Belief Dynamics for Conditionals

Ronnie Hermens & Jan-Willem Romeijn

University of Groningen

August 14, 2012

Outline

The example of Sly Pete

Gibbard [1980] nicely illustrates that the interpretation of conditionals may depend on context. His story involves Sly Pete and Mr. Stone playing poker on a Mississippi river boat.

Henchmen Red and Green have differing pieces of information: Red knows that Stone has the upper hand, while Green knows that Pete knows Stone's hand.

Conditionals for Green and Red

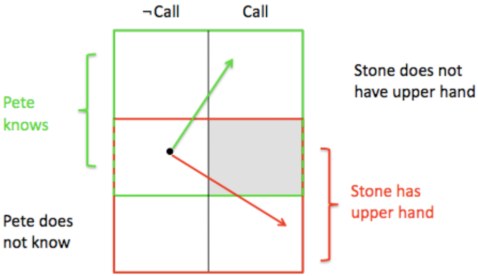
We can simultaneously entertain the truth and falsity of a counterfactual conditional:

1. It is true that If Pete called, he would have won.
2. It is false that If Pete called, he would have won.

Received wisdom has it that, if these conditionals obtain a truth value, they do so relative to a context. The two conditionals somehow express different propositions.

Conditionals are context-sensitive

The seemingly opposite conditionals are true in virtue of satisfying the antecedent of the conditional sentence in different ways.



In the actual world Pete does not call and Stone has the upper hand. For Red, if Pete had called he would have lost, period. But for Green, if Pete had called, this would have been because he had the upper hand after all.

The triviality results of Lewis

Context dependence also shows up in the debate over Lewis' [1975] triviality arguments against Stalnaker's hypothesis. For all P we have

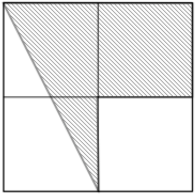
$$P(A \to B) = P(B|A).$$

Applying the law of total probability (TP), Bayes' rule (BR), and then applying Stalnaker's hypothesis (SH) leads to triviality:

$$\begin{aligned} P(A \to B) &\overset{\text{TP}}{=} P(A \to B|B)P(B) + P(A \to B|\neg B)P(\neg B) \\ &\overset{\text{BR}}{=} P_B(A \to B)P(B) + P_{\neg B}(A \to B)P(\neg B) \\ &\overset{\text{SH}}{=} P_B(B|A)P(B) + P_{\neg B}(B|A)P(\neg B) \\ &\overset{\text{BR}}{=} P(B|A \wedge B)P(B) + P(B|A \wedge \neg B)P(\neg B) \\ &= 1\,P(B) + 0\,P(\neg B) = P(B) \end{aligned}$$

Stalnaker's hypothesis as independence

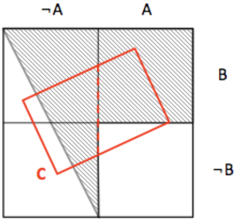
It is useful to consider Stalnaker's hypothesis as an independence requirement on probability assignments. We assume that within A the truth condition of $A \to B$ is B, so that $P(A \to B|A) = P(B|A)$.



Stalnaker's hypothesis then is that the conditional is independent of its antecedent: $P(A \to B|\neg A) = P(A \to B|A)$.

Independence and Bayes' rule

On the assumption of Bayes' rule, the independence also holds within every context proposition C. We should be able to apply Stalnaker's hypothesis to $P(\cdot|C)$ as well:

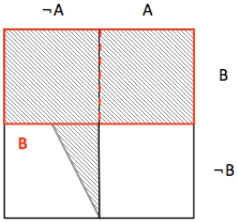


This is the starting point for the resolutions of triviality by McGee [1989] and by Stalnaker and Jeffrey [1994]:

$$P(A \rightarrow B|\neg A \wedge C) = P(A \rightarrow B|A \wedge C).$$

Triviality and updating

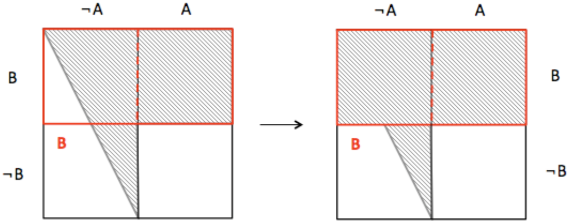
Lewis' triviality result follows if we fill in B and $\neg B$ for the context C. For an update with B, Stalnaker's hypothesis entails $P(A \rightarrow B|\neg A \wedge B) = 1$.



