Learning by Questions and Answers:

Belief Dynamics under Iterated Revision

Sonja Smets, University of Amsterdam

Based on joint work with Alexandru Baltag, University of University

Iterated Revision with Doxastic Information

PROBLEM: investigate the **long-term behavior of iterated learning** of **higher-level doxastic information**.

Learning: belief revision with new *true* information.

Certain or uncertain information about the answer to some specific *question*.

Long-term behavior: whether the learning process *comes to an end, stabilizing* the doxastic structure, or *keeps changing it forever*.

Higher-level (doxastic) information: may refer to the agents' own *beliefs*, or even to her *belief-revision policy* (her contingency plans for belief change).

Iteration

By "iteration", one may mean two things (and I mean them both!):

1. iterating the application of **the same belief-revision method**, but with possibly **different new inputs** (new true sentences);
2. repeatedly revising with **the same input** (the same new true sentence).

But why would anyone need to keep re-learning the same true sentence?!

Contrast with Classical Theory

Classical literature on Belief Revision deals only with **propositional** information.

In that context, the process of learning the same, true information **always comes to an end**: *the most* one can learn by iterated revisions is *the correct valuation* (which atomic sentences are true in the real world).

Moreover, in that context **it is useless to repeatedly revise** with the **same** information: after learning a *propositional* sentence φ once, learning it again would be **superfluous** (leaving the doxastic state *unchanged*).

The "Success" Axiom

This uselessness of repeated learning is captured by one of the AGM (Alchourrón, Gärdenfors, Makinson) axioms:

The **"Success" Postulate**

$$\varphi \in T * \varphi$$

(Here, T is an *initial "theory"* (belief set), φ a *new propositional formula* and $T * \varphi$ is the *new theory after learning* φ.)

Meaning: **"After learning φ, one believes φ."**

Hence, *any further learning of* φ *is superfluous.*

Why bother?

QUESTION: *Why should we worry about revision with higher-level doxastic sentences? Why would an agent revise her beliefs about her own beliefs?*

After all, *an introspective agent already knows what she believes and what not*! It may seem there is no new information, so there is no need for revision!

ANSWER: Because the new information may come "packaged" in this way, *explicitly refering to the agents' beliefs* in order to *implicitly convey some new information about reality*.

Example: Learning you are wrong

Suppose somebody truthfully tells you the following sentence φ:

"**You are wrong about p.**"

We interpret φ as saying that: $Bp \leftrightarrow \neg p$

"**Whatever you currently believe about (whether or not) p is false.**"

This is a *doxastic* sentence, but it *does convey new information about the real world*: after learning φ and using *introspection* (about your own current beliefs), you will *come to know whether p holds or not*, thus correcting your mistaken belief about p.

"Success" is a failure!

Note that φ *changes its value* by being learned: after learning φ, your new belief about p is true, so now φ has obviously become *false*!

But the Success Postulate asks you to believe (after learning φ) that φ is true! In other words, it forces you (as a principle of rationality!) to acquire false beliefs!

Conclusion: the "Success" axiom fails for doxastic sentences.

Repeated learning is impossible?!

After learning φ once, after that φ becomes false, but moreover you **know** it is false. So **you cannot possibly "learn" it again**: you cannot accept again that φ is true, if you have already accepted it as such before!

Learning twice a sentence such as φ ("Moore sentences") is not superfluous, but on the contrary: it is impossible.

So *repeated learning is still trivial* in this case, but *trivial in a different sense* than in the case of propositional information.

What is the general picture?

BUT THIS IS NOT TRUE IN GENERAL!

As we'll see, **repeated learning of the same (true) doxastic information is not always trivial**: it may give rise to "doxastic loops"!

More generally, **iterated revision with truthful higher-level information can be highly non-trivial**.

The long-term behavior will turn out to **depend both on the type of sentence that is learnt**, and on **the specific way in which the "learning" takes place** (in particular, the *reliability of the source*).

Belief-Revision "Policies" (or "Methods")

Since the AGM postulates do not uniquely determine the belief-revision operation, there are various proposals in the literature.

I will adopt a **SEMANTIC** point of view: we are given a **(pointed) model**, describing both the "*truth*" and the agent's "*beliefs*". Typically, this consists of a *structured* set of *possible worlds*, together with a designated world (the *"real" world*).

A belief-revision policy will thus be given by a "**semantic upgrade**": a systematic **model transformation**, that maps any such model into a new model.

SEMANTICS: Plausibility (Grove) Structures

A **(finite) plausibility frame** is a *finite set* S of "*states*" (or "*possible worlds*") together with a **connected preorder** $\leq \subseteq S \times S$, called *plausibility relation*.

"**Preorder**": *reflexive* and *transitive*.
"**Connected**": $\forall s, t(s \leq t \vee t \leq s)$.

Read $s \leq t$ as "*state s is at least as plausible as state t*".

NOTE: This is *the same* as a finite **Grove model of "spheres"**: the *minimal* (=most plausible) states form the *first sphere*, and together with the next most plausible states form the *second sphere* etc.

