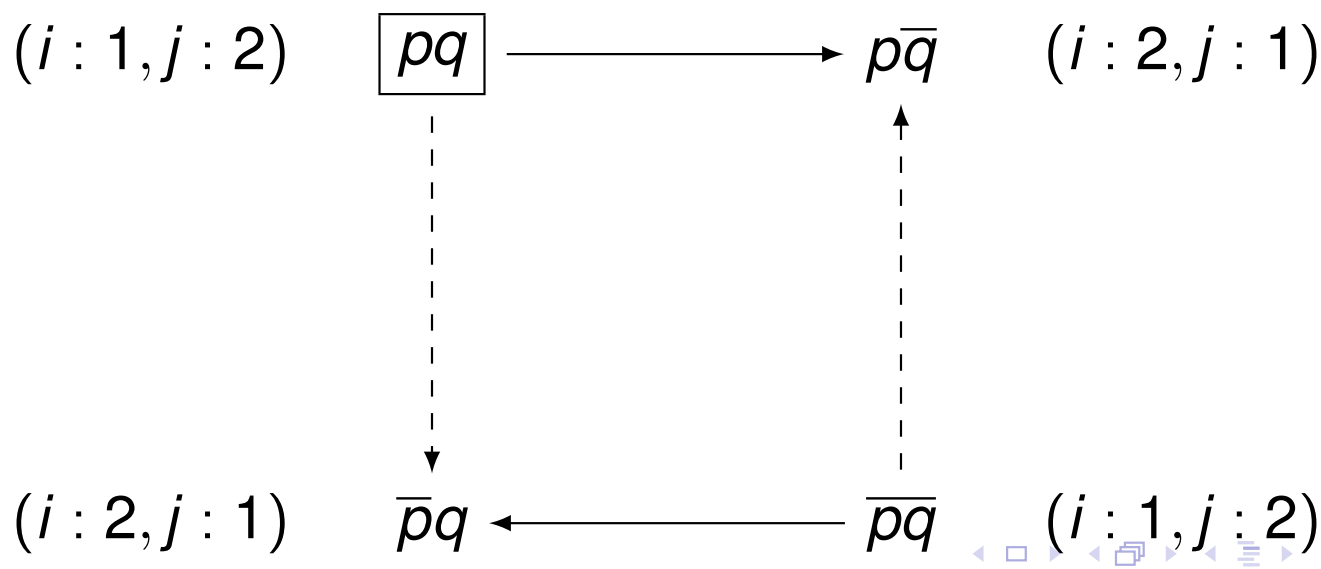


$[\sharp(p \leftrightarrow \neg q)]$

Effect of Public Lie that $p \leftrightarrow \neg q$

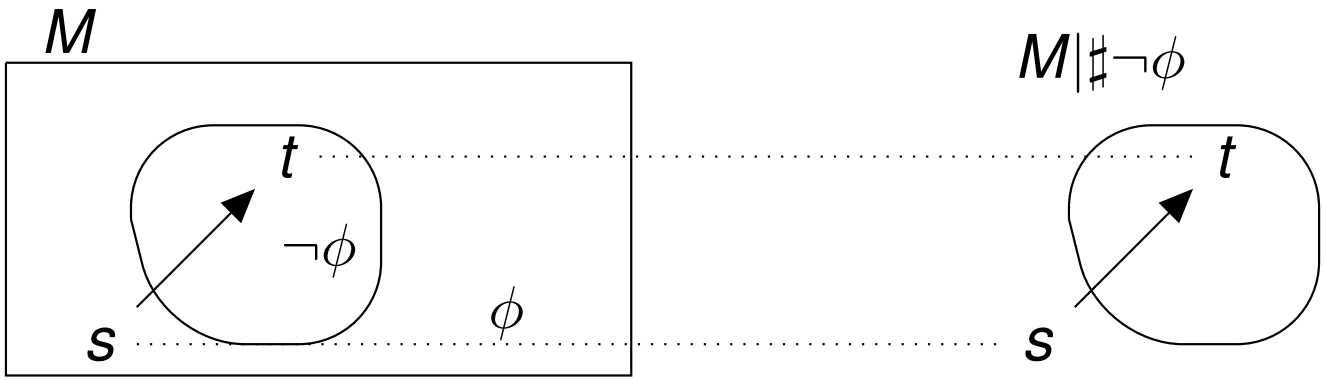


$[\sharp(p \leftrightarrow \neg q)]$



Key Validity for Public Lying

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



The formula $[\sharp\neg\phi]K_a\psi$ says that, in $M|\sharp\neg\phi$, all worlds t that are i-accessible from s satisfy ψ.
The corresponding worlds t in M are those i-accessible from s which satisfy $\neg\phi$.
We make our assertion on the right (the assertion about the model after the update) conditional on $\sharp\phi$ being executable, i.e., on $\phi \wedge \neg K_a\neg\phi$ being true.

From Weighted Belief Models to Conditional Displaced Neighbourhood Models

- Start out from accessibility relations R_i that are serial, transitive and euclidean (D45 relations).
- A conditional displaced neighbourhood model is a model where the neighbourhoods of w given ϕ need not include w.
- Given world w and agent i, $[w]_i$ is the set of worlds that are i-accessible from w. Note that w need not be in $[w]_i$.
- If R_i is serial and euclidean, then $[w]_i$ is an equivalence.
- The neighbourhoods of w for i, given ϕ, are the subsets X of $Y = [\![\phi]\!] \cap [w]_i$ with the property that the i-weight of X is larger than the i-weight of $Y - X$.