A similar argument leads to $P(A \rightarrow B|\neg A \wedge \neg B) = 0$. This means that $P(A \rightarrow B|B) = 1$ and $P(A \rightarrow B|\neg B) = 0$ so that triviality follows.

Triviality and context shifting

We can, however, maintain Stalnaker's thesis if we adapt the extension of the conditional in the possible world semantics halfway the update:



This is a direct violation of Bayesian updating. Next to eliminating possible worlds, we adapt the proposition associated with the conditional sentence.

Modelling context-sensitivity

We have two independent motivations for incorporating context-sensitivity in the semantics of conditionals:

1. There are convincing examples in which the truth values of a conditional depend on context.
2. We can avoid the triviality result of Lewis if we allow the interpretation of conditionals to shift during an update.

Our main objectives are to bring this context-sensitivity to bear on the semantics and belief dynamics of conditionals.

Earlier formalizations of context

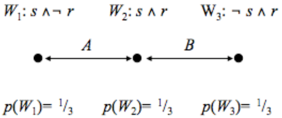
There are a number of earlier formalizations of the context-sensitivity of conditionals.

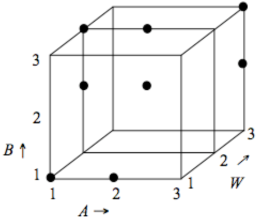
- Van Fraassen [1976] presents a minimal logic for maintaining Stalnaker's hypothesis and a context-sensitive semantics for conditionals.
- McGee [1989] defines an update rule for beliefs about conditionals that respects the aforementioned independence.
- Stalnaker and Jeffrey [1994] present conditionals as random variables, and thereby make them explicitly context-sensitive.
- Lindström [1996] gives a treatment of context-sensitivity with the machinery of belief revision.

Van Fraassen has an elaborate possible worlds semantics but does not present a dynamics for beliefs. Others present the dynamics but do not provide a semantics, or lack on both counts.

Interpretation shifts

We take inspiration from earlier work on violations of Bayesian updating due to interpretation shifts. Van Benthem [2003] considers violations of conditionalization in dynamic epistemic logic.





Romeijn [2011] models this violation by so-called knowledge structures: relations among worlds are captured by their internal structure. These are updated on new information, causing shifts in interpretation.

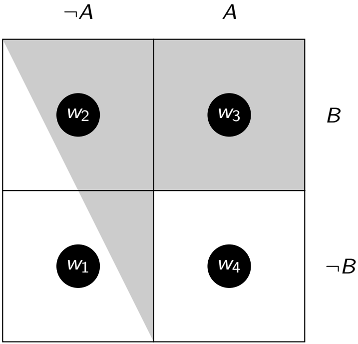
Advantages of this approach

We believe we thereby improve on earlier models of context-sensitivity.

- The semantics of conditional sentences is intuitive.
  - Possible worlds are states of affairs squared with states of mind, which are captured in the internal structure of worlds.
  - Conditionals are sets of such worlds, whose extension is determined by a Ramsey test.
- The belief dynamics for conditionals is defined properly and fits the semantics.
  - Updating on states of affairs goes by standard Bayesian conditionalization.
  - A full update also involves conditionalization over possible states of mind, and thus over the internal structure of worlds.
  - This second Bayesian update captures the context-sensitive interpretation of conditionals.

Making room for conditionals 1

- A model with 4 possible worlds is enough to capture all Boolean combinations of two propositions A and B.
- But for conditionals we need more!
- Idea: instead of adding more possible worlds, capture the notion of conditionals by adding structure to the possible worlds.
- Prototype: ‘knowledge structures’.