Plausibility Models

A **(finite, pointed) plausibility model** is just a pointed Kripke model $(S, \leq, \| \cdot \|, s_0)$ having a (finite) plausibility structure as its underlying frame.

I.e. a plausibility frame

$$(S, \leq)$$

together with

a *designated world* $s_0 \in S$, called the "**real world**",

and

a **valuation map**, assigning to each atomic sentence p (in a given set At of atomic sentences) some set $\|p\| \subseteq S$.

Example 1

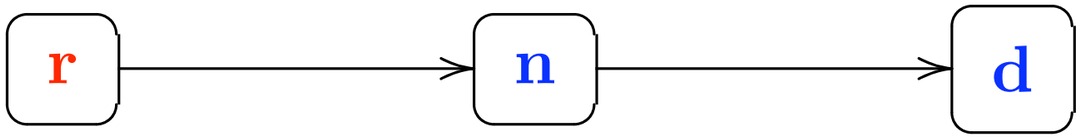
Consider a pollster (Charles) with the following **beliefs** about how a given voter (Mary) will vote:

He believes she will **vote Democrat**.

But in case this turns out wrong, he'd rather believe that she **won't vote** than accepting that she may vote Republican.

Let us assume that, **in reality** (unknown to Charles), Marry will **vote Republican**!

A Model for Example 1



Here, the valuation is trivial: each atom r, n, d is true at the corresponding world. The **real world** is r: Mary will vote Republican. We use *arrows* to represent **the converse plausibility relation** $\geq$ (going from less plausible to more plausible worlds), but **we skip all the loops and composed arrows** (obtainable by reflexivity and transitivity).

So Charles considers world d (voting Democrat) to be the most plausible, and world n (not voting) to be more plausible than world r.

(Conditional) Belief in Plausibility Models

$B\varphi$

A sentence φ is **believed** in (any state of) a plausibility model $(S \leq)$ if φ **is true in all the "most plausible" worlds**; i.e. in all "minimal" states in the set

$$Min_{\leq}S := \{s \in S : s \leq t \text{ for all } t \in S\}.$$

$B^P\varphi$

A sentence φ is **believed conditional on** P if φ **is true at all most plausible worlds satisfying P**; i.e. in all the states in the set

$$Min_{\leq}P := \{s \in P : s \leq t \text{ for all } t \in S\}.$$

Contingency Plans for Belief Change

We can think of conditional beliefs $B^{\varphi}\psi$ as "*strategies*", *or "contingency plans" for belief change*:
in case I will find out that φ was the case, I will believe that ψ was the case.

They can also be understood as a subjective ("doxastic") type of **non-monotonic conditionals**.

EXAMPLE: In Example 1, we have $Bd \wedge B^{\neg^d}n$.

Modelling Higher-Level Belief Revision

From a *semantic* point of view, higher-level belief revision is about "revising" the whole relational structure: *changing the plausibility relation (and/or its domain).*

A **relational transformer** is a *model-changing operation* α, that takes any plausibility model $\mathbf{S} = (S \leq, \|\cdot\|, s_0)$ and returns a new model $\alpha(\mathbf{S}) = (S', \leq', \| \cdot \| \cap S', s_0)$,

having as set of states some *subset* $S' \subseteq S$,
as valuation *the restriction of the original valuation to* S',
the same real world s_0 as the original model
(but *possibly a different order relation*).

Examples of Transformers

(1) Update $!\varphi$ (conditionalization with φ):

all the non-φ states are deleted and *the same plausibility order is kept between the remaining states.*

(2) Lexicographic upgrade $\Uparrow \varphi$:

all φ-worlds become "better" (more plausible) than all $\neg\varphi$-worlds, and *within the two zones, the old ordering remains.*

(3) Conservative upgrade $\uparrow \varphi$:

the "best" φ-worlds become better than all other worlds, and *in rest the old order remains.*

Explanation

• After a *conservative upgrade* $\uparrow \varphi$, the agent only comes to **believe** that φ (was the case); i.e. to allow only φ-worlds as the most plausible ones.

• The *lexicographic upgrade* $\Uparrow \varphi$ *has a more "radical" effect*: the agent comes to "**strongly believe**" φ; i.e. **accept** φ **with such a conviction** that *she considers all* φ*-worlds more plausible than all non-*φ *ones.*

• Finally, after an *update*, the agent comes to "**know**" φ in an **absolute, irrevocable sense**, so that *all non-*φ *possibilities are forever eliminated.*

A new Modality for Strong Belief

A sentence φ is **strongly believed** in a model $\mathbf{S}$ if the following two conditions hold

1. φ **is consistent with the agent's knowledge**:

$$\|\varphi\|_{\mathbf{S}} \neq \emptyset,$$

2. **all φ-worlds are strictly more plausible than all non-φ-worlds**:

$$t < s \text{ for every } t \in \|\varphi\|_{\mathbf{S}} \text{ and every } s \notin \|\varphi\|_{\mathbf{S}}.$$

It is easy to see that **strong belief implies belief**.

