THE RELEVANCE OF AWARENESS

MICHAEL FRANKE AND TIKITU DE JAGER

ILLC, Universiteit van Amsterdam
M.Franke@uva.nl, S.T.deJager@uva.nl

Relevance is a crucial concept in linguistics, but also a notoriously vague notion. Recently, some formal decision-theoretic notions of relevance have been applied successfully to linguistics (van Rooij 2003; van Rooij 2004), but none of these captures the impact of modalized sentences, in particular possibility statements. We treat decision problems of possibly *unaware* agents (cf. Fagin and Halpern 1988; Modica and Rustichini 1994) and give an update procedure that captures *becoming aware* of further contingencies. We define the relevance of such updates, and hint at the pragmatic reasoning surrounding possibility statements in dialog.

1. Decisions, relevance and awareness

It's Alice's birthday. Bob, our Bayesian baker, is uncertain whether Alice likes cake and thus faces a DECISION PROBLEM, i.e. a structure $D = \langle S, P, A, U \rangle$, where

S is a set of relevantly distinct STATES OF THE WORLD,
P is a PROBABILITY DISTRIBUTION on S,
A is a set of ACTIONS, and
U is a UTILITY FUNCTION, $U \colon S \times A \rightarrow \mathbb{R}$.

Bob's decision problem (shown in Figure 1(a)) contains states s_1 and s_2 for Alice's preferences, probabilities for these possibilities (e.g. $P(s_2) < P(s_1)$, not shown), actions a_c ('bake a cake') and $a_\varnothing$ ('do nothing'), and utilities for all possible combinations of states and actions ($U(s_1, a_c) = 0.5$: even if Alice doesn't like cake, Bob still does). Bob quickly checks the EXPECTED UTILITY of his actions

$$\mathrm{EU}_D(a) \stackrel{\text{def}}{=} \sum_{s \in S} P(s) \times U(s, a)$$

and concludes that a cake it will be, since baking the cake is an action which maximizes expected utility (in this case it is the only one). If he learns that the actual state is in $T \subseteq S$, he will update his beliefs and recalculate his actions' EXPECTED UTILITY AFTER LEARNING that $T \subseteq S$:

$$\mathrm{EU}_D(a, T) \stackrel{\text{def}}{=} \sum_{s \in S} P(s|T) \times U(s, a).$$

A reasonable measure for the relevance of such information is the following variant

	a_c	$a_\varnothing$
$s_1 : c$	1	0
$s_2 : \neg c$	0.5	0

(a) Care for cake?

	a_c	$a_\varnothing$
$s_1 : c, \neg e$	1	0
$s_2 : \neg c, \neg e$	0.5	0
$s_3 : c, e$	-5	0
$s_4 : \neg c, e$	-5	0

(b) Bad eggs

	a_c	a_s	$a_\varnothing$
$s_1 : c, \neg e$	1	0.4	0
$s_2 : \neg c, \neg e$	0.5	0.4	0
$s_2 : c, e$	-5	0.4	0
$s_4 : \neg c, e$	-5	0.4	0

(c) Shortbread

	a_c	a_s	$a_\varnothing$
$s_1 : c, \neg e$	1	0.4	0
$s_2 : \neg c, \neg e$	0.5	0.4	0
$s_3 : c, e$	-5	0.4	0
$s_4 : \neg c, e$	-5	0.4	0
$s_5 : c, \neg e$	1	-10	0
$s_6 : \neg c, \neg e$	0.5	-10	0
$s_7 : c, e$	-5	-10	0
$s_8 : \neg c, e$	-5	-10	0

(d) Allergies

a_c	Bake chocolate cake
a_s	Bake shortbread
$a_\varnothing$	Do nothing
c	Alice likes chocolate cake
e	The eggs are off
-5	Eating rotten eggs (nasty)
-10	Allergic reaction (dangerous)

(e) Key

Figure 1: Decision problems resulting from various updates.

of the VALUE OF SAMPLE INFORMATION (cf. Raiffa and Schlaifer 1961):[1]

$$\mathrm{VSI}_D(T) \stackrel{\text{def}}{=} \max_{a \in A} \mathrm{EU}_D(a, T) - \mathrm{EU}_D(\mathrm{BA}(D), T),$$

where $\mathrm{BA}(D)$ is the set of actions with maximal expected utility in D.[2] But then, how to deal with a case where Alice brings up something genuinely new as in (1)?