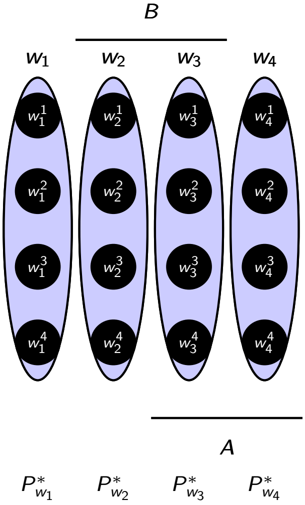


Making room for conditionals 2: worlds with structures

- Start with the simple worlds.
- And add some structure.
- Every world has its own copy of the total set of possible worlds.
- And a (lexicographic) probability function $P^*_{w_i} = (P^*_{w_i,0}, P^*_{w_i,1})$ over this set.
- $P^*_{w_i,0}$ expresses the simple belief one 'should' have in world w_i:

$$P^*_{w_i,0}(w_i^i) = 1.$$

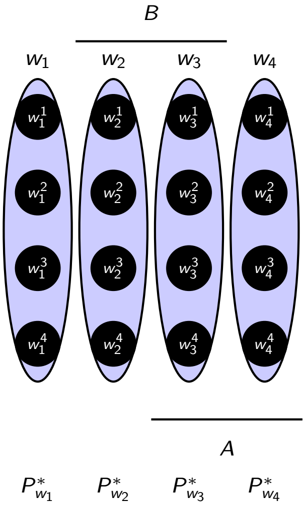
- But there is an (infinitesimal) probability that w_i is not the case, and $P^*_{w_i,1}$ expresses the belief given that w_i is not the case.



The probability of a conditional

- Evaluate the conditional in each of the possible worlds and weigh according to the probability of that world:

$$P(A \to B) = \sum_{i=1}^{4} P(w_i) P^*_{w_i}(B|A)$$
$$= P(w_1) P^*_{w_1,1}(B|A)$$
$$+ P(w_2) P^*_{w_2,1}(B|A)$$
$$+ P(B \wedge A).$$



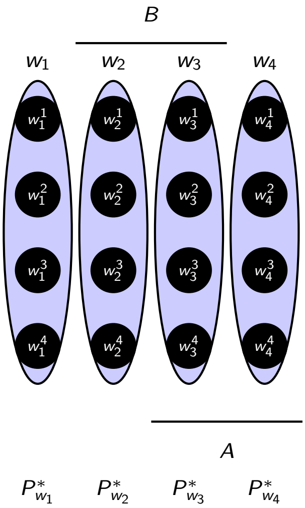
- Taking $P^*_{w_i,1}(.) = P(.|\neg w_i)$ one obtains Stalnaker's hypothesis:

$$P(A \to B) = P(B|A).$$

Comments on the framework 1

- Conditionals are not understood as propositions in the sense that they are not associated with a set of possible worlds.
- Rather, they resemble the stochastic variables as understood by Stalnaker-Jeffrey [1994]:

$$X_{A\to B}(w_1) = X_{A\to B}(w_2) = P(B|A),$$
$$X_{A\to B}(w_3) = 1,\ X_{A\to B}(w_4) = 0,$$
$$P(A \to B) = E(X_{A\to B}).$$



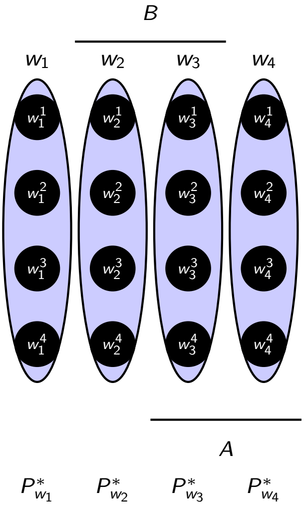
Comments on the framework 2

- But unlike Stalnaker-Jeffrey, we do have a clear update rule: double Bayesian updating.
- An update of P, and an update of the P^*'s.
- Thus updating on X leads to