Strong Belief is Believed Until Contradicted by Evidence

Actually, strong belief is so strong that **it will never be given up except when one learns information that contradicts it!**

More precisely:

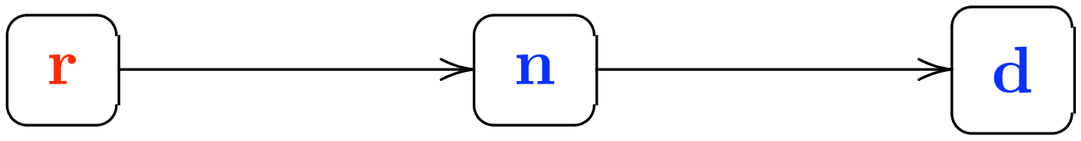
φ is **strongly believed** iff φ **is believed and is also conditionally believed given any new evidence (truthful or not) EXCEPT if the new information is known to contradict** φ; i.e. if:

1. $B\varphi$ holds, and
2. $B^{\theta}\varphi$ holds for every θ such that $\neg K(\theta \Rightarrow \neg\varphi)$.

Example

The "presumption of innocence" in a trial is a rule that asks the jury to hold a **strong belief in innocence** at the start of the trial.

In our example



• **The sentence** $n \vee d$ **is a strong belief** (although it is a **false** belief).

The sentence $r \vee d$ is not a strong belief.

The sentence Bd is itself a strong belief.

Updates are closed under composition

It is easy to see that a sequence of successive updates is equivalent to only one update:

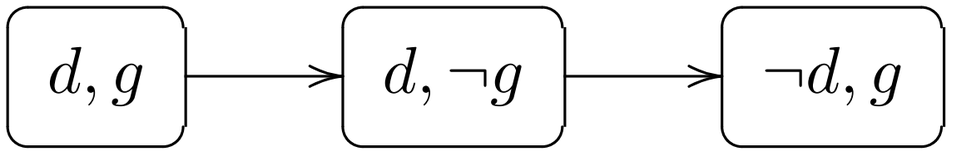
the effect of doing first an update $!\varphi$ then an update $!\psi$ is the same as the effect of doing the update $!(<!\varphi>\psi)$.

So instead of first announcing that φ is the case and then announcing that ψ is the case, I just announce from the start that φ is the case AND that ψ WOULD be the case AFTER I'd announce you φ.

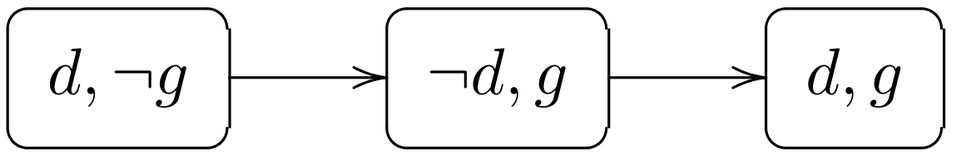
Upgrades are not closed under composition

Lexicographic upgrades are not closed under composition, a sequence of two such upgrades is not itself a lexicographic upgrade.

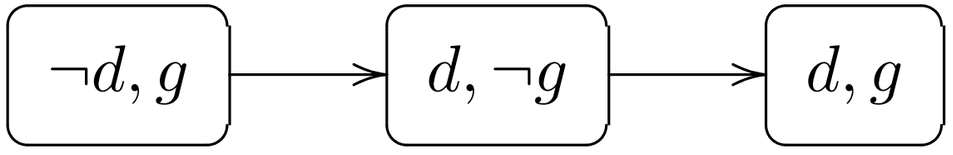
Counterexample Model



Do $\Uparrow (d \wedge g)$.



Then perform an upgrade $\Uparrow d$:



NO single (radical or conservative) upgrade can get us from the first to the last model!

NEEDED: A More General Notion of "Upgrade"

But doing two successive belief upgrades IS a meaningful belief change.

So we need to define **a more general notion of "belief upgrades"**, one that **subsumes radical and conservative upgrades**, and that **is closed under sequential composition**.

For this we'll work with **questions and answers**.

Revision as Partial Answer to a Binary Question

Till now have seen that there are various types of "revising" with φ.

They can now be understood as ways of **learning the answer to a binary question** ("Is φ the case?"), with different degrees of conviction.

Let's now generalize this setting.

Questions and Answers

Questions in a given language are **partitions of the state space** ($A^1, \ldots, A^n$, such that each cell A_i of the partition is definable by a sentence φ^i in the language.

Formally, a **question** is a (finite) family

$$\mathbf{Q} = \{\varphi^1, \ldots, \varphi^n\}$$

of sentences that are *exhaustive* and *mutually disjoint*, i.e.

$$\bigvee_{i=1,n} \varphi^i = True,$$

$$\varphi^i \wedge \varphi^j = False\,, \text{ for all } i \neq j.$$

Binary Questions

In particular, a **binary question** (corresponding to a Boolean attribute)

"Is it the case that φ or not?"

is given by a **two-cell partition**

$$\{A, \neg A\}$$

definable in the language by sentences

$$\{\varphi, \neg\varphi\}.$$

General Questions

A general (n-ary) question

"Which of the following is the case: $\varphi_1, \ldots, \varphi_n$?"