(1) ALICE: Hmm, the eggs might be off, did you think of that?

Intuitively, Alice's remark in (1) does not eliminate previously considered alternative states, but *brings new options into consideration*. The decision problem in Figure 1(a) represented Bob's awareness of the situation, and his implicit assumption that the eggs are fresh. After Alice mentions the possibility that they might be off, $\Diamond e$, Bob's decision problem is updated to (something like) Figure 1(b). Another possibility is that Alice could suggest a new action ($\Diamond a_s$) leading to Figure 1(c):

(2) ALICE: You could make shortbread instead.

Or she might give a possible consequence of a given action ($\Diamond(a_s; r)$, where r stands for 'have an allergic reaction'), bringing Bob to Figure 1(d):

(3) ALICE: But hold it! Will your allergies react to shortbread?
BOB: Glad you reminded me to check, but no: I'm only allergic to nuts.

All of these changes to Bob's representation of the situation seem to be additive. The difficulty is simply stated: where do we find the elements being added? In particular,

[1] The idea is that information T is *irrelevant* iff all of your behavior lacking T is never a mistake in the light of T (i.e. you are not doing anything 'wrong' without the information T, so you don't need it).
[2] The expected utility of a set of actions is the average of the expected utilities of its elements.

new states need probabilities and the results of new actions need utilities. The notion of awareness (cf. Fagin and Halpern 1988) suggests an answer: these elements were already present 'in the background', but not yet explicitly brought into consideration.

2. Formal system

The system we construct has two components: a BACKGROUND MODEL $\mathfrak{M}$ and an AWARENESS STATE $\mathfrak{A}$ which filters out certain possibilities that the agent is not explicitly considering: we write $\mathfrak{M}{\restriction}\mathfrak{A}$ for such a FILTERED MODEL, where the function-restriction notation is not intended literally but rather to suggest this filtering process. From any model M (background or filtered) we can 'read off' a decision problem $\delta(M)$. While $\delta(\mathfrak{M})$ describes the *actual* decision problem faced by the agent, $\delta(\mathfrak{M}{\restriction}\mathfrak{A})$ describes the decision problem the agent is *aware* of. Anticipating somewhat, the update with a possibility formula φ will be performed on the awareness state only: $\delta(\mathfrak{M}{\restriction}\mathfrak{A})$ gives the decision problem before the update, $\mathfrak{A}[\varphi]$ is the awareness state updated with φ, so $\delta(\mathfrak{M}{\restriction}\mathfrak{A}[\varphi])$ is the decision problem we get after filtering the background model through this updated awareness state.

We assume throughout finite sets Φ of primitive propositions and Γ of actions. A MODEL M is a structure $\langle \mathcal{P}, \mathcal{A}, \mathcal{W}, \mathcal{O}, P, \mathcal{U} \rangle$, where

$\mathcal{P}$ is a subset of Φ (primitive propositions);
$\mathcal{A}$ is a subset of Γ (actions);
$\mathcal{W}$ is a (finite) set of WORLDS;
$\mathcal{O}$ is a (finite) set of OUTCOMES (think: 'future states of the world');
P is a PROBABILITY DISTRIBUTION on $\mathcal{W}$;
$\mathcal{U}$ is a UTILITY FUNCTION for outcomes: $\mathcal{U} \colon \mathcal{O} \to \mathbb{R}$.

A world $w \in \mathcal{W}$ is a pair $\langle V_w, R_w \rangle$ where $V_w \colon \Phi \to \{0, 1\}$ is the VALUATION FUNCTION (the current state of the world) and $R_w \colon \Gamma \to \mathcal{O}$ is the RESULT FUNCTION telling the outcome of each action; an outcome $\omega \in \mathcal{O}$ is simply a valuation function $V_\omega \colon \Phi \to \{0, 1\}$. (The current state of the world does not necessarily define its future evolution: there might be multiple worlds with the same propositional valuation but where actions have different outcomes.) Probabilities are defined on *worlds* and thus embrace future contingencies, while utilities are given on *outcomes*. We also define $\mathcal{V}_\mathcal{W} \stackrel{\text{def}}{=} \{V_w \ ; \ w \in \mathcal{W}\}$ for convenience in referring only to the valuations in some set of worlds.