$$P_X(A \to B) = \sum_w P(w|X) P^*_w(B|A \wedge X).$$

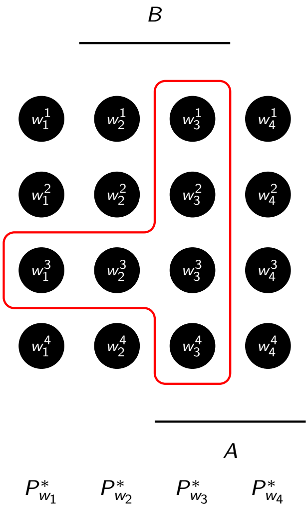
- In particular:

$$P_B(A \to B) = 1.$$



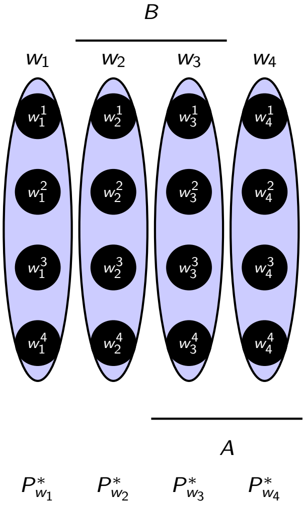
Modifying the framework 1

- To have conditionals be propositions more worlds have to be added.
- The cheapest way is to give up the idea of 'internal structure'.
- Obtaining 16 possible worlds.
- And the $A \rightarrow B$-worlds within the red border.
- But by giving up sructure, we feel we are giving up understanding of the model. Making it more artificial.



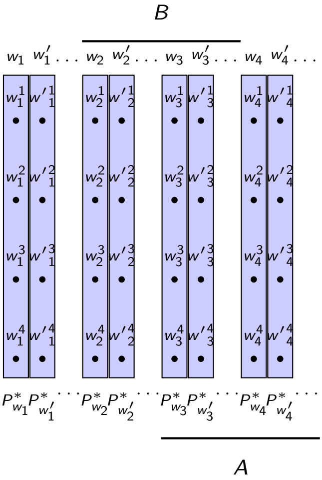
Modifying the framework 2

- $P^*_{w_i}$ models the epistemic state in w_i but... many possible epistemic states may be compatible with a single world.
- There may be $B \wedge \neg A$-worlds in which one believes $A \to B$ to be true, and ones in which one believes it to be false.
- And those will be considered to be distinct possible worlds.
- For every w_i, for every admissible lexicographic probability function $P^*_{w_i}$ there is a possible world which is a copy of w_i in which $P^*_{w_i}$ is the true epistemic state.



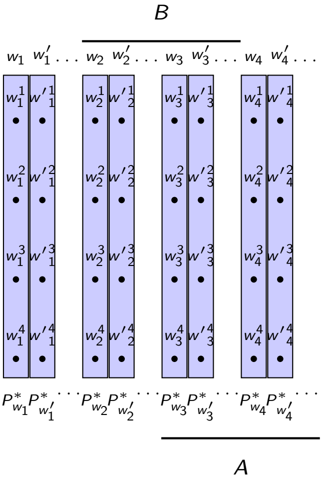
Modifying the framework 2

- $P^*_{w_i}$ models the epistemic state in w_i but... many possible epistemic states may be compatible with a single world.
- There may be $B \wedge \neg A$-worlds in which one believes $A \rightarrow B$ to be true, and ones in which one believes it to be falls.
- And those will be considered to be distinct possible worlds.
- For every w_i, for every admissible lexicographic probability function $P^*_{w_i}$ there is a possible world which is a copy of w_i in which $P^*_{w_i}$ is the true epistemic state.



Modifying the framework 3

- In every possible world w_i' either $A \to B$ is true or false depending on $P^*_{w_i'}$, making conditionals 'epistemic propositions'.
- But we still have a double update, thus learning X changes both P and all the $P^*_{w_i'}$.
- The set of possible worlds in which $A \to B$ is true thus changes as belief changes.
- Within this framework conditionals are context-sensitive propositions.