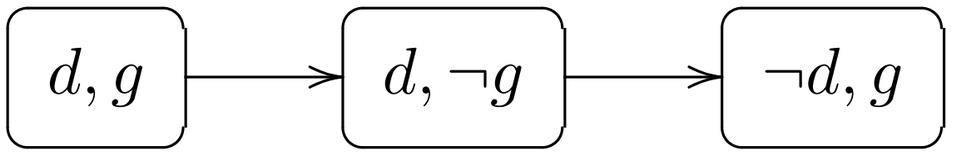
may have more cells

$$\{A^1, \ldots, A^n\},$$

each corresponding to a possible answer A^i defined by a sentence φ^i.

Examples

In the following example



the question "**Is g the case**?" is a binary question

$$\{g, \neg g\}$$

defining the partition

$$\{\,\{(d, g), (\neg d, g)\}, \{(d, \neg g)\}\,\}.$$

A ternary question

The question

"Which of the following is the case: d and g, not d and g, or not g?"

is a ternary question

$$\{d \wedge g, \neg d \wedge g, \neg g\}$$

defining the partition

$$\{\,\{(d, g)\}, \{(\neg d, g)\}, \{(d, \neg g)\}\,\}.$$

Learning uncertain answers: Upgrades

If the agent **learns some uncertain information about the answer** to a question $Q = \{\varphi^1, \ldots, \varphi^n\}$,
we may think that what is actually learnt *is a plausibility relation* $\leq$ *on a subset* $\mathcal{A} \subseteq \{\varphi^1, \ldots, \varphi^n\}$ *of the set of all possible answers.*

The agent learns with certainty that the answer is **not** one of the excluded ones in $Q \setminus \mathcal{A}$, and in rest she just comes to **assign some plausibility ranking between the remaining answers**, saying which answers are more plausible to her than others.

General Upgrades

We encode such an action as a plausibility frame

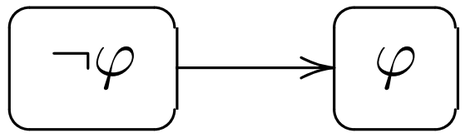
$$(\mathcal{A}, \leq)$$

whose **"worlds"** $\varphi_i \in \mathcal{A}$ **are mutually disjoint sentences**. For now, we'll assume that $\leq$ is a total **order** (rather than preorder): anti-symmetric.
We call such a frame a **(general) belief "upgrade"**

An upgrade describes a *type of belief change* induced by **gaining (both some certain and uncertain) information about the answer to a specific question**.

Examples: answers to binary questions

The plausibility frame



encodes **the radical upgrade** $\Uparrow \varphi$**:** it tells us that the "new information" learned by the agent is that φ is more plausible than $\neg\varphi$.

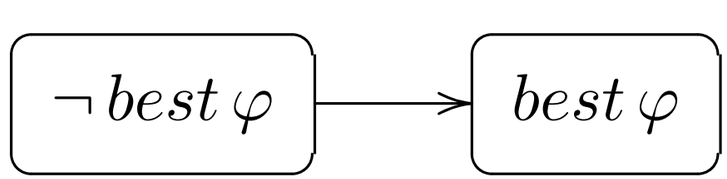
The plausibility frame

encodes **the update** $!\varphi$. The new information is just φ: no plausibility arrows means no uncertainty, so the agent is certain that the new info is correct.

More examples of answers to binary questions

The previous two upgrades describe ways to gain information or beliefs about the answer to the question $Q = \{\varphi, \neg\varphi\}$.

But we can also represent in this way the *conservative upgrade* $\uparrow \varphi$ in terms of **answering a DIFFERENT question** $\{best\,\varphi, \neg\, best\,\varphi\}$:



The agent "learns" (truthfully or not) that $best\,\varphi$ is more plausible than $\neg\, best\,\varphi$.

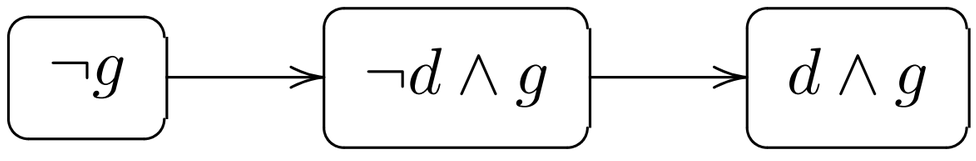
Examples of uncertain answers to ternary questions

Consider the ternary question $\{d \wedge g, \neg d \wedge g, \neg g\}$

Suppose our agent learns the answer:

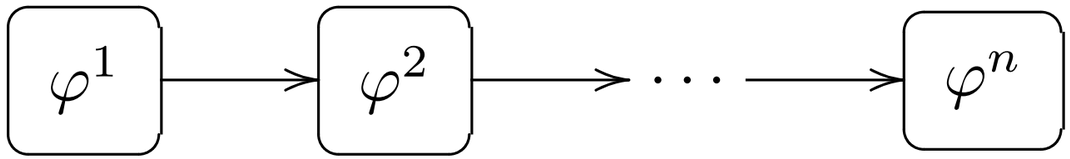
"I believe that d and g. But, well, even if not d, I still believe g"

Representing the Answer as a plausibility ordering of the three possible (definite) answers :



General upgrades

So, in general an upgrade will look like this:



and it represents an uncertain answer to a question of the form

$$Q = \{\varphi^1, \ldots, \varphi^n, \ldots \varphi^m\},$$

where $m \geq n$ and $\varphi^1, \ldots, \varphi^m$ define a partition.