(C1) **Representing awareness.** Given a valuation V and a set $\mathcal{P} \subseteq \Phi$ of primitive propositions, we write $V^\mathcal{P}$ for the largest set of valuations that agree with V on all propositions in $\Phi \setminus \mathcal{P}$. If $\mathcal{V}$ is a set of valuations and we can find some $\mathcal{P} \subseteq \Phi$ such that $\mathcal{V} = V^\mathcal{P}$ for some $V \in \mathcal{V}$, then we say that $\mathcal{V}$ REPRESENTS AWARENESS of the primitive propositions $\mathcal{P}$ and UNAWARENESS of (and implicit belief about) all others.[3] A set $\mathcal{W}$ of worlds represents awareness of $\mathcal{P}$ iff $\mathcal{V}_\mathcal{W}$ does; any

[3] This conflation of awareness and implicit belief is not as restrictive as it might seem, since we can still

structure containing both worlds and propositions "satisfies constraint (C1)" if the worlds represent awareness of the propositions.

The background model. The background model $\mathfrak{M}$ is a model as defined above, with $\mathcal{P}_{\mathfrak{M}} = \Phi$ and $\mathcal{A}_{\mathfrak{M}} = \Gamma$. It satisfies (C1) (i.e., $\mathcal{W}_{\mathfrak{M}}$ represents awareness of $\mathcal{P}_{\mathfrak{M}}$) and in addition we require that every possible result function occur with every possible valuation in some world. We associate with $\mathfrak{M}$ a STEREOTYPICAL CAUSALITY FUNCTION $\mathcal{S}_{\mathfrak{M}} : \mathcal{V}_{\mathcal{W}_{\mathfrak{M}}} \rightarrow \wp(\mathcal{W}_{\mathfrak{M}})$, which gives for each valuation the worlds with that valuation whose outcomes are *subjectively stereotypical* according to the agent. These outcomes 'spring to mind' when the agent entertains a possibility (see the stereotypicality constraint (C2) below).[4]

The awareness state. An awareness state $\mathfrak{A} = \langle \mathcal{P}_{\mathfrak{A}}, \mathcal{A}_{\mathfrak{A}}, \mathcal{W}_{\mathfrak{A}} \rangle$ is defined with reference to a background model $\mathfrak{M}$:

$\mathcal{P}_{\mathfrak{A}}$ is the set of primitive propositions being attended to, with $\mathcal{P}_{\mathfrak{A}} \subseteq \Phi$;
$\mathcal{A}_{\mathfrak{A}}$ is the set of actions being explicitly considered, with $\mathcal{A}_{\mathfrak{A}} \subseteq \Gamma$
$\mathcal{W}_{\mathfrak{A}}$ is the set of worlds being entertained, with $\mathcal{W}_{\mathfrak{A}} \subseteq \mathcal{W}_{\mathfrak{M}}$;

We require in addition two consistency constraints: the awareness state should satisfy (C1) (otherwise the model we read off from it will not do so), and also the following:

(C2) **Stereotypicality constraint.** For every world $w \in \mathcal{W}_{\mathfrak{A}}$, $\mathcal{S}_{\mathfrak{M}}(V_w) \subseteq \mathcal{W}_{\mathfrak{A}}$. That is, among the outcomes an agent entertains for any current state of affairs, she must always entertain at least the causally stereotypical ones.

The filtered model, $\mathfrak{M}{\restriction}\mathfrak{A}$, carries over the propositions, actions and worlds from $\mathfrak{A}$ and is defined as follows: $\mathfrak{M}{\restriction}\mathfrak{A} \stackrel{\text{def}}{=} \langle \mathcal{P}_{\mathfrak{A}}, \mathcal{A}_{\mathfrak{A}}, \mathcal{W}_{\mathfrak{A}}, \mathcal{O}', P', \mathcal{U}' \rangle$, where[5]

$$\begin{aligned} \mathcal{O}' &\stackrel{\text{def}}{=} \bigcup \{ R_w[\mathcal{A}_{\mathfrak{A}}] \; ; \; w \in \mathcal{W}_{\mathfrak{A}} \}, \\ P'(w) &\stackrel{\text{def}}{=} P_{\mathfrak{M}}(w | \mathcal{W}_{\mathfrak{A}}), \\ \mathcal{U}' &\stackrel{\text{def}}{=} \mathcal{U}_{\mathfrak{M}} {\restriction} \mathcal{O}_{\mathfrak{A}}. \end{aligned}$$