Acknowledging the context

- The context is tied up with ones belief state. That is, the context changes if and only if the state of belief changes.
- Upon learning X the degrees of belief shift from P to P_X. And the context $\mathfrak{C}$ shifts to $\mathfrak{C}_X$.
- There are as many conditionals as there are contexts and this context will be specified as

$$A \rightarrow B \;\rightsquigarrow\; A \stackrel{\mathfrak{C}}{\rightarrow} B$$

- Thus upon learning X the degrees of belief shift from $P(A \stackrel{\mathfrak{C}}{\rightarrow} B)$ to $P_X(A \stackrel{\mathfrak{C}_X}{\rightarrow} B)$.
- Allowing the conditional to be context-sensitive on these context-changes already blocks the famous triviality results.

Blocking Lewis' proof

The main step in Lewis' proof involves showing that the probability of a conditional equals the probability of its consequent. Besides the law of total probability it uses the following step:

(BR=Bayes' Rule, CI=Context Independence, SH=Stalnaker's Hypothesis)

$$
\begin{aligned}
P(A \overset{\mathfrak{C}}{\to} B|B) &\overset{\text{BR}}{=} P_B(A \overset{\mathfrak{C}}{\to} B) \overset{\text{CI}}{=} P_B(A \overset{\mathfrak{C}_B}{\to} B) \\
&\overset{\text{SH}}{=} P_B(B|A) \equiv \frac{P_B(B \wedge A)}{P_B(A)} \overset{\text{BR}}{=} \frac{P(B \wedge A|B)}{P(A|B)} \\
&\equiv \frac{P(B \wedge A \wedge B)}{P(B)} \frac{P(B)}{P(A \wedge B)} = 1
\end{aligned}
$$

$\Rightarrow$ We deny CI but also the idea that the expression $P_B(A \overset{\mathfrak{C}}{\to} B)$ even makes sense since it mixes the two contexts $\mathfrak{C}$ and $\mathfrak{C}_B$.

Updating belief in a conditional

- What is $P_X(A \overset{\mathfrak{C}_X}{\to} B)$ if not $P(A \overset{\mathfrak{C}}{\to} B|X)$?
- Only one option if one wishes to uphold Bayes' Rule for simple propositions and the Thesis:

$$
\begin{aligned}
P_X(A \overset{\mathfrak{C}_X}{\to} B) &\overset{\text{SH}}{=} P_X(B|A) \equiv \frac{P_X(B \wedge A)}{P_X(A)} \\
&\overset{\text{BR}}{=} \frac{P(B \wedge A|X)}{P(A|X)} \equiv \frac{P(B \wedge A \wedge X)}{P(X)} \frac{P(X)}{P(A \wedge X)} \\
&\equiv P(B|A \wedge X)
\end{aligned}
$$

- Conversely, taking up this update rule as an axiom is sufficient for obtaining Stalnaker's hypothesis for updated probability functions.
- Needless to say, the update rule holds in our framework.

Sly Pete and the update rule v2.0

- Back to the poker game.

Sly Pete and the update rule v2.0

- We are interested in the conditional $C \to W$ with

$$C = \text{“Pete calls”},$$
$$W = \text{“Pete wins”}.$$

- There are two contexts (Red and Green):

$$\color{red}{\mathfrak{C}} = \text{“Stone has the upper hand”},$$
$$\color{green}{\mathfrak{C}} = \text{“Pete knows Stone's hand”}.$$

- With the previous formula we find

$$P_{\color{red}{\mathfrak{C}}}(C \stackrel{\color{red}{\mathfrak{C}}}{\to} W) = P(W|C \wedge \color{red}{\mathfrak{C}}) = 0,$$
$$P_{\color{green}{\mathfrak{C}}}(C \stackrel{\color{green}{\mathfrak{C}}}{\to} W) = P(W|C \wedge \color{green}{\mathfrak{C}}) = 1.$$