We can write this upgrade formally as $(\varphi^1, \ldots, \varphi^n)$.

Representing the Revised Beliefs

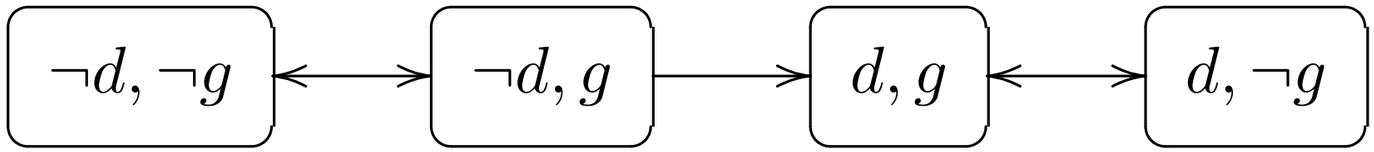
Note that all the models in the last few slides are **only representing the "answer"**, i.e. the **new information**.

They **do NOT** fully **represent the agent's new beliefs** after hearing the answer.

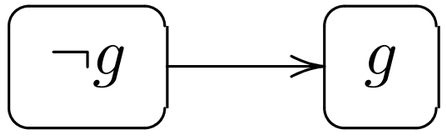
For this, we will have to somehow "**compose" the old model** (representing the agent's **prior beliefs**) with the model of the answer (representing **the answer** and **the agent's current belief about it**), to obtain a **new model** for the agent's **newly revised beliefs**.

Example

How does the original model of our agent



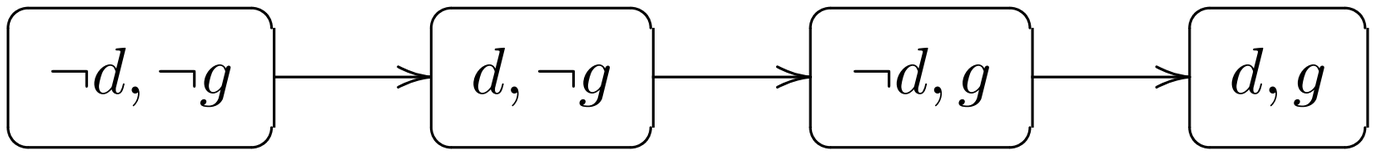
gets changed after she receives the answer



to the question $Q = \{g, \neg g\}$?

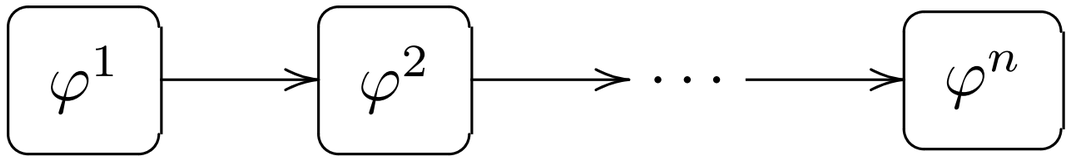
But we know this!

Well.., we already know how! This upgrade is nothing but the radical upgrade $\Uparrow g$, so all the g-worlds become more plausible than all non-g-worlds:



The General Rule

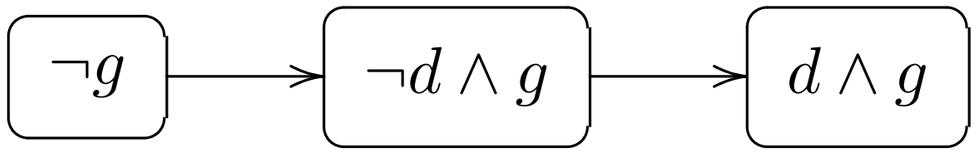
More generally, an upgrade $\alpha = (\varphi^1, \ldots, \varphi^n)$ of the form



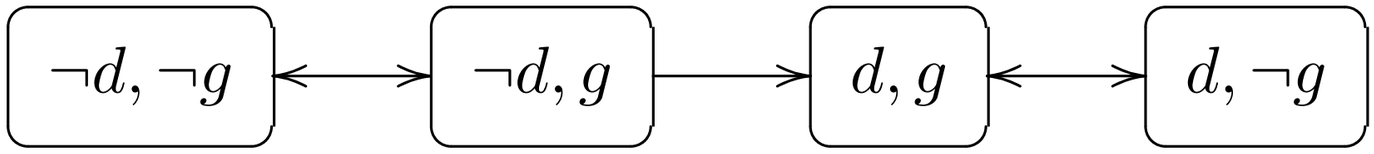
changes a given model **S** to a new model $\alpha(\mathbf{S})$, simply by **making all φ^{k+1}-worlds more plausible than all φ^k-worlds** (for every $k = 1, n-1$), while **keeping the old ordering within each φ^k-zone**.