Reading off a decision problem. When reading off $\delta(M)$ from a model M, we cannot always take worlds in $\mathcal{W}_M$ as states. Worlds that only differ in their outcomes for actions that are *not being considered* (if the model is filtered) should be combined into the same state. We do this via a partition on $\mathcal{W}_M$, and the complete decision

represent beliefs probabilistically: an agent can ENTERTAIN a possible state of affairs (explicitly consider a world where that state obtains) while assigning it probability zero.

[4] Inasmuch as the stereotypical causality function represents the agent's expectations, we expect it to be closely related to her probability distribution. The details of the relation are somewhat unclear, and seem not to be important for our purposes.

[5] Here $R_w[\mathcal{A}_{\mathfrak{A}}]$ and $\mathcal{U}_{\mathfrak{M}} {\restriction} \mathcal{O}_{\mathfrak{A}}$ refer to ordinary function image and restriction, not update and filtering.

problem $\delta(M) = \langle S, P, A, U\rangle$ is given by:

$$\begin{aligned}
S &\stackrel{\text{def}}{=} \big\{\{w' \in \mathcal{W}_M \,;\, V_w = V'_w \text{ and } R_w{\restriction}\mathcal{A}_M = R_{w'}{\restriction}\mathcal{A}_M\} \,;\, w \in \mathcal{W}_M\big\};\\
P(s) &\stackrel{\text{def}}{=} P_M(s) = \textstyle\sum_{w\in s} P_M(w);\\
A &\stackrel{\text{def}}{=} \mathcal{A}_M;\\
U([w]_{\equiv_M}, a) &\stackrel{\text{def}}{=} \mathcal{U}_M(R_w(a)).
\end{aligned}$$

Now that we can read off a decision problem from a filtered model, all that remains is to define the three kinds of awareness updates exemplified by (1)–(3).[6]

Updating with $\Diamond p$. In becoming aware of p, the agent realizes "p might be differently valued to what I have been assuming". So she adds worlds with valuations the same as those she already entertains, except for the value of p. This leaves unspecified the outcomes: $\mathcal{S}_{\mathfrak{M}}$ picks out only the *stereotypical* worlds.[7] (See Figure 1(b).)

$$\langle \mathcal{P}, \mathcal{A}, \mathcal{W}\rangle[\Diamond p] = \langle \mathcal{P} \cup \{p\}, \mathcal{A}, \mathcal{W} \cup \mathcal{W}'\rangle$$
$$\text{where } \mathcal{W}' = \bigcup\{\mathcal{S}_{\mathfrak{M}}[V_w^{\{p\}}] \,;\, w \in \mathcal{W}\}.$$

This update preserves (C1), and the new awareness state will satisfy (C2). Moreover, updating more than once with $\Diamond p$ will have no additional effect.

Updating with $\Diamond a$. $\langle \mathcal{P}, \mathcal{A}, \mathcal{W}\rangle[\Diamond a] = \langle \mathcal{P}, \mathcal{A} \cup \{a\}, \mathcal{W}\rangle$. The worlds already come equipped with their (stereotypical) a-outcomes (see Figure 1(c)), so (C1) and (C2) are trivially preserved, and applying the update repeatedly has no additional effect.

Updating with $\Diamond(a;p)$. The possibility that a might lead to p introduces *non*-stereotypical worlds: it might be that a leads to p for very strange reasons. The update is explicitly concerned with the possibility that p should be *brought about by* a, not simply hold in the current state of the world.

$$\langle \mathcal{P}, \mathcal{A}, \mathcal{W}\rangle[\Diamond(a;p)] = \langle \mathcal{P}, \mathcal{A}, \mathcal{W} \cup \mathcal{W}'\rangle$$
$$\text{where } \mathcal{W}' = \{\langle V_w, R_w^{(a;p)}\rangle \,;\, w \in \mathcal{W}\}$$

and $R_w^{(a;p)}$ is the same outcome function as R_w except that at the outcome of a the valuation of p is inverted. This update again preserves (C1) and (C2), and will not change the awareness state the second time if performed twice in succession.[8]

[6] We define awareness updates only for primitive propositions and actions. That is, we do not identify any natural-language modality with our $\Diamond$ operator — after all, simply mentioning a proposition should also induce awareness of it. Some modals, however, may be used specifically to induce awareness updates.