CDNL Calculus for Conditional D Neighbourhood Logic

(Taut)	All instances of propositional tautologies
(Def-K)	$K_a\phi \leftrightarrow \neg B_a(\neg\phi, \top)$
(Dist-K)	$K_a(\phi \to \psi) \to K_a\phi \to K_a\psi$
(Ser-K)	$K_a\top$
(PI-KB)	$B_a(\phi, \psi) \to K_a B_a(\phi, \psi)$
(NI-KB)	$\neg B_a(\phi, \psi) \to K_a \neg B_a(\phi, \psi)$
(M)	$K_a(\phi \to \psi) \to B_a(\chi, \phi) \to B_a(\chi, \psi)$
(EC)	$K_a(\phi \leftrightarrow \psi) \to B_a(\phi, \chi) \to B_a(\psi, \chi)$.
(D)	$B_a(\phi, \psi) \to \neg B_a(\phi, \neg\psi)$
(N)	$B_a(\phi, \psi) \to B_a(\phi, \phi \wedge \psi)$

$$\frac{\phi \to \psi \quad \phi}{\psi}\ \text{(MP)} \qquad \frac{\phi}{K_a\phi}\ \text{(Nec-K)}$$

CDNL Calculus for Conditional D Neighbourhood Logic

(Taut)	All instances of propositional tautologies
(Def-K)	$K_a\phi \leftrightarrow \neg B_a(\neg\phi, \top)$
(Dist-K)	$K_a(\phi \to \psi) \to K_a\phi \to K_a\psi$
(Ser-K)	$K_a\top$
(PI-KB)	$B_a(\phi, \psi) \to K_a B_a(\phi, \psi)$
(NI-KB)	$\neg B_a(\phi, \psi) \to K_a \neg B_a(\phi, \psi)$
(M)	$K_a(\phi \to \psi) \to B_a(\chi, \phi) \to B_a(\chi, \psi)$
(EC)	$K_a(\phi \leftrightarrow \psi) \to B_a(\phi, \chi) \to B_a(\psi, \chi)$.
(D)	$B_a(\phi, \psi) \to \neg B_a(\phi, \neg\psi)$
(N)	$B_a(\phi, \psi) \to B_a(\phi, \phi \wedge \psi)$

$$\frac{\phi \to \psi \quad \phi}{\psi}\ \text{(MP)} \qquad \frac{\phi}{K_a\phi}\ \text{(Nec-K)}$$

Conjecture

This calculus is sound and complete for conditional displaced neighbourhood models.

Recovery from Public Lies

What Does It Mean to Recover from a Lie?

- Let's not worry about lie *detection* for now.
- We just assume that the source of the public update with $\neg\phi$ gets distrusted.
- This does not mean we recover and update with ϕ instead.
- We wish to model just the retraction from the public update with $\neg\phi$.
- Public opening of the mind for ϕ again; common realization that ϕ might be true.
- Notation for this: $\natural\phi$.
- $\natural\phi; !\phi$ models recovery from the $\neg\phi$ lie followed by public update with ϕ.

Recovery from Public Lies

- Suppose current epistemic state is given by a.
- The act of recovering from the public lie $\neg\phi$ is given by the relational change

$$a := a \cup a^{\smile} \cup (?\phi; a; ?\neg\phi; a^{\smile}; ?\phi)$$

- Explanation: we need to get back to an equivalence. To restore symmetry, add all $a^{\smile}$ arrows. Next, if you are in a ϕ situation s that is disbelieved, then you can recover the connection to any ϕ situation t that is disbelieved by first taking an a step inside the lollipop to a $\neg\phi$ world, and next taking a reverse a step out of the lollipop again to t.
- Use $\natural\phi$ for the operation of public recovery from the lie that $\neg\phi$.
- Precondition for this to succeed is that ϕ is actually true while the agents are certain of $\neg\phi$, that is: precondition is $\phi \wedge K_a\neg\phi$. Recall that K_a is interpreted as a D45 modality.

Key Validity for Recovery

- Use $K_a\check{}$ for the modal operator interpreted as the reverse of the knowledge relation for a.
- $$[\natural\phi]K_a\psi \leftrightarrow ((\phi \wedge K_a\neg\phi) \rightarrow K_a(\psi \wedge K_a\check{}\psi))$$
- This says: After recovery from the lie that $\neg\phi$, $K_a\psi$ is true iff the fact that the lie is believed $(\phi \wedge K_a\neg\phi)$ implies that ψ and $K_a\check{}\psi$ are known. The latter fact means that every a; $a\check{}$ step ends in a ψ world.
- Note that this opens the way for an axiomatisation with reduction axioms.

Further Work

- Prove the completeness conjecture
- Public lying and recovery from lies for conditional displaced neighbourhood models: calculus, completeness.
- Public lying and recovery from lies for weighted models: calculus, completeness.
- Result of public lying: the community of agents loses touch with reality.
- Why is this bad? Because the utilities of our actions in the world are determined by properties of the world, not by what agents *believe* about the world.
- Add agent-utilities to the model and connect up to Paolo Galeazzi's world [Gal17].

Bibliography

Hannah Arendt.

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

Thomas Agotnes, Hans van Ditmarsch, and Yanjing Wang.

True lies.
https://arxiv.org/abs/1606.08333, 2016.

Johan van Benthem, Hans van Ditmarsch, Jan van Eijck, and Jan Jaspars.

Logic in Action.
Internet, 2013.
Electronic book, available from www.logicinaction.org.

Jan van Eijck and Kai Li.

Conditional belief, knowledge and probability.
manuscript, CWI, 2016.

Paolo Galeazzi.

Play Without Regret.
PhD thesis, ILLC, University of Amsterdam, 2017.

David J.C. MacKay.

Information Theory, Inference, and Learning Algorithms.
Cambridge University Press, 2003.
Available from http://www.inference.phy.cam.ac.uk/mackay/itila/.

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*, pages 201–216, 1989.