Example

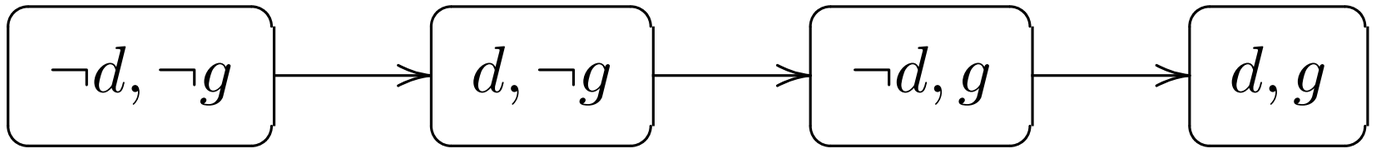
The answer $\alpha = (\neg g, \neg d \wedge g, d \wedge g)$



to the ternary question $\{d \wedge g, \neg d \wedge g, \neg g\}$ changes a given original belief model



to the new model



Iterating Upgrades

To study iterated belief revision, consider a **finite model** $\mathbf{S}_0 = (S, \leq_0, \|\cdot\|_0, s_0)$, and an **(infinite) sequence of upgrades**

$$\alpha_0, \alpha_1, \ldots, \alpha_n, \ldots$$

In particular, these can be updates

$!\varphi_0, !\varphi_1, \ldots, !\varphi_n, \ldots$

or conservative upgrades

$\uparrow \varphi_0, \uparrow \varphi_1, \ldots, \uparrow \varphi_n, \ldots$

or lexicographic upgrades

$\Uparrow \varphi_0, \Uparrow \varphi_1, \ldots, \Uparrow \varphi_n, \ldots$.

The iteration leads to **an infinite succession of upgraded models**

$$\mathbf{S}_0, \mathbf{S}_1, \ldots, \mathbf{S}_n, \ldots$$

defined by:

$$\mathbf{S}_{n+1} = \alpha_n(\mathbf{S}_n).$$

Iterated Updates Always Stabilize

OBSERVATION: For every initial finite model $\mathbf{S}_0$, every infinite sequence of updates

$$!\varphi_0, \ldots, !\varphi_n, \ldots$$

stabilizes the model after finitely many steps.

I.e. there exists n such that

$$\mathbf{S}_n = \mathbf{S}_m \text{ for all } m \geq n.$$

This is a *deflationary* process: the model keeps contracting until it eventually must reach a fixed point.

Iterated Upgrades Do Not Necessarily Stabilize!

This is NOT the case for arbitrary upgrades.

First, it is obvious that, if we allow for **false** upgrades, the revision may oscilate forever: the sequence

$$\Uparrow p, \Uparrow \neg p, \Uparrow p, \Uparrow \neg p, \ldots$$

will forever **keep reverting back and forth the order between the p-worlds and the non-p -worlds.**

Tracking the Truth

This is to be expected: such an "undirected" revision with mutually inconsistent pieces of "information" is not real learning.

But, surprisingly enough, **we may still get into an infinite belief-revision cycle, even if the revision is "directed" towards the real world**: i.e. even if we allow only upgrades that are always **truthful**!

Iterated Learning can produce Doxastic Cycles

PROPOSITION **For some initial finite models, there exist infinite cycles of truthful upgrades (that never stabilize the model).**

Even worse, this **still holds** if we restrict to iterations of **the same** truthful upgrade (with **one fixed sentence**): no fixed point is reached.

Moreover, when iterating **conservative** upgrades, **even the (simple, unconditional) beliefs may never stabilize, but may keep oscillating forever.**

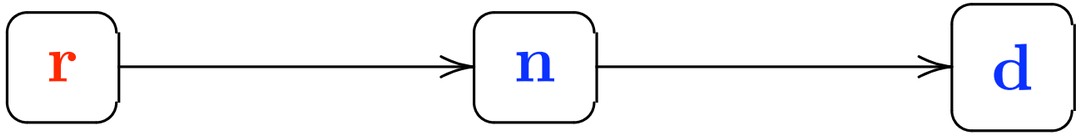
Iterating a Truthful Conservative Upgrade

In Example 1, suppose a trusted informer tells Charles the following true statement φ:

$$\mathbf{r} \vee (\mathbf{d} \wedge \neg \mathbf{Bd}) \vee (\neg \mathbf{d} \wedge \mathbf{Bd})$$

"*Either Mary will vote Republican or else your beliefs about whether or not she votes Democrat are wrong*".

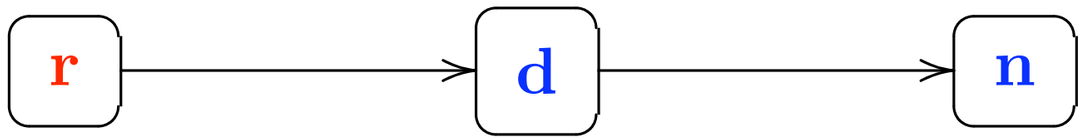
In the original model



the sentence φ is true in worlds r and n, but not in d.

Infinite Oscillations by Truthful Upgrades

Let's suppose that Charles **conservatively upgrades** his beliefs with this new true information φ. The most plausible state satisfying φ was n, so this becomes now the most plausible state overall:



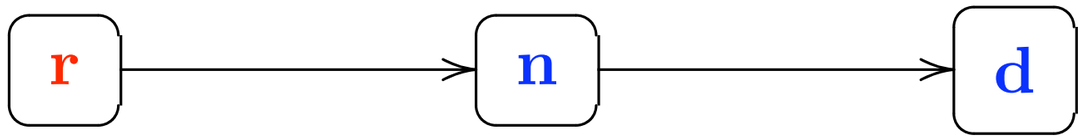
In this new model, the sentence φ is *again true at the real world (r)*, as well as at the world d. So **this sentence can again be truthfully announced**.

However, if Charles **conservatively upgrades again** with this new true information φ, he will promote d as the most plausible state, **reverting to the original model**!

Moreover, **not only the whole model (the plausibility order) keeps changing**, but Charles' (simple, un-conditional) **beliefs keep oscillating forever** (between d and n)!

Iterating Truthful Lexicographic Upgrades

Consider the same original model:

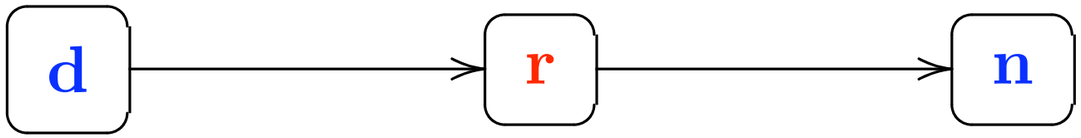


But now consider the sentence

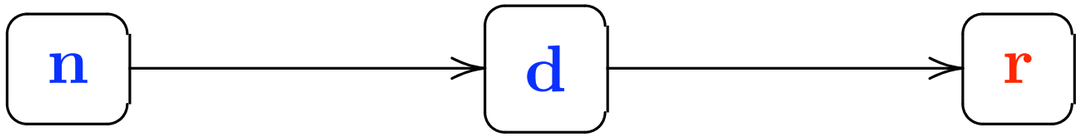
$$\mathbf{r} \vee (\mathbf{d} \wedge \neg \mathbf{B}^{\neg \mathbf{r}} \mathbf{d}) \vee (\neg \mathbf{d} \wedge \mathbf{B}^{\neg \mathbf{r}} \mathbf{d})$$

"*If you'd truthfully learn that Marry won't vote Republican, then your resulting belief about whether or not she votes Democrat would be wrong*".

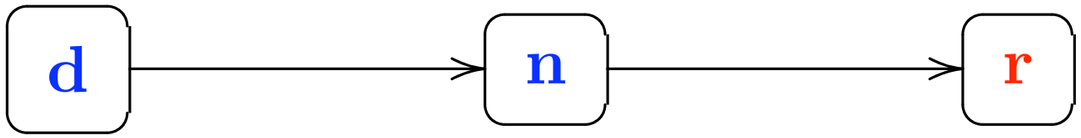
This sentence is true in the real world r and in n but not in d, so a **truthful lexicographic upgrade** will give us:



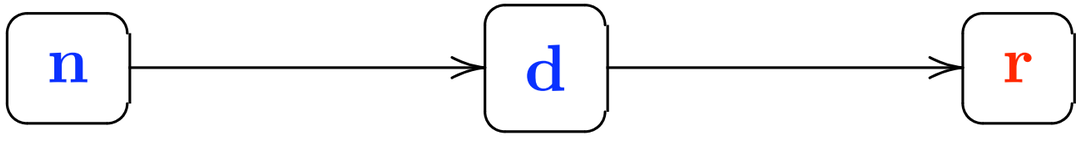
The same sentence is again true in (the real world) r and in d, so it can be again truthfully announced, resulting in:



Another truthful upgrade with this sentence produces



then another truthful upgrade with the same sentence **gets us back** to



Stable Beliefs in Oscillating Models

Clearly from now on the last two models **will keep reappearing, in an endless cycle**: as for conservative upgrades, the process never reaches a fixed point!

However, *unlike in the conservative upgrade example*, **in this example the simple (unconditional) beliefs eventually stabilize**: from some moment onwards, Charles correctly believes that the real world is r (vote Republican) and he will never lose this belief again!

This is a symptom of a more general phenomenon:

Beliefs Stabilize in Iterated Lexicographic Upgrades

THEOREM:

In any infinite sequence of truthful lexicographic upgrades $\{\Uparrow \varphi_i\}_i$ on an initial (finite) model $\mathbf{S}_0$, the set of most plausible states stabilizes eventually, after finitely many iterations.

From then onwards, the simple (un-conditional) beliefs stay the same (despite the possibly infinite oscillations of the plausibility order).

Upgrades with Un-conditional Doxastic Sentences

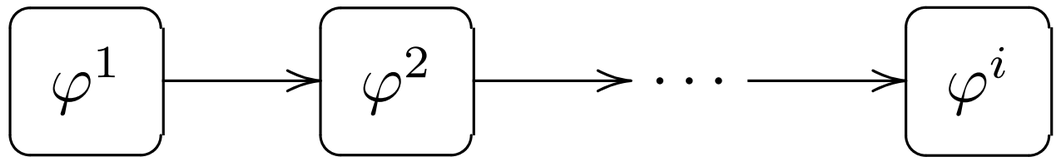
Moreover, if the infinite sequence of lexicographic upgrades $\{\Uparrow \varphi_i\}_i$ consists only of sentences belonging to the language of basic doxastic logic (allowing only for simple, un-conditional belief operators) then the model-changing process eventually reaches a fixed point: after finitely many iterations, the model will stay unchanged.

As we saw, this is not true for conservative upgrades.

Generalization

In our paper, we also **generalize this theorem to arbitrary upgrades, provided they are "correct"**.

An upgrade α, given by



is **"correct"** if **the most plausible answer** φ^i (according to α) **is true** (in the real world).

For lexicographic upgrades, "correct"=truthful; but in general (e.g. for conservative ones) this is not the case.

Converging to the Truth?

So simple beliefs stabilize after an infinite series of truthful lexicographic upgrades (not so with conservative upgrades). **But under what conditions do these beliefs stabilize on the Truth?**

Strongly informative upgrade streams

An upgrade with φ is called "**strongly informative**" on a pointed model **S** iff φ is **not already believed** at (the real world of) **S**. I.e. **S** satisfies $\neg B\varphi$.

Now, an upgrade stream $\{\Uparrow \varphi_n\}_n$ is "**strongly informative**" if *each of the upgrades is strongly informative at the time when it is announced:*

i.e. in the iteration, we have that

$$\mathbf{S}_n \models \neg B\varphi_n$$

Belief correcting upgrade and streams

Call an upgrade $\Uparrow \varphi$ "**belief-correcting**" on $\mathbf{S}$ iff φ is actually believed to be FALSE at $\mathbf{S}$. I.e.

$$\mathbf{S} \models B\neg\varphi.$$

Now, an upgrade stream is called "**belief-correcting**" if each of the upgrades is belief-correcting at the time when it is announced:

$$\mathbf{S}_n \models B\neg\varphi_n.$$

NOTE: "belief correcting" $\Rightarrow$ "strongly informative" (The converse fails.)

Maximal Strongly informative streams

An upgrade stream is a **"maximally" strongly-informative** (OR **"maximally belief-correcting"**), truthful stream if:

- (1) it is strongly-informative (OR belief-correcting) and truthful, and
- (2) it is maximal with respect to property (1): it cannot be properly extended to any stream having property (1).

So a strongly informative truthful stream is "maximal" iff it is **either infinite or** if, in case it is finite (say, of length n) then **there exists NO upgrade** $\Uparrow \varphi_{n+1}$ which would be **truthful and strongly informative** on the last model $\mathbf{S}_n$.

The results

1. **Every maximally belief-correcting lexicographic upgrade stream** $\{\Uparrow \varphi_n\}_n$ (starting on a given finite model $\mathbf{S}$) **is finite and converges to true beliefs**; i.e. in its final model $\mathbf{S}_n$, all the beliefs are true.

2. **Every maximally strongly-informative lexicographic upgrade stream** $\{\Uparrow \varphi_n\}_n$ (starting on a given finite model $\mathbf{S}$) **is finite and stabilizes the beliefs on FULL TRUTH**; i.e. in its final model $\mathbf{S}_n$, all beliefs are true and all true sentences are believed.

Note

But note that the last conclusion is NOT necessarily equivalent to saying that the set of most plausible worlds coincides in the end with only the real world!

The reason is that the language may not be expressive enough to distinguish the real world from some of other ones; and so the conclusion of 2 can still hold if the most plausible worlds are these other ones...

The above results do NOT hold for any other belief-revision methods except lexicographic (and conditioning).

Conclusions

- Iterated upgrades may never reach a fixed point: conditional beliefs may remain forever unsettled.

- When iterating truthful lexicographic upgrades, the simple (non-conditional) beliefs converge to some stable belief.

- If we repeatedly (lexicographically) upgrade with THE SAME sentence in BASIC DOXASTIC logic, then all conditional beliefs eventually stabilize.

- In iterated truthful radical upgrades that are maximal strongly-informative, all believes converge to the truth and all true sentences are believed.

Other types of upgrades do not have these last positive properties.

This is not the full story!

We can extend the above positive result regarding repeated upgrades **beyond** basic doxastic logic, allowing *various forms of "knowledge" operators in the language*.

Still, there exist *important conditional-doxastic sentences* lying outside this fragment (e.g. "**Surprise**"-sentence in the Surprise Examination Puzzle) for which repeated lexicographic upgrades nevertheless stabilize the whole model!