[7] It may help to gloss this in procedural terms. First we add p to the propositions the agent is aware of. Next, for each world w in $\mathcal{W}$, take the valuations that agree with V_w except for possibly at p (that is, trivially V itself and one other), collect their stereotypical worlds (according to $\mathcal{S}_{\mathfrak{M}}$) and add them all.

[8] The non-stereotypicality of these worlds means that the same update *may* be performed informatively more than once, if new worlds are added in the meantime.

3. Relevance, awareness and pragmatic inference

Given these awareness updates, we can take over the notion of relevance that we introduced in Section 1 without substantial amendment. We only have to define the expected utility after becoming aware of contingency x: $\mathrm{EU}_{\delta(\mathfrak{M}\upharpoonright\mathfrak{A})}(a, \Diamond x) = \mathrm{EU}_{\delta(\mathfrak{M}\upharpoonright\mathfrak{A}[\Diamond x])}(a)$. This measure of RELEVANCE OF AWARENESS may take a pivotal role in the pragmatic reasoning triggered when agents are purposefully *made aware* of contingencies in dialog.[9]

For instance, Bob in example (1) may reason as follows: Alice is trying only to make me aware of a possibility; the awareness update had better be relevant; this is so only if it changes my course of action, so it should convince me not to bake a cake.[10] Such reasoning has no place yet in our model, since we deal only with the single-agent perspective. An obvious extension is to include explicit uncertainty about (higher-order beliefs about) the background model. That is, from assuming that Alice believes her information is relevant, Bob can conclude that she also believes sufficiently strongly that the eggs are off. If Bob considers Alice expert (well-informed; competent) then his conclusions about what she believes influence his own beliefs. Similar pragmatic reasoning lets Bob conclude in (2) that baking shortbread does not require eggs: the uncertainty and expertise relates only to the outcomes of actions rather than to the probabilities of worlds.

Example (3) involves a different kind of uncertainty, and here separating Alice's belief that her move is relevant from her expertise in the matter is crucial. The awareness update is in fact irrelevant, since Bob's allergies (as he knows) will not be triggered by anything he bakes.[11] Alice is nonetheless motivated by relevance: it is her uncertainty about Bob's awareness that leads her to mention the possibility.

Reasoning about beliefs about awareness requires fully-fledged awareness models in the style of Fagin and Halpern 1988. Our system focusses on dynamic awareness updates; explicit modeling of the entire relevance-based pragmatic reasoning process in a single framework seems a most promising direction for further research.

Bibliography

Fagin, R. and Halpern, J. Y.: 1988, Belief, awareness and limited reasoning, *Artificial Intelligence* 34, 39–76

Modica, S. and Rustichini, A.: 1994, Awareness and partitional information structures, *Theory and Decision* 37, 107–124

Ozbay, E. Y.: 2007, Unawareness and strategic announcements in games with uncertainty, in D. Samet (ed.), *Proceedings of TARK XI*, pp 231–238

[9] Just as with an eliminative update, awareness updates are relevant iff some action which maximized expected utility before the update no longer does so after. In particular, since updating with $\Diamond a$ doesn't change existing probabilities or utilities, $\Diamond a$ is relevant iff a becomes the unique best action after update.

[10] The "reasoning refinement" of Ozbay 2007 can be recast in similar terms as a relevance-based inference.

[11] That is, the update merely transforms his implicit belief into *the same* explicit belief.

Raiffa, H. and Schlaifer, R.: 1961, *Applied Statistical Decision Theory*, MIT Press
van Rooij, R.: 2003, Questioning to resolve decision problems, *Linguistics and Philosophy* 29, 727–763
van Rooij, R.: 2004, Utility, informativity and protocols, *Journal of Philosophical Logic* 33(4), 